Gene-environment and gene-gene interactions in myopia

Alfred Pozarickij

A thesis submitted for the degree of Doctor of Philosophy at

Cardiff University in October 2019

School of Optometry and Vision Sciences

Abstract

Motivated by the release of the UK Biobank data and the lack of documented gene-environment (GxE) and gene-gene (GxG) interactions in myopia, I sought to apply various statistical tools to provide a quantitative assessment of the interplay between environmental and genetic risk factors shaping refractive error.

The comparison between the two different risk measurement scales with which GxE interactions can be identified suggested that the additive risk scale can lead to a more informative perspective about refractive error aetiology.

The evaluation of two indirect methods for detecting genetic variants affecting refractive error via interaction effects suggested the enrichment of GxG and GxE among the variants that display marginal SNP effects.

For genetic variants already known from prior GWAS studies to influence refractive error, genetic effect sizes were highly non-uniform; individuals from the tails of the refractive error distribution (i.e. high myopes and hyperopes) displayed much larger effects compared to individuals in the middle of the distribution (i.e. emmetropes).

Prediction of refractive error using GxE interactions indicated that although some of the variance of refractive error could be explained by a risk score constructed using interaction effects, the contribution of GxE was already accounted for by a risk score constructed using marginal SNP effects only.

Although a handful of candidate genes were identified using multifactor dimensionality reduction technique, none displayed compelling evidence of involvement in a GxG interaction. There was, however, suggestive evidence that the candidate genes constitute a genetic interaction network which is regulated by hub gene *ZMAT4*.

In summary, the analyses reported in this thesis provide further support for the challenging nature of definitively identifying loci involved in GxE and GxG interactions. The thesis provides several guidelines that future studies could take into account to obtain more insightful results regarding the extent of interactions in refractive error.

Acknowledgements

I would like to thank my supervisors Professor Jeremy Guggenheim and Dr Cathy Williams for allowing me to explore my ideas during this studentship. I would like to thank Prof Guggenheim for his patience, regular feedback and stimulating discussions.

I would like to thank my fellow PhD students, Neema Ghorbani, Denis Plotnikov, Yu Huang, Rupal Shah, Ryan Bartlett and Petar Markov, who helped me to get through my studentship as well as providing a relaxed and fun environment to work in.

I would like to thank my family and Inna Korableva for providing support and encouragement during the most difficult times of my PhD.

Finally, I would like to thank the National Eye Research Centre for sponsoring my PhD programme.

Table of Contents

List of Tables

List of Figures

List of abbreviations

AOSW	Age of onset of spectacle wear
AUC	Area under the curve
avMSE	Average mean spherical equivalent
BP	Base pair
CADD	Combined annotation dependent depletion
CHR	Chromosome
CI	Confidence interval
CQR	Conditional quantile regression
CREAM	Consortium for Refractive error and myopia
D	Dioptres
EduYears	Age at the completion of education
eQTL	Expression quantitative trait locus
GxE	Gene-environment interaction
GxG	Gene-gene interaction
GWAS	Genome-wide association study
LD	Linkage disequilibrium
LRT	Likelihood ratio test
MAF	Minor allele frequency
MDR	Multifactor dimensionality reduction
MR	Meta-regression
OLS	Ordinary least squares
P	P-value
PolyPhen	Polymorphism phenotyping
PC	Principal component
PRS	Polygenic risk score
PRS-I	Interaction polygenic risk score

PRS-G	Marginal polygenic risk score
RERI	Relative excess risk due to interaction
RINT	Rank inverse normal transformation
SIFT	Sorting intolerant from tolerant
SNP	Single nucleotide polymorphism
UM-MDR	Unified model-based multifactor dimensionality reduction
UniEdu	Education exposure representing degree status
vQTL	Variance quantitative trait locus

Chapter 1

Introduction to myopia and interactions

1.1. Myopia

1.1.1. Definition and classification

Myopia, also known as short-sightedness or near-sightedness, is among the most common eye disorders in the world (Rudnicka et al., 2016). Figure 1.1. illustrates the contribution from most common eye disorders leading to visual impairment. In healthy patients with clear vision, distant objects focus directly on or near the photoreceptors, whereas in myopic eyes the image of distant objects is focused in front of the retina before the light reaches the photoreceptors (Morgan et al., 2012). This leads to a blurred image which cannot be brought back into focus by the variable power of the lens (accommodation), thereby causing visual impairment. The larger the distance between the image focusing point and the retina, the more myopic the individual will become.

Several distinct mechanisms can lead to an imbalance between the axial length of the eye and the eye's optical power. Most commonly, myopia arises as a result of greater than normal elongation of the eye (Morgan and Rose, 2019). In some cases, changes in the cornea can lead to the development of myopia, as is the case in keratoconus (Gordon-Shaag et al., 2015), an eye disorder characterised by a progressive thinning of the cornea. In some rare cases, cataract-associated myopia can arise as a result of the development of a small central zone that increases the refractive index (Iribarren, 2015). From a biological perspective, myopia is thought to result from a failure of 'emmetropisation' (see section 4.4.), which is characterised by a light-dependent retina-to-sclera signalling cascade that ties together various mechanisms such as retinal cell physiology, light processing, rod-and-cone bipolar synaptic

neurotransmission, anterior-segment morphology and angiogenesis (Tedja et al., 2018).

Refractive error is defined in terms of spherical equivalent (SE) and is expressed in measurement units called dioptres (D) (Morgan et al., 2012). Although the range of refractive error can vary considerably, in practice, the frequency distribution is typically leptokurtic (Cook and Glasscock, 1951). Refractive error can change substantially during one's lifetime. For example, at birth, refractive error is approximately normally distributed with a mean level in the low hyperopic range. However, by one or two years of age, the distribution narrows and refractive errors become concentrated in the range of +1.00 to +2.00 D (Morgan et al., 2012). This active process leading the change in refractive error distribution is known as 'emmetropisation' (Morgan et al., 2012). At older ages, the rate of change depends on the susceptibility to different risk factors (discussed in section 1.1.3).

Given the continuous nature of refractive error measurement, there has been confusion regarding the best criterion by which to define myopia. Suggestions have been made to group individuals based on myopia aetiology (e.g. hereditary, degenerative) (Saw et al., 1996), time of development (e.g. juvenile-onset, adult-onset) (Grosvenor, 1987), severity (e.g. low, medium, high) and with or without degenerative changes (e.g. physiological, pathological) (Tano, 2002). Thresholds used to define myopia vary considerably across studies (Holden et al., 2016). For example, based on a systematic review of myopia literature (Holden et al., 2016), 1.7% of studies defined high-myopia as -3.00 D or less, and 1.7% as -8.00 D or less. The heterogeneity of classification can have a large impact on meta-analyses and can lead to reduced power, e.g. the inability to identify associations. One recently published paper has outlined a set of standards for defining and classifying myopia (Flitcroft et al., 2019). These authors proposed to classify myopia based on presumed aetiology, age at onset, progression pattern, amount of myopia and structural complications. For each of these categories, a list of descriptive terms was assigned. For example, if epidemiologists describe myopia as permanently progressive, progressive high, stationary or temporarily progressive, then this type of myopia belongs to the classification based on progression pattern. In addition, they proposed four quantitative definitions of

myopia: pre-myopia defined as < -0.50 D and ≤ +0.75 D; myopia ≤ -0.50 D; low myopia ≤ -0.50 D and > -6.00 D and high myopia as ≤ -6.00 D.

As discussed in section 1.1.4., correction measures to halt myopia progression are available. However, if left untreated, a range of complications can arise. Highly myopic individuals are more likely to develop other eye disorders such as cataracts, myopic macular degeneration, glaucoma and chorioretinal abnormalities, which can result in irreversible vision loss (Saw et al., 2005; Morgan and Rose, 2019). The Blue Mountains Eye Study reported a maculopathy prevalence of 0.42% in myopes of less than -5.00 D and 25% for myopes greater than -5.00 D (Vongphanit et al., 2002). In addition to health problems, myopic patients report lower quality of life, especially those with a high degree of myopia (Rose et al., 2000). It is also important to consider the economic burden of myopia, given that half of the population is estimated to be myopic within the next 30 years and 10% are expected to be highly-myopic (Chua and Foster, 2019).

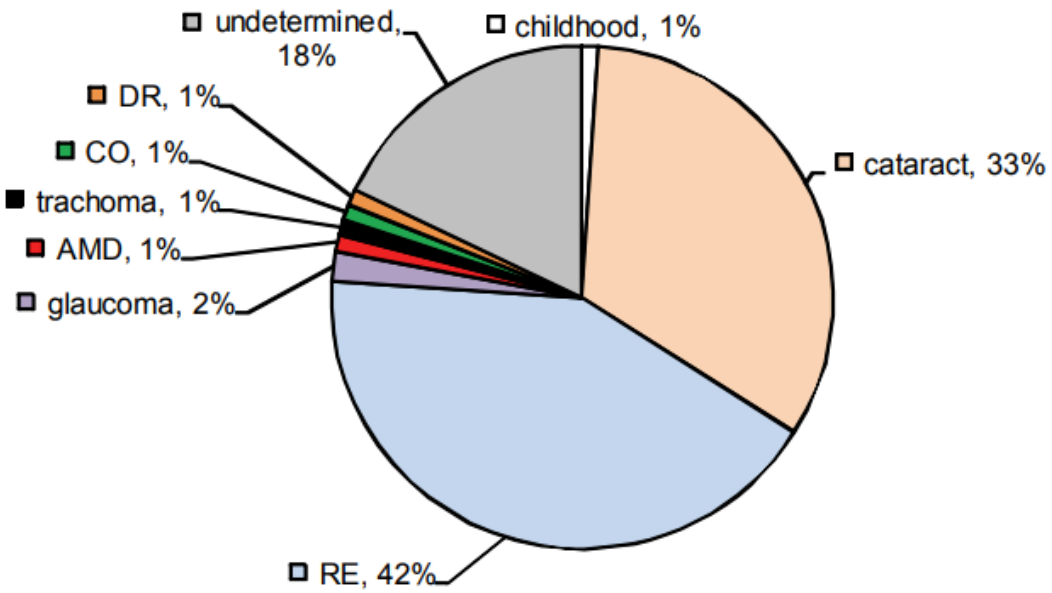


Figure 1.1. Major causes of visual impairment in the year 2010. The figure was copied from http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf. Accessed: 29 October 2019. Abbreviations: RE - refractive error, AMD - age-related macular degeneration, CO - corneal opacity, DR - diabetic retinopathy.

1.1.2. Demographics of myopia prevalence

The prevalence of myopia varies widely across geographical regions (Rudnicka et al., 2016). Figure 1.2. provides an overview of myopia prevalence rates for different parts

of the world. It is worth noting that the calculation of prevalence rates is sensitive to the threshold used to classify myopia. For example, using a random sample of 1,985 individuals from the 1958 British cohort and shifting the threshold towards 0.00 D changed prevalence from 28% to 47% (Cumberland et al., 2018). Nevertheless, a common feature across the countries is that the number of individuals with myopia is expected to increase. It is estimated that by the year 2050, approximately half of the population will become myopic and approximately one billion individuals (10%) will be highly-myopic (Figure 1.3.) (Holden et al., 2016). The region with the highest prevalence rates includes East and Southeast Asia. More specifically, it is the 'school myopia' seen in young adults in developed parts of this region that is of greatest concern. For example, up to 90% of teenagers and young adults in mainland China and 96.5% of 19-year-old men in Seoul are myopic compared to only 10-20% sixty years ago (Dolgin, 2015). Similar rates (70-90%) are observed in Hong-Kong, Taiwan, Singapore, Japan and the Republic of North Korea (Morgan et al., 2018).

A large, high quality meta-analysis carried out by Rudnicka et al. investigated trends in the prevalence of childhood myopia (1 to 18 years) in 42 different countries (Rudnicka et al., 2016). The combined meta-analysis comprised of 374,349 individuals. In total, 74,847 (20%) individuals were considered to be myopic (defined as < -0.50 D). Although the number of individuals with refractive error assessment for meta-analysis was large, sufficient data were only available for whites, East Asians and South Asians. After accounting for confounding factors such as age and environmental setting (i.e. urban, rural or combined), they observed an increase in prevalence only in East Asians (OR 1.23, 95% Credible Interval 1.00 to 1.55). The highest prevalence was observed for 18-year old children from Singapore (92%), about 87% in children from Hong Kong and Taiwan and 72% in children from China. The lowest rates were observed for 5-year old children in rural Mongolia (0.3%). Rates across all studies were the lowest in Mongolians from rural areas irrespective of the age group. The same study did not find convincing evidence to suggest that prevalence rates vary between Europe, USA and Oceania, although a trend of increasing prevalence was noticeable. An interesting observation was made for South Asian migrant children residing in Australia, England or Singapore. At the age of 15 years-old, the South Asian migrant children were five times more likely to be myopic compared to their counterparts from Nepal or India

who had not migrated and were living in South Asia at the time the study was conducted: the myopia prevalence was 40% in migrants vs. 9% in South Asians who had not migrated (Rudnicka et al., 2016).

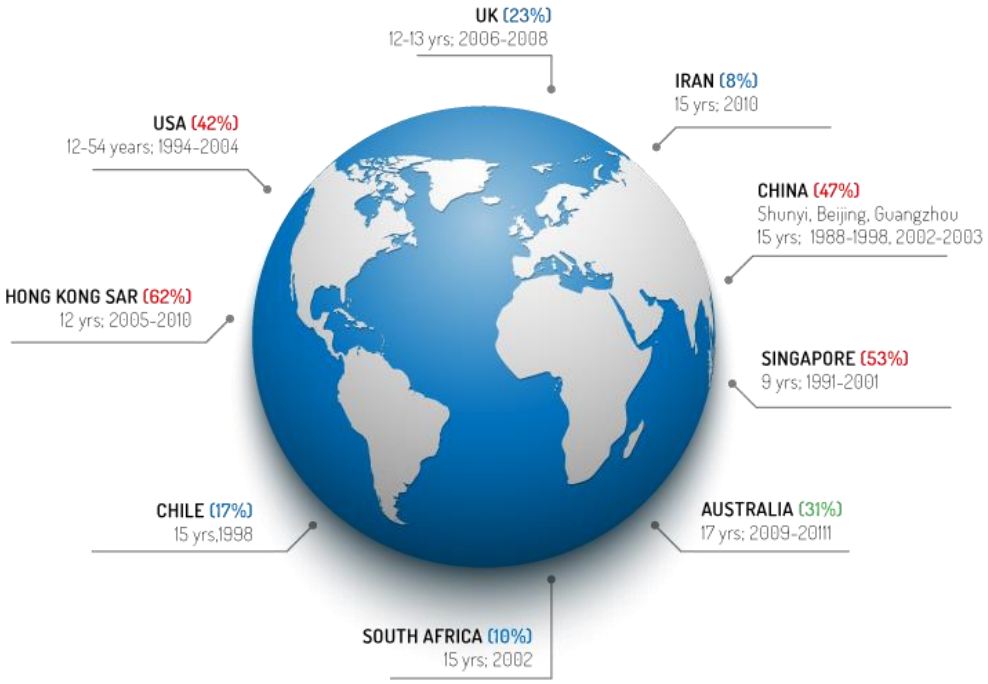


Figure 1.2. Summary of myopia prevalence in different parts of the world. The numbers in brackets represent prevalence rates. The figure was copied from https://www.myopiainstitute.org/prevalence.html. Accessed: 29 October 2019.

In conclusion, the prevalence rates around the globe are not uniform and some regions, such as East Asia, display substantially higher rates. In these regions, the incidence rate of myopia is the highest in teenagers and young adults. Collectively, this information can provide some insight as to what causes myopia. Given that predictions anticipate continued steady rise in new myopia cases, it is important to develop effective management strategies.

1.1.3. Risk factors associated with myopia

Fifty years ago, myopia was believed to be mostly caused by genetic determinants, with only minor environmental influences (Sorsby, 1962). However, during the following years it became apparent that the sharp increase in myopia prevalence in urbanised regions of East and Southeast Asia could not be explained by genetic changes on their own. As a result, there has been a shift from viewing myopia as being predominantly genetically determined to predominantly environmentally controlled.

Now it is accepted that both genetic and environmental risk factors and the interplay between the two are involved in shaping myopia's development and progression.

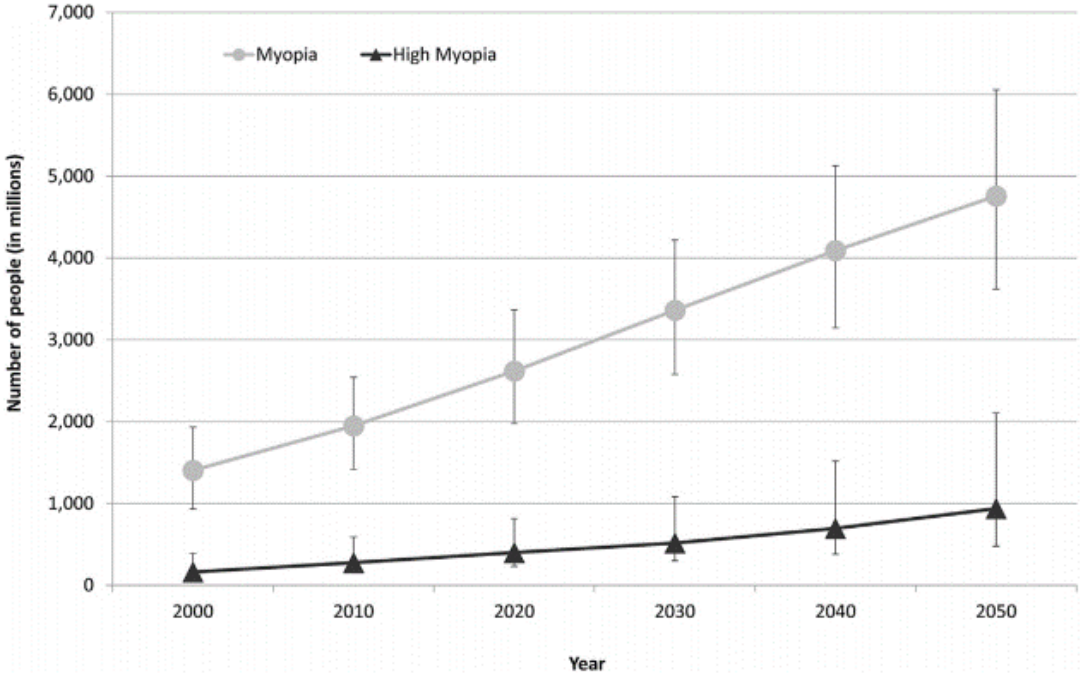


Figure 1.3. Global myopia and high-myopia prevalence rates from year 2000 to 2050. The projection for next 30 years is based on the meta-analysis of 145 studies that reported prevalence rates for myopia or refractive error in different parts of the world. Error bars represent the 95% confidence intervals. The figure is copied from Holden et al., 2016.

One of the most compelling pieces of evidence regarding environmental cues comes from the studies examining educational attainment. Observational studies have consistently shown that individuals who spend more time in education tend to have a more negative refractive error (Tay et al., 1992; Au Eong et al., 1993; Morgan and Rose, 2013; Williams and Hammond, 2014; Nickels et al., 2019). For instance, in Asian populations, individuals aged 40 years or older with a university education are nearly four times more likely to have myopia compared with those with primary or no formal education (Pan et al., 2012). The low prevalence of myopia in African populations has been attributed to low literacy rates given that most children among the countries investigated do not start formal education until the age of 6-8 years (Rudnicka et al., 2016). Several studies have suggested a causal relationship between refractive error and educational attainment (Cuellar-Partida et al., 2016; Mountjoy et al., 2018). For example, Mountjoy et al. performed a Mendelian Randomization (MR) analysis

suggesting that one additional year spent in education was associated with -0.27 D change in refractive error (95% CI -0.37 to -0.17, $p = 4 \times 10^{-8}$) (Mountjoy et al., 2018). No evidence of reverse causality was observed. Frequently, instead of using educational attainment as a risk factor, highly correlated proxies such as socio-economic status (SES), intelligence and near-work activities have been explored to test for association with myopia (Parssinen, 1987; Rosner and Belkin, 1987; Teasdale and Goldschmidt, 1988; Saw et al., 2000). Generally, there is a tendency for higher SES, longer time spent reading and having a higher IQ to be associated with more negative refractive error, however, the effects of these risk factors vary across studies. An additional confounding factor is likely to be the steady rise in the use of modern electronic devices over the past three decades (Wojciechowski, 2011). Because the aforementioned environmental risk factors are correlated, it is difficult to tease apart their individual effects. For example, educational attainment has shown to have a causal effect on refractive error; however, this might be mediated either through the amount of time an individual spends outdoors or the amount of time performing near-work tasks.

As mentioned above, another environmental factor known to influence refractive error is time spent outdoors (Mutti et al., 2002; Jones et al., 2007; Guggenheim et al., 2012). This risk factor has been shown to have a protective effect; thus, a shorter time spent outdoors is associated with a more negative refractive error. The strongest evidence for a protective effect of time outdoors comes from randomised controlled trials, in which the intervention was an extra 20-80 minutes outdoors during the school day during a 1-3 year period (Wu et al., 2013; He et al., 2015). The light stimulation of retinal dopamine, which discourages axial elongation, has been proposed as the mechanism through which time outdoors controls the rate of incident myopia (Read et al., 2014).

Another line of evidence to support an environmental basis for myopia comes from studies of form deprived or lens-induced myopia in animal models (Morgan et al., 2013). In juvenile animals from a wide range of species, covering one or both eyes with a frosted diffuser or a ‘minus’ power spectacle lens induces an accelerated rate of axial elongation leading to myopia. As regards lens-induced myopia, the accelerated rate of

eye growth in response to the imposed defocus continues until the eye reaches an optimal 'compensated' state, where the light is focused on the retina with the spectacle lens in place.

Many other environmental risk factors not mentioned here have been associated with myopia and include birth order (Guggenheim et al., 2013a), number of myopic parents (Kurtz et al., 2007), diet (Lim et al., 2010) and smoking (Iyer et al., 2012).

The first line of evidence of a genetic contribution towards myopia came from twin studies, where a greater concordance of refractive error was observed in monozygotic compared to dizygotic twins (Sorsby, 1962). Indeed, the first ever 'classical' twin study was an investigation of myopia (Liew et al., 2005). Subsequently, estimates for 'broad sense' heritability across populations were consistently found to be high (Guggenheim et al., 2000; Hammond et al., 2001; Lyhne et al., 2001; Dirani et al., 2006). For example, Dirani et al. studied 345 monozygotic and 267 dizygotic twin pairs aged 18-88 years and concluded that heritability of spherical equivalent was 88% in men and 75% in women (Dirani et al., 2006).

It has been noted that myopia accompanies syndromic refractive errors with a known genetic predisposition. These include X-linked and autosomal recessive congenital stationary night blindness (Zeitz et al., 2015), X-linked retinitis pigmentosa (Parmeggiani et al., 2016), X-linked Bornholm eye disease (Schwartz et al., 1990), Marfan syndrome (Verstraeten et al., 2016) and Stickler syndrome (Snead and Yates, 1999). According to the Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov/omim, accessed: 29 October 2019), there are a total of 100-200 rare genetic disorders, where myopia is described as a secondary feature. There is also evidence to suggest that non-syndromic forms of early-onset myopia are under genetic control (Hornbeak and Young, 2009; Morgan and Rose, 2019). For example, Cai et al. provide a list of 25 monogenic forms of myopia or high myopia (categorized as MYP1 - MYP25) that have been discovered using linkage analysis (Cai et al., 2019). More recently, the Consortium for Refractive Error And Myopia (CREAM) has conducted the largest meta-analysis to date (Tedja et al., 2018). The analysis of 160,420 individuals from European and Asian populations identified 161 genomic loci, demonstrated a high degree of genetic overlap between Europeans and Asians, and

highlighted light-induced signalling as a driver mechanism for refractive errors. In addition, the heritability explained by common genetic variants (SNP-heritability) was found to be 0.17-0.21% in Europeans, while it was much lower in the Asian sample (0.05%). Note that this is substantially lower than the typical 80-90% estimated in twin studies. The discrepancy between the two could be due to shared environmental effects, a larger number of variants with smaller effects, rare and structural variants or interactions, and lack of statistical power to detect loci of small effect (Manolio et al., 2009). In general, linkage studies and genome-wide association studies (GWAS) of myopia have pinpointed different set of genes. An explanation for this is that common single nucleotide polymorphisms (SNPs) with small effects are typically identified in GWAS analyses, whereas linkage studies focus on high-myopia and find loci that have large effects. In addition, several whole-exome sequencing experiments for early-onset high-myopia (eoHM) have been conducted (Mordechai et al., 2011; Jiang et al., 2015; Wan et al., 2018). Mutations in genes such as *LRPAP1* were found to influence eoHM (Aldahmesh et al., 2013; Jiang et al., 2015). However, other genes, such as *SCO2* await further validation. Although different study designs used to investigate myopia have discovered different sets of genes, thus making the interpretation of the genetic component more difficult to understand, these various study designs provide a more comprehensive overview of the heterogenous nature of myopia.

The age at which GWAS-identified variants exert their effects is not known. One study examined the age-of-onset correlation between refractive error and 39 GWAS-identified variants in 5,200 children from the Avon Longitudinal Study of Parents and Children (ALSPAC) (Fan et al., 2016). The children had their refractive error measurements taken from ages 7 to 15 years. Many differences between specific SNPs and their effects during childhood were observed. For example, 5 SNPs displayed early-onset effects and showed progressively stronger effects during later childhood, 10 SNPs demonstrated early-onset effects with stable effects through childhood, while 11 SNPs had later onset in childhood.

Relatively few examples of gene-environment (GxE) interactions have been described for myopia. A study by Chen et al. explored the possibility of GxE interactions through two rounds of selective breeding of White Leghorn chicks treated with monocular form

deprivation (Chen et al., 2011a). Although the genes responsible were not identified, susceptibility to this environmentally induced type of myopia found to be heritable (h^2 = 0.50). GxE interaction involving *APLP2* gene in mice has been identified and subsequently replicated in humans (Tkatchenko et al., 2015). In the Tkatchenko et al. study, interaction of the risk variant rs188663068 with age revealed that individuals carrying the risk allele "A" experienced more rapid, age-dependent progression towards myopia. At age 8, children with genotypes *GG* and *GA* were emmetropic (~0 D), whereas when measured at 15 years of age, the same children had refractive error -0.75 D and -1.25 D for the *GG* and *GA* genotypes, respectively. A model with a 3-way SNP × Age × Time spent reading interaction was found to be significant in children who spent >1.0 hrs/day reading. Amongst 'heavy' readers, children with the *GA* genotype were on average one dioptre more myopic compared to children with *GG* genotype at the age of 15 years (-1.4 D vs. -2.5 D). This demonstrates that the refractive error trajectory of children during the 8 - 15 age period depends on the genotype they carry, and that this trajectory can be modified by the amount of reading done per day. Another example of a GxE interaction comes from a study were individuals from several Asian countries were assessed for gene × education interaction (Fan et al., 2014). The authors explored 40 candidate SNPs that have been previously associated with refractive error. They identified three variants that conferred an increased risk of becoming myopic in the 'high education' group (i.e. high schoolers, diploma holders or university graduates) compared to those who completed lower secondary education, primary education or had no formal education. The strongest interaction effect was observed for rs2969180 variant near *SHISA6-DNAH9* genes *(*β_{GxE} = −0.28 ± 0.08 D**).** In the replication analysis, only the *GDJ2* gene showed evidence of interaction with education. A variant nearby another gene, *TJP2*, displayed significant interaction only in Europeans ($p = 6.91 \times 10^{-3}$).

To date, no compelling evidence to support the involvement of gene-gene (GxG) interactions in myopia has been observed.

Overall, the evidence suggests the environmental influences to be the major driver in refractive error variation between populations and different ethnic groups. Environmental risk factors are often confounded (e.g. education and intelligence,

occupation and near-work), making it difficult to unravel individual environmental effects. The prevalence of myopia seems to depend on where children grow up and the environments to which they are exposed, rather than aspects of genetic ancestry (Morgan et al., 2012). While genetics cannot explain the rapid increase in myopia over the last few decades in high prevalence parts of the world, variation within populations is likely to have an important genetic contribution (Wojciechowski, 2011).

1.1.4. Methods used to control myopia development

Currently, there is no strategy to stop myopia progression completely. However, intervention approaches have been proposed that aim to either slow down progression or to delay myopia onset. Given the fact that the environment plays a major role in determining refractive error, it seems reasonable to assume that prevention focused on regulating (i.e. increasing or decreasing) exposure time to an environmental risk factor, would be a good first step to control myopia development and progression. As discussed in the previous section, time spent outside has a protective effect. Hence, strategies encouraging children to spend more time outdoors during school hours would be beneficial (He et al., 2015). For example, the addition of 40 minutes of outdoor activity was found to reduce the incidence rate of myopia over a 3-year period by 10% among 6-year olds in China (He et al., 2015). Attempts to increase the time children spend outdoors, however, have had little success (Verkicharla et al., 2017).

Myopic patients need to wear spectacles or contact lenses to see clearly in the distance. Figure 1.4. lists several types of specialist lens that can be used not only to correct the defocus of myopic children but also to slow their myopia progression. For example, orthokeratology involves wearing rigid gas-permeable lenses overnight to flatten the cornea. The use of these lenses was shown to lead to a consistent reduction in myopia progression of approximately 45% over a two-year period and 30% over five years (Si et al., 2015). Additionally, pharmaceuticals such as the antimuscarinic agents pirenzepine, and atropine have also shown promising results. Use of 0.01% atropine is the most common treatment regimen for the management of myopia in children in a number of Asian countries, such as Singapore, Taiwan and China (Wu et al., 2019).

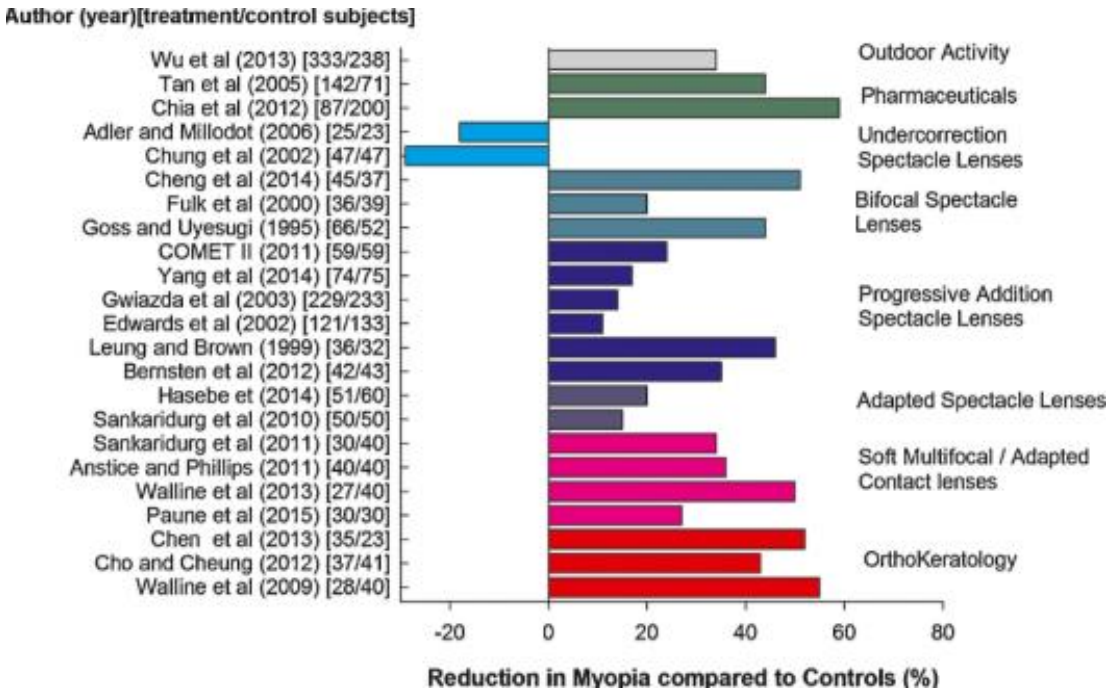


Figure 1.4. Summary of different treatment strategies to reduce myopia onset and progression. This figure was copied from (Wolffsohn et al., 2016).

1.2. Gene-environment interactions

1.2.1. Introduction to gene-environment interaction

Over recent years, advancements in sequencing technologies have made genotyping more accessible and cost-effective, allowing researchers to study entire genomes in a hypothesis-generating fashion. The rapid adoption of genome-wide association studies (GWAS) superseded previous methodologies such as linkage and candidate gene studies of complex traits and led to several, now well established and robust conclusions. Some of the key findings include a high degree of polygenicity for common complex traits, very small effect sizes for common genetic variants and a high degree of genetic overlap across different traits (Visscher et al., 2017). Although there is undeniable evidence that large-scale genomic studies have provided valuable insight regarding the genetic architecture of common traits (Visscher et al., 2012; Visscher et al., 2017), many challenges remain to be resolved (Manolio et al., 2009; Parker et al., 2009; Platt et al., 2010; Korte and Farlow, 2013; Tam et al., 2019). For example, some have argued that the majority of the currently used tools are too simplistic and do not accurately capture the relationship between the phenotype and the genotype due to the failure to account for the effect of biologically plausible mechanisms such as

epigenetics or post-transcriptional modification (Nelson et al., 2013). Consequently, a step in this direction has been made by utilising -omics and next-generation sequencing data (Ritchie et al., 2015; Ritchie et al., 2017). Since it is well accepted that complex traits are influenced by a multitude of genetic and environmental factors (Dick et al., 2015), an intuitive approach is to build statistical models that estimate the joint effect of genetic variants and environmental exposures.

From a statistical point of view, gene-environment (GxE) interaction is characterized by the lesser or greater joint effect of environmental exposure and genetic susceptibility compared to the expected effect of each risk factor if they were studied independently. In this context, the environment can refer to a wide range of exposures, including biological (e.g. virus and disease vectors), chemical (e.g. pesticides and persistent organic pollutants), physical (e.g. radiation and temperature), behavioural (e.g. late age at first pregnancy) or some kind of life event (e.g. injury and job loss) (Ottman, 1996).

Some of the best-characterized evidence of GxE interaction comes from studies of Mendelian disorders, where the phenotype is influenced by the action of one or a few genes that have a strong effect on the phenotype. For example, phenylketonuria (PKU) is a metabolic disorder that arises as a consequence of a defect in the enzyme that metabolises the amino acid phenylalanine (Pietz et al., 1999). Affected children display intellectual and developmental disability. However, the condition can be reversed by monitoring diet and limiting the intake of phenylalanine. Ritz et al. provide a comprehensive list of disorders that are amenable and preventable by the environmental intervention (Ritz et al., 2017). In summary, the authors suggest that the most compelling evidence of the successful discovery of GxE interaction to date comes from the studies of proteins that have enzymatic or metabolic functions. However, evidence of GxE interaction in common complex traits is scarce (Risch et al., 2009; Karg et al., 2011; Dick et al., 2015).

When considering GxE interaction, a question regarding the direction of its effect arises (Dick et al., 2015). There are several points to consider. First, it might be of interest to understand whether genetic susceptibility predisposes to enhanced effects of the environmental exposure or whether environmental exposure leads to greater changes

in the phenotype, for example, through the altered gene expression (Gauderman et al., 2017). This is important if the goal is to understand the biological mechanism by which GxE interaction exerts an effect on the phenotype. Ottman has described five biologically plausible mechanisms (Ottman, 1990, 1996). For example, model A in Figure 1.5. represents a situation where the genotype of an individual modifies the risk of environmental exposure. An example of this type of mode of action is PKU, as mentioned above. In model B, the effect of genotype is exerted in individuals who are exposed to an environmental risk factor. Certain statistical tests, such as case-only analysis (discussed below), have been tailored to detect this specific type of GxE interaction. An example of this type of interaction is xeroderma pigmentosum (Ottman, 1996). Affected individuals lack the enzyme that repairs DNA damage after exposure to ultra-violet radiation. In model C, the effect that the genotype has on a trait is exacerbated by environmental exposure. In this case, individuals who carry a low-risk genotype do not show a change in the phenotype upon being exposed. An example of this mode of action can be seen in individuals with porphyria variegata, an autosomal dominant disorder affecting the skin. Exposure to barbiturates can lead to paralysis and even death. Model D represents a situation, where both genetic and environmental risk factors are required to observe an effect (i.e. no effect is seen when only one risk factor is present). A classic example is glucose-6-phosphate dehydrogenase (G6PD) deficiency. Individuals with this condition are asymptomatic until they eat fava beans, which leads to haemolytic anaemia. In contrast, individuals without G6PD deficiency do not display adverse effects upon eating fava beans. Model E represents a scenario where both genetic and environmental risk factors have a certain effect on the phenotype. However, the effect of the two is greater or lesser only if both factors are present. An example would be chronic obstructive pulmonary disease (COPD). The risk of developing COPD is increased in non-smokers with hereditary α1-antitrypsin deficiency and smokers without α-1-antitrypsin deficiency.

In general, the actual biological mechanisms underlying GxE interactions can be explored by functional studies. Since the underlying relationship between the interaction effect and the trait of interest is rarely known, several types of statistical interpretation have been described (Figure 1.6.). For example, a quantitative interaction (Figure 1.6. left panel) is described by a change in effect magnitude of one exposure in the presence of another (Hutter et al., 2013). This type of interaction might be of interest if the goal is to identify a group of individuals at highest risk and the majority of methods discussed in section 1.2.3. are developed to detect this type of interaction. On the other hand, a qualitative interaction (also known as crossover interaction) (Figure 1.6. right panel) is defined by the reversal of one exposure's effect by the other (Hutter et al., 2013) and a method to detect this type of interaction using a likelihood ratio test has been proposed (Gail and Simon, 1985).

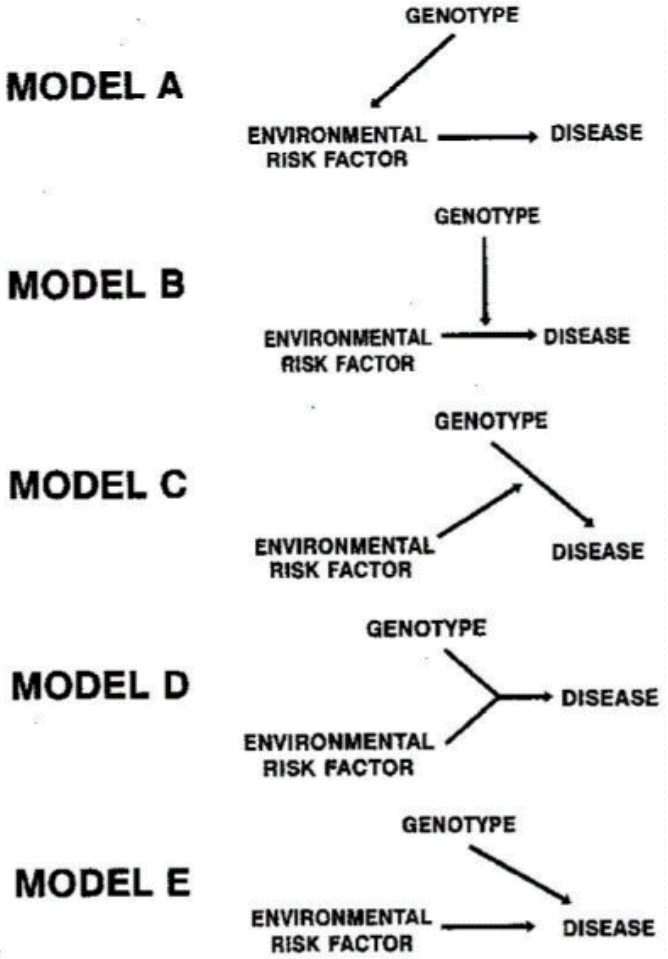


Figure 1.5. Summary of five different biologically plausible GxE interaction mechanisms by which genetic and environmental risk factors can produce changes in the phenotype (copied from Ottman, 1996).

Dempfle et al. provide a detailed discussion regarding different types of interactions (Dempfle et al., 2008). Typically, environmental variables are considered as binary factors. Taking, for example, university education, we could represent a group of individuals who have a university degree as $E = 1$ or $E = 0$. Although the difference in coding does not affect the significance of the interaction effect, the magnitude and direction of effect can change, and this arbitrary choice of coding will determine the interpretation of interaction coefficients (Gauderman et al., 2017). To avoid the confusion due to arbitrary classification, variable coding should always be stated explicitly (VanderWeele and Knol, 2014).

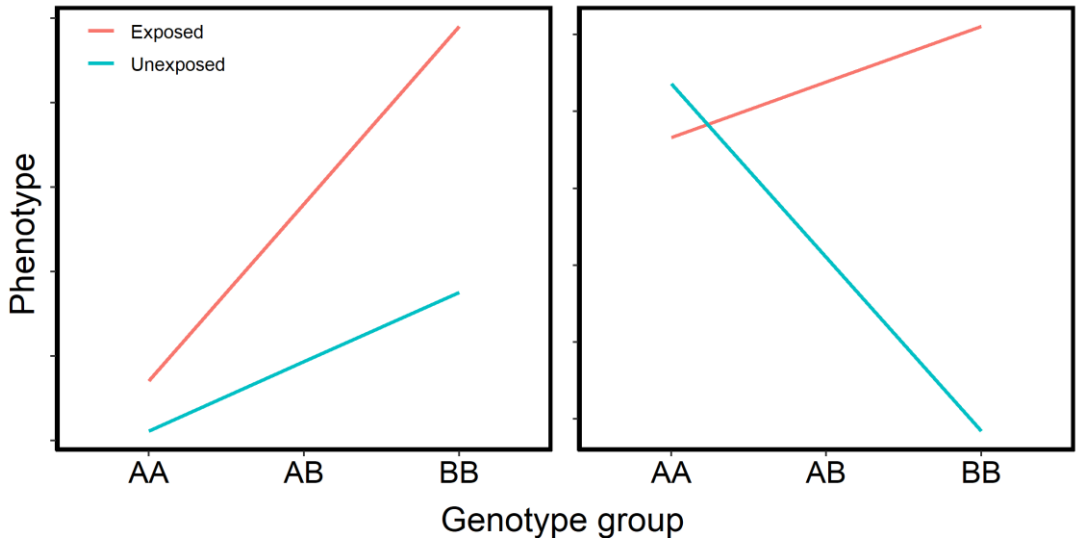


Figure 1.6. Statistical interpretation of plausible GxE interaction effects on the phenotype. An example is shown for GxE interaction involving a biallelic genotype and a binary environmental exposure. The left panel shows concordant direction of GxE interaction (i.e. the effect of interaction increases with additional copy of *B* allele in both exposed and unexposed individuals) and divergent effect magnitude (i.e. the risk is greater in individuals that have a *BB* genotype and are exposed to an environmental risk factor compared to unexposed and *BB* genotype carriers). In the right panel, I show a scenario where GxE interaction has an opposite effect in *BB* genotype carriers.

The focus of this thesis revolves around the statistical detection of interaction effects. In other words, I utilise available bioinformatics approaches to identify genetic variants that show a departure from additivity and do not take into account the mechanism by which GxE interaction influences refractive error. Where

appropriate, I use functional annotation tools to provide biological interpretation for newly-discovered associations.

1.2.2. Motivation for studying gene-environment interactions

There are several reasons to consider GxE interactions. First, it is essential to understand the interplay between genetic and environmental risk factors because they operate together to influence changes in the phenotype. For example, studying GxE interaction could provide novel insight into the biology of disease beyond that offered by studying marginal SNP effects (McAllister et al., 2017). The greatest benefit would be obtained in the case where the genetic variants that are involved in GxE interaction are distinct from those that show marginal association with the trait of interest. This is an example of so called *pure* interaction (Hutter et al., 2013). Additionally, novel environmental exposures that mediate the changes in the phenotype could be identified (Patel and Ioannidis, 2014; Patel and Manrai, 2015). Collectively, such information could lead to improved treatment and prognosis of patients. Second, GxE interaction could be used to identify a high-risk subgroup of individuals that is at greater risk of developing the disease (Figure 1.6. left panel). This could arise as a consequence of increased genotype-specific or increased environmental-specific risk. Such information can influence decisions regarding treatment strategy. For example, suppose a limited dose of medication is available. We would prefer to administer drugs to individuals who are at the greatest risk (VanderWeele and Knol, 2014). In the pharmacogenomics setting, GxE interaction information can guide the decision behind the development of tailored treatment for each patient (Ritz et al., 2017). In the case of continuous traits, such as refractive error, the rate of change in the phenotype could be exacerbated by GxE interaction. For example, in a longitudinal study, where relevant information is collected at different time points, we could observe a faster rate of disease progression that is caused by innate genetic susceptibility or by more prolonged exposure to adverse a specific environmental condition. In any case, this would provide motivation for developing early treatment strategies before the manifestation of serious complications. Third, current studies fail to provide accurate phenotype prediction in independent samples. It has been argued that the inclusion of interactions leads to better prediction of extreme phenotypes in model organisms (Forsberg et al., 2017). However, whether this is also true in humans remains to be

determined. Lastly, Mendelian Randomization (MR) has gained popularity in recent years (Davies et al., 2018; Grover et al., 2018; Bowden and Holmes, 2019; Lor et al., 2019) and become a standard method of choice to strengthen causal inference. However, MR is prone to biases that can arise by virtue of several assumptions that need to be satisfied in order to obtain valid measures of causality (Hemani et al., 2018). For example, SNPs must associate with the phenotype through a particular exposure but not by a direct association with the trait or via an unmeasured confounding factor. It has been suggested that inclusion of GxE interaction could detect and correct for pleiotropic bias (i.e. one gene many functions) in MR studies (Spiller et al., 2019).

1.2.3. Methods for detecting gene-environment interactions

Recent years have seen a large increase in the number of distinct analytical methodologies to study GxE interaction in large-scale studies. This section provides an overview of the available modelling strategies. I begin by describing simple regression-based methods and finish with a discussion of more complicated approaches.

The simplest method includes a standard logistic or linear regression method that takes into account a GxE interaction term. In this context, interaction is assessed using the multiplicative risk scale (see Chapter 3 for more details). An exhaustive search for interaction is typically performed for all genotyped and imputed genetic variants. These models can be extended to estimate relative risks in cohorts (Breslow et al., 1987), or hazard rate ratios for time-to-disease data (Kalbfleisch and Prentice, 2011). In addition, environmental exposure does not have to be restricted to binary classification and can be specified as continuous, ordinal or categorical (Tchetgen Tchetgen and Kraft, 2011). A common feature for all these methods is that the null hypothesis being tested is H_0: $\beta_{GxE} = 0$. A two degree of freedom (2-df) approach could be used to test the joint effect of a SNP's main and interaction effect (H_0: $\beta_G = \beta_{GxE} = 0$) (Kraft et al., 2007; Aschard et al., 2010; Manning et al., 2011; Dai et al., 2012b; Hancock et al., 2012). Piegorsch et al. suggested using a case-only analysis when both *G* and *E* are binary (Piegorsch et al., 1994), which was shown to be a more powerful method compared to logistic regression. However, this method has a few limitations. First, a case-only design cannot estimate the main effects of β_G and β_E (Gauderman et al., 2017). To resolve this issue, methods that use controls, instead of cases, to estimate

main effects have been developed (Umbach and Weinberg, 1997; Chatterjee and Carroll, 2005). Second, the estimates for β_{GxE} will be biased if there is a gene-environment correlation (rGE) (Albert et al., 2001). To partially address this issue, an empirical Bayes approach (Mukherjee and Chatterjee, 2008; Chen et al., 2009a) and Bayesian model averaging (Li and Conti, 2009) have been proposed. These methods provide more flexible modelling that exploit properties of both case-only and case-control analyses. To further increase the statistical power of GxE interaction tests, several two-step methods have been proposed (Murcray et al., 2008; Gauderman et al., 2010; Pare et al., 2010; Murcray et al., 2011; Dai et al., 2012a; Hsu et al., 2012; Gauderman et al., 2013; Zhang et al., 2016). Generally, two-step methods involve the following two stages. First, an exhaustive interaction search is performed using one of the previously described methods for all *M* available genetic variants. Next, only *m* SNPs that pass a given threshold (e.g. $p < 0.05$) are selected, and GxE interaction discovery is carried out in this smaller subset of variants that passed the initial screening stage. These methods are statistically more powerful because the significance threshold in stage two is adjusted by the fewer number of SNPs (i.e. $m < M$). In essence, two-step methods resemble candidate gene approach, where a few SNPs, within or nearby a gene, are selected for testing based on the previous evidence of their association with the phenotype. Alternatively, a prioritization strategy for the second stage could include a selection of variants that are known to be associated with environmental factors or variants located in the functional regions of the genome (Ritchie et al., 2017). A key assumption for any two-step method is that each stage should be independent. In other words, the calculation of test statistic in the second stage should not be influenced by the results from stage one. One complication arises when selecting a threshold used to prioritize variants for the second stage. If this threshold is too stringent, pure interactions will be missed. If the threshold is closer to nominal significance, statistical power will be reduced.

A different category of GxE interaction tests includes methods that use multiple SNPs within a biologically defined set (e.g. gene, pathway) (Tzeng et al., 2011; Jiao et al., 2013; Lin et al., 2013; Jiao et al., 2015; Zhao et al., 2015; Lin et al., 2016; Liu et al., 2016; Su et al., 2016). In other words, a single interaction test between environmental exposure and multiple SNPs is performed. The motivation behind this strategy is that

when SNPs display only weak interaction effects individually, the signal could be much stronger when multiple SNPs are considered in aggregate. Set-based methods can be broadly classified into three categories: variance component tests, burden tests, and their combination. Methods based on variance components are more powerful when the magnitude and direction of effect across SNPs are heterogeneous (Lin et al., 2013), while burden tests perform better when genetic variants in a set display the same effect direction and are causal (Jiao et al., 2013; Liu et al., 2016). Most of the time it is unknown which of the two hypotheses regarding the direction of effect is true. Therefore, a hybrid method that combines variance component and burden tests that may perform well under a range of scenarios has been proposed (Jiao et al., 2015; Lin et al., 2016; Su et al., 2016). Another method that utilizes a set of SNPs to find evidence of GxE is based on polygenic risk scores (PRS) (Garcia-Closas et al., 2013; Garcia-Closas et al., 2014; Lin et al., 2018, 2019). Although similar in spirit (i.e. aggregate effect across multiple SNPs being tested), there are key differences that separate it from other set-based methods. For example, studying SNPs within a gene might yield a more insightful view regarding the biological function of the gene, and it can incorporate LD information accounting for untyped SNPs.

The methods discussed so far assume a linear relationship between the phenotype and risk factors. However, this assumption might not hold in cases where the risk of both genetic and environmental factors varies with time. For example, the incidence of myopia between ages 10 and 15 years was shown to depend on the time spent outdoors across the 3 to 9 years age range (Shah et al., 2017). For this reason, non-linear tests to study GxE interaction have been proposed (Ma et al., 2011; Wu and Cui, 2013; Wu et al., 2014; Ma and Xu, 2015; Sa et al., 2016). Both single variant and gene-based tests under this framework are available. Due to their design, these tests can only accommodate continuously measured environmental factors.

Finally, a number of non-parametric methods have been proposed. Collectively, these approaches can be described as data mining analyses. The advantage of these methods lies in their relaxed assumption regarding the genetic model or distribution of effect sizes, which leads to improved computational efficiency. Typically, they aim to find a combination of genetic or environmental factors that best describe the phenotype.

These methods are usually designed to avoid multiple testing; however, they require follow-up analysis to obtain the optimal model and a measure of statistical significance (Steen, 2012). One popular method is the multifactor dimensionality reduction algorithm (Ritchie et al., 2001; Mei et al., 2005; Chen et al., 2014; Yu et al., 2016; Yang et al., 2019), which is discussed in more detail in Chapter 6. Other methods are based on random forests (Schwarz et al., 2010; Wolf et al., 2010) and pattern recognition (Zhang et al., 2014).

Given the large number of tests developed to investigate GxE interaction, one might wonder which method is the most appropriate. It is unlikely that any given test will perform well under different scenarios and its efficiency will be determined by the underlying causal mechanism (Hsu et al., 2012; Gauderman et al., 2013). Factors such as study design attributes and the population being studied could also influence the decision of which test to use (McAllister et al., 2017). To avoid reporting spurious findings, caution should be taken when applying multiple methods. For example, replication analysis using an independent sample should be used to confirm the validity of newly identified GxE interactions. It should be stressed that the list of GxE interaction methods provided in this section is by no means exhaustive. The number of available tests is likely to increase in the future, to accommodate the improved understanding of biological mechanisms shaping complex phenotypes.

1.2.4. Current challenges related to gene-environment interactions

Undoubtedly, the biggest challenge and the key explanation for poor detection and replication of GxE interaction is low statistical power. There are many factors that determine the power of the study. First, GxE effect sizes are hypothesized to be much smaller compared to marginal SNP effects (Bureau et al., 2015; Aschard, 2016). Thus, larger sample sizes are required to detect an interaction effect compared to a marginal effect. Environmental exposure variability may be another explanation for the inability to identify GxE interaction (Kraft and Aschard, 2015; Stenzel et al., 2015; Gauderman et al., 2017). This is possibly one of the reasons why studies of model organisms, such as mouse or *Drosophila*, have been more successful at identifying interactions compared to human observational studies (Aschard et al., 2012b; Mackay, 2014). Another explanation for the relative paucity of compelling GxE interaction in human

genetic epidemiology could be due to exposure misclassification (Stenzel et al., 2015). Measured exposures are rarely etiologically relevant to the trait under consideration as these are most often replaced by inexpensive proxies (Kraft and Aschard, 2015). For example, in the studies of dietary measures, self-reported questionnaires are typically assayed instead of the analysis of a dietary record. Frequently, there is a low correlation between the two (Gauderman et al., 2017). The effect of environmental exposures might vary considerably throughout human lifespan and might be spatially or culturally dependent (Gauderman et al., 2017). It is also apparent that complex traits are influenced by multiple environmental factors (e.g. refractive error) and so it might be difficult to dissect an individual effect of any given exposure in the case where there is a correlation between exposures (Patel and Ioannidis, 2014; Patel and Manrai, 2015). Methodological challenges remain regarding the best approach to characterize the joint effect of multiple genetic and environmental risk factors. Some studies have used a linear model where a term for "Gene x Environment x Environment" interaction was examined (Grabe et al., 2012; Fan et al., 2016). For example, Grabe et al. suggested the presence of a synergistic effect between childhood abuse and adult traumatic events using this statistical framework (Grabe et al., 2012). Moore et al. have developed a mixed-model approach that simultaneously includes many environmental exposures and ranks them according to their relevance as regards GxE interaction (Moore et al., 2019). Enthoven created environmental risk score (ERS) using outdoor activity, reading time and books read per week and tested the effect of ERS interaction with polygenic risk score for myopia (Enthoven et al., 2019). In practice, we do not know how many genetic variants and environmental exposures influence changes in the phenotype, nor do we know the underlying relationship between these factors (linear vs non-linear). The aforementioned examples illustrate a multitude of ways GxE interaction could be explored in the future. Measurement error also affects the statistical methods used to detect GxE interaction. As a consequence of linkage disequilibrium (LD), the causal locus is rarely measured directly, and therefore we rely on marker loci that are correlated (Gauderman et al., 2017). It is also likely that model misspecification contributes to the failure to detect evidence of GxE interaction (Sitlani et al., 2015; Sun et al., 2018; Ueki et al., 2019). As discussed in the previous section, a wide range of methods to test for GxE interaction exist, each of which is based on different assumptions regarding the biology of the trait. If the method of choice used to discover

GxE interaction does not reflect the underlying genetic architecture, the results might be enriched for false-positive findings that do not replicate in independent samples. On the other hand, standard approaches might not detect evidence of interactions with opposite effects (Murcray et al., 2008), leading to false-negative findings. Additional considerations regarding statistical power should be made when replication and meta-analysis using GxE interaction is performed (McAllister et al., 2017). For example, the power in the independent studies will be influenced by differences in the distribution of environmental risk factors. In some cases, where a rare disease is being studied, or environmental exposure is unique to a specific population, a replication sample might not be available (Mechanic et al., 2012). Further complications might arise when gene- and pathway-based methods are used to study GxE interaction due to the heterogeneous nature of risk factors (Ritchie et al., 2017). All of these aforementioned considerations suggest that an increase in statistical power might be gained by cost-effective study designs rather than simply by increasing sample size (Wong et al., 2003).

Keller argues that it is important to consider interactions not only between genetic and environmental risk factors (GxE) but also between genetic factors and confounding covariates (GxC) and between an environmental factor and confounding covariates (ExC) (Keller, 2014). However, including a statistical term for every possible SNP-environment-covariate interaction is statistically burdensome due to a large number of degrees of freedom used. In other words, a reduction in statistical power will be observed in situations where there are no GxC and ExC interaction effects. This, coupled with the fact that current sample sizes are rarely large enough to detect interaction effects is why this suggestion is often ignored in practice when performing a genome-wide environment interaction study (GWEIS). In addition, sometimes it is not obvious if adjustment for certain factors is required. For example, in the study of gene-diet interaction, adjusting for body mass index (BMI) and including SNP x BMI, and diet x BMI may be appropriate but might result in reduced ability to detect SNP x diet interaction (Gauderman et al., 2017).

Compared to standard GWAS analyses of marginal SNP effects, GxE interaction studies are likely to have additional confounders. One example is gene-environment

correlation (rGE) (Avinun, 2019). This phenomenon refers to the situation where environmental exposures depend on an individual's genetic makeup (Jaffee and Price, 2007; Kendler and Baker, 2007). The evidence for potential bias due to rGE is most evident in the case-only type of analysis where independence between the two factors needs to be guaranteed. There are three types of rGE: passive, evocative and active. An example of passive rGE is when the environment of a child depends on the genetic characteristics of the parents. For example, a meta-analysis of educational attainment focused on interpreting the effect of non-transmitted SNPs in children and suggested that this source of information could explain some of the variation in the child's phenotype (Kong et al., 2018). Evocative rGE arises when heritable characteristics evoke an environmental response. An example of this was demonstrated in the study of interpersonal behaviour between the mother and the child (Klahr et al., 2013). The investigators concluded that maternal control, but not maternal warmth, was influenced by rGE. On the other hand, heritable characteristics that influence the selection of a particular environment (e.g. friends, activities) leads to an active rGE (Goldman et al., 2013).

In summary, analysis of GxE interaction requires additional care when considering the joint effect of genetic and environmental risk factors. Some complications described in this Chapter and discussed further in later Chapters, e.g. the scale on which interaction is measured (VanderWeele and Knol, 2014), make the search for gene-environment interactions more difficult compared to marginal SNP testing in a standard GWAS setting.

1.3. Gene-gene interactions

1.3.1. Introduction to gene-gene interactions

Many arguments raised in the previous section can be easily extrapolated to gene-gene interactions (GxG). Therefore, in this section, I aim to discuss additional ideas that are unique to interactions across genes.

GxG interaction refers to the situation, where the gene effect is modified by the effect of another gene. Similar to GxE interaction, a biological and statistical interpretation of this phenomenon has been described. Biological interaction, also known as

epistasis, has been characterised by the physical suppression of gene expression at one locus by a gene at another locus (Bateson, 1907). A classic example is eye colour determination in *Drosophila* (Tyler et al., 2009). The mutations in *scarlet*, *brown*, and *white* genes lead to different eye pigmentation that cannot be explained by the effect of each individual gene. On the other hand, statistical interaction was coined by Fisher to account for potential departure from his multi-locus additive genetic model (Fisher, 1919). The two were shown not to be equivalent (Norton and Pearson, 1976; Moore and Williams, 2005). In other words, the presence of biological GxG interaction does not necessarily imply the presence of statistical GxG interaction and vice versa (Wei et al., 2014). One explanation for this is that a wide range of statistical genetic architectures determining gene regulatory network is possible in theory (i.e. there are multiple mathematical models which describe how two or more genes could interact in gene regulatory network) (Neuman et al., 1992; Gjuvsland et al., 2007; Wirapati et al., 2011). For example, Hallgrímsdóttir and Yuster explored 69 unique models that could summarize two-locus epistasis (Hallgrímsdóttir and Yuster, 2008). However, which of these models are real in practice is yet to be fully understood. Another explanation is that some types of statistical GxG interaction are dependent on the measurement scale, while biological GxG is not (Cordell et al., 2001). It is crucial to realise that biological GxG interaction is best understood by physical interactions at the cellular level in an individual whereas statistical GxG interaction relates to genotype-phenotype relationship on a population scale (Moore and Williams, 2009).

It has been suggested that GxG interactions are common in human diseases (Moore, 2003). Several arguments to support this idea have been proposed. First, the ubiquity of biomolecular interactions observed in biochemical and metabolic systems indicates the importance of biological GxG interaction (Moore and Williams, 2009). Second, poor replication rate across independent samples could be potentially attributed to GxG interaction (Moore and Williams, 2009). Third, GxG interaction is commonly found in model organisms (e.g. *Drosophila* example above) when properly investigated (Moore and Williams, 2009; Hemani et al., 2014; Mackay, 2014). However, it remains a major challenge to have a full understanding of GxG interaction networks in humans. For example, it is not clear if GxG interaction is equally ubiquitous across common complex traits and empirical evidence of transferability of gene networks across populations is

lacking (Wei et al., 2014). An evolutionary interpretation has been put forward to explain the ubiquitous nature of GxG interaction (Waddington, 1942). This perspective is guided by the ability of natural selection to evolve systems to a robust level that is resistant to most genetic and environmental perturbations (Gibson, 2009). Therefore, destabilization leading to the changes in the phenotype might be a consequence of the accumulation of multiple mutations in different parts of a gene network (Moore and Williams, 2009). This might explain why analysis of SNPs explain little of the disease risk and support the idea that such variation is an emerging property of interaction networks (Huang and Mackay, 2016).

1.3.2. Motivation for studying gene-gene interactions

The advantages of considering GxG interactions are largely the same as for GxE interaction. For example, having a well-characterized network of GxG interactions could be useful in personalized medicine (Moore and Williams, 2009). Suppose a pair of SNPs exert a strong effect on the phenotype so that individuals who inherit an effect allele at each locus has a substantially greater risk of developing disease or rate of change in the case of a continuous phenotype. This knowledge could influence our decision regarding diagnosis and disease management. However, public treatment strategies will have to be different compared to those implemented based on the knowledge gained from studying GxE interactions. Similarly, the inclusion of GxG interaction in the statistical model can aid with the discovery of novel variants that are either rare or have negligible marginal SNP effects and empirical evidence to support this idea has been provided (Wei et al., 2014).

Analogous to GxE interaction, GxG interaction effects are anticipated to be of several magnitudes smaller compared to additive SNP effects. Therefore, in order to take full advantage of GxG interactions, an extensive characterization of genome-wide interactive networks is required. This will provide a more insightful view of the genetic homeostasis and thus would solve some of the outstanding questions concerning the genetic architecture of complex traits (Waddington, 1942; Gibson, 2009). For example, classical quantitative genetics theory suggests that epistatic variance (V_i) is relatively small as GxG interactions contribute towards additive genetic variance (Hill et al., 2008), while the more recent analysis, using alternative parameterization scheme,

proposes that V_i could be much greater (Huang and Mackay, 2016). Although the disparity between the two is purely theoretical, this information could potentially have direct implications for phenotypic prediction (Forsberg et al., 2017), refined breeding programs tailored to maximise response to natural selection (Mackay, 2014) and could be used to highlight mechanisms responsible for speciation (Ayala and Fitch, 1997).

1.3.3. Methods for detecting gene-gene interactions

Regression-based methods described in section 1.2.3. can be extended to find GxG interactions and the software such as *PLINK* (Purcell et al., 2007) and *CASSI* (URL: http://www.staff.ncl.ac.uk/richard.howey/cassi/, accessed: 29 October 2019) provide a good starting point for performing an exhaustive or two degrees of freedom tests. However, as discussed in the following section, if applied on a genome-wide scale, these methods lead to high computational and statistical burdens. Therefore, most of the new methods that are being developed rely on certain shortcuts that make the search for statistical GxG interactions more manageable. In general, these methods use shortcuts to either filter out redundant SNPs before the analysis of GxG interaction (Zhang et al., 2008; Greene et al., 2009; Dorani and Hu, 2018; El-Rashidy, 2019) or utilise efficient algorithms that are based on different classification rules (Zhang and Liu, 2007; Tang et al., 2009; Schwarz et al., 2010; Zhang et al., 2011b; Knights et al., 2013). Wei et al. provide a comprehensive summary of different methods that can be used to detect GxG interactions (Wei et al., 2014).

1.3.4. Current challenges related to gene-gene interactions

There are many challenges associated with reliably detecting and confirming GxG interactions on a genome-wide scale. Nowadays, a typical GWAS analysis uses in excess of one million genotyped and imputed genetic variants. To test for SNP x SNP interaction means to test every possible combination between a pair of SNPs. For a study that assays one million variants, this requires a total of 5 x 10^{11} comparisons. In general, the number of resulting combinations between a pair of SNPs can be calculated using C_k^n, where *n* is the number of genetic variants used in the study and *k* is the degree of interaction (e.g. *k* = 2 for SNP x SNP interaction, *k* = 3 for SNP x SNP x SNP interaction etc.). Given that sequencing costs will decrease in the future (Levy and Myers, 2016), this problem will only become more difficult to solve. First, the large

number of possible GxG interaction combinations requires stringent adjustment for multiple comparisons. It has been proposed that the threshold defining significant GxG interaction should be set to 1.00×10^{-13} (Wei et al., 2014). Therefore, the power to detect GxG interaction association with the trait is low *a priori*.

Second, the exhaustive search for interactions is computationally time-consuming. Methods that select a subset of SNPs or rely on strategies that use efficient algorithms and parallelize GxG interaction test into multiple batch jobs could reduce the computational time (Moore and Williams, 2002; Moore and Ritchie, 2004). However, even in such cases, the anticipated increase in the available genetic data can make this problem intractable. This is evident if the higher-order interactions (i.e. $k > 2$) are considered. If complex traits are influenced by large genetic networks, as seems to be the case, higher-order interactions will have to be considered in order to gain a meaningful representation of genetic interactome. The issue of computational burden can be further exacerbated if non-parametric methods, such as multifactor-dimensionality reduction, are used, that rely on expensive cross-validation and permutation procedures to evaluate the significance of GxG interaction (Ritchie et al., 2001).

Another layer of complexity stems from the fact that in the case of two loci, four different interaction terms (additive-by-additive (AxA), additive-by-dominant (AxD), dominant-by-additive (DxA) and dominant-by-dominant(DxD)) are possible (Wei et al., 2014). Similarly, this can be extended to higher-order interactions. Currently, this concern remains theoretical. Typically, studies in humans do not distinguish between different modes of inheritance and explore additive-by-additive coding. In situations where dense genotyping, or high-quality imputed genotypes are not available, the selection of coding has a direct consequence on the amount of variance explained (Wei et al., 2014). In other words, the proportion of variance depends on LD (r) and will on average decrease by r^4 for AxA, r^6 for AxD and r^8 for DxD (Wei et al., 2014). As a consequence of increased dependence on LD, replication of epistasis is expected to be substantially lower than that for marginal SNP effects (Hemani et al., 2014). Furthermore, from studies of marginal SNP effects, it is known that statistical power depends on the allele frequency (Myles et al., 2009). Therefore, when testing for GxG

interaction, the frequency at each locus will determine the success of identifying GxG interaction.

In summary, the detection of GxG interaction is challenging due to large computational and statistical demands. Incredibly large samples are required to detect GxG interaction effects that are equivalent to those observed for marginal SNP effects, and dense sequence data are needed in order to address the issues related to model complexity and replication. The vast number of available methods to detect GxG interactions aim to discover a combination of SNPs that best describe the phenotype but do not provide biological interpretation for this association. As a result, studies of GxG interactions will have to supplement the findings with biologically derived information from publicly curated databases.

1.4. Description of UK Biobank data

In this thesis, I utilize information provided by the UK Biobank project, which recruited approximately 500,000 UK based individuals aged 40 to 70 years-old (2006-2010) (Sudlow et al., 2015), who were registered with the UK National Health Service and living within a 25 mile radius of one of the 22 study assessment centres. Ethical approval for the UK Biobank was granted by the National Health Service National Research Ethics Service (Ref 11/NW/0382). Written informed consent was provided by all participants. DNA was extracted and genotyped, from a blood sample of each participant, using either the UK BiLEVE Axiom array or the UK Biobank Axiom Array (Bycroft et al., 2018). Refractive error measurement was carried out using non-cycloplegic autorefraction (Tomey RC 5000 auto refkeratometer; Tomey GmbH Europe, Erlangen-Tennenlohe, Germany). Refractive error was defined as the sum of the sphere power plus half the cylinder power, and then averaged between the two eyes. Up to 10 measurements of each eye were taken and the most representative result automatically recorded. All participants completed a touchscreen questionnaire, reporting on demographic, socio-economic and medical factors (Cumberland et al., 2015), followed by a face-to-face interview where additional information such as the age of onset of spectacle wear, ocular history (described in section 1.1.2.), family

history of diseases and early life events was recorded. Participants provided information regarding their education level.

A comparison of the UK Biobank data to the UK Census 2011 data (URL: https://www.nomisweb.co.uk/census/2011, accessed: 29 October 2019) showed similar ethnic composition (90% White, 3.8% Asian/Asian British including 0.5% Chinese, 3.5% Black/Black British, 0.9% of Mixed and 1.5% Other ethnicity) (Cumberland et al., 2015). However, the UK Biobank data is enriched in older, more affluent individuals who have higher educational qualifications, who live in less socioeconomically deprived areas, are less likely to be obese, to smoke, and to drink alcohol on a daily basis (Fry et al., 2017). There were fewer males than females.

1.4.1. Inferring refractive error

Only approximately 25% of participants, from England and Wales, underwent refractive error measurement as a result of an ophthalmic assessment not being introduced until the latter stages of UK Biobank recruitment, presence of eye infection during recruitment or due to having undergone an eye surgery in the preceding 4 weeks of recruitment (Cumberland et al., 2015). Hence, to increase statistical power, I selected as the discovery sample the (bulk of) UK Biobank participants who did not undergo the ophthalmic assessment, but who did answer a questionnaire item asking the age of onset of spectacle (or contact lens) wear. The mean spherical equivalent (MSE) refractive error of participants who reported their age-at-onset of spectacle wear (AOSW), but who did not undergo autorefraction was imputed as follows. First, a statistical model was derived for participants who *did* undergo autorefraction and who reported their AOSW. Then, this model was used to predict the "AOSW-inferred MSE" of individuals who did not undergo autorefraction. The statistical model took the form:

$$MSE_i = \mu + (\alpha \times Sex_i) + (\beta_1 \times Age_i) + (\beta_2 \times Age_i^2) + \ldots. + \left(\beta_j \times Age_i^j\right) + (\delta_1 \times AOSW_i) + (\delta_2 \times AOSW_i^2) + \ldots. + \left(\delta_k \times AOSW_i^k\right) (Eq. 1)$$

Where, MSE_i is the autorefraction-measured refractive error of individual *i*, Sex_i is a binary variable indicating the gender of individual *i*, Age_i is the age in years of

individual *i* when autorefraction was performed, $AOSW_i$ is the age-at-onset of spectacle wear in years of individual *i*.

To optimize the model fit, the sample of participants who self-reported as white-British, did not have ocular history (see section 1.4.2.), who had autorefraction measurement taken and who reported their AOSW was split into a training sample (*n* = 49,435) and a test sample (*n* = 49,435). Polynomials (*j*, *k*) of increasing order were fit using the training sample until the model fit showed no improvement. Specifically, nested models with polynomial order *j* vs. *j* + 1 (or *k* vs. *k* + 1) were compared using a likelihood ratio test (*LRT*), and polynomial order was increased until the *LRT* test suggested no improvement (*LRT* test, $p > 0.05$). The performance of the final model was assessed by calculating the adjusted R^2 between autorefraction-measured MSE and AOSW-inferred MSE in the test sample. The AOSW phenotype yielded an adjusted $R^2 = 0.30$. Moreover, inclusion of interaction terms did not improve model fit, nor did models using year-of-birth in place of *Age*. Models fit in samples with truncated upper age or AOSW ranges also produced similar performance.

AOSW (or, more specifically, age-at-onset of myopia) has been previously shown to have a strong genetic correlation to refractive error (Wojciechowski and Hysi, 2013) as well as strong phenotypic correlation with refractive error (Williams et al., 2013). Furthermore, age-at-onset of myopia has been used previously as a proxy phenotype in a meta-analysis with refractive error (Tedja et al., 2018). To evaluate the genetic overlap between the AOSW-inferred refractive error phenotype and autorefraction-measured refractive error, I carried out a genome-wide association study (GWAS) analysis for each trait (discussed in sections 2.3.1., 2.3.2., 3.3.2. and 3.3.3.) and then performed LD score regression (Bulik-Sullivan et al., 2015) (discussed in section 3.3.2.). Approximately one million 'high-confidence' genetic variants (discussed in section 1.4.2.) were used to estimate the genetic correlation for marginal SNP effects. A high degree of overlap between the two traits was found (genetic correlation $r_g = 0.93$, 95% CI 0.88 to 0.97, $p < 2.2 \times 10^{-16}$). This confirmed that the AOSW-inferred refractive error phenotype was a suitable surrogate for refractive error in UK Biobank participants who did not undergo autorefraction.

1.4.2. Sample quality control for Chapters 2,3 and 6

Of 488,363 individuals with genetic information available, samples were excluded due to: withdrawal of consent (*n* = 68), self-report of non-white-British ethnicity (*n* = 78,661) and outlying level of genetic heterozygosity (beyond ±4 standard deviations from the mean; n = 731). Relatedness was determined by the UK Biobank (Bycroft et al., 2018). To select unrelated participants amongst those who had at least one related member, a genetic relationship matrix (GRM) was created. This GRM was created using a linkage disequilibrium (LD)-pruned set of well-imputed variants (IMPUTE2 $r^2 > 0.9$), minor allele frequency (MAF) > 0.005, missing genotype rate ≤ 0.01, and variants with 'rs' ID prefix). LD-pruning was accomplished by using the --indep-pairwise 50 5 0.1 command in *PLINK* v2 (Chang et al., 2015). From a subset of participants with at least one related member, unrelated individuals were inferred using a genomic relatedness cut-off of less than 0.025 in *PLINK* (Purcell et al., 2007). This sample was combined with participants who did not have any related members according to the UK Biobank. The final 'unrelated' sample had 336,258 individuals from which further exclusions were made due to: AOSW-inferred refractive error measurement not available (*n* = 118,122) and ocular history (*n* = 20,170), which included individuals who self-reported any of the following eye disorders: cataracts, "serious eye problems", "eye trauma", a history of cataract surgery, corneal graft surgery, laser eye surgery, or other eye surgery in the past 4 weeks. Individuals whose hospital records (ICD10 codes) indicated a history of the following were also excluded: cataract surgery, eye surgery, retinal surgery, or retinal detachment surgery. This resulted in a final 'discovery sample' of 197,966 individuals used in this study. An illustration of the above quality control steps is provided in Figure 1.7.

As a 'replication sample', I used individuals who had their refractive error measured using non-cycloplegic autorefraction. The same sample selection procedure as for the discovery sample (see above) was performed, except for an additional step to exclude individuals who did not have a refractive error measurement taken (Figure 1.7.). The final replication sample was comprised of 73,174 individuals.

In the discovery sample of 197,966 unrelated, white-British participants whose genotype data passed quality control and had phenotype information predicted from

the age of onset of spectacle wear (AOSW), the mean ± SD "AOSW-inferred refractive error" was -0.36 ± 1.5 dioptres (D) and the average age was 57.7 ± 7.5 years, while in the replication sample of 73,174 unrelated, white-British participants with autorefraction-measured refractive error, the mean ± SD refractive error was -0.26 ± 2.7 dioptres (D) and the average age was 57.7 ± 7.9 years. More detailed demographics for the discovery and replication samples and their comparisons are presented in Table 1.1. For continuous variables such as age and BMI, comparisons between discovery and replication samples were made using non-parametric Mann-Whitney U test, while variables representing proportions such as prevalence of university education were assessed using two-proportion z-test. All comparisons showed a significant difference at the nominal significance threshold ($p < 0.05$). These observed demographic differences could reflect geographic UK region differences since autorefraction was measured in only 6 out of the 22 UK Biobank assessment centres (due to the late introduction of the eye assessment component into the UK Biobank protocol). The late introduction of the eye measurements component of the UK Biobank study would also have resulted in a small "cohort effect" such that participants who underwent autorefraction had a slightly later year-of-birth (the mean age was 57.8 vs 57.7 years, $P = 6.26 \times 10^{-11}$, in the replication vs. discovery samples, respectively).

1.4.3. Sample quality control for Chapters 4 and 5

The sample used for conditional quantile regression (Chapter 4) and for polygenic interaction score (Chapter 5) analyses differed slightly. Genetic variants were imputed to the haplotype reference consortium panel (the Haplotype Reference et al., 2016) by Bycroft et al. (Bycroft et al., 2018). Exclusion of participants was due to history of ocular condition (*n* = 48,145), withdrawal of consent (*n* = 8), self-report of non-white British ethnicity (*n* = 69,938), outlying level of genetic heterozygosity (*n* = 648), or refractive error not measured (*n* = 283,352). From the remaining sample of 86,286 participants, one of each pair of relatives was removed, as follows. For the subset of participants with at least one relative present, unrelated individuals were inferred using genomic relatedness cut-off of less than 0.025 in *PLINK* (Purcell et al., 2007). In total, the final sample size comprised of 72,985 unrelated, white-British individuals. An illustration of quality control steps for this analysis is provided in Figure 1.8. In the final sample comprising of 72,985 unrelated, white-British participants with autorefraction-

measured refractive error, the mean ± SD refractive error was -0.25 ± 2.67 dioptres (D) and the average age was 57.8 ± 7.8 years.

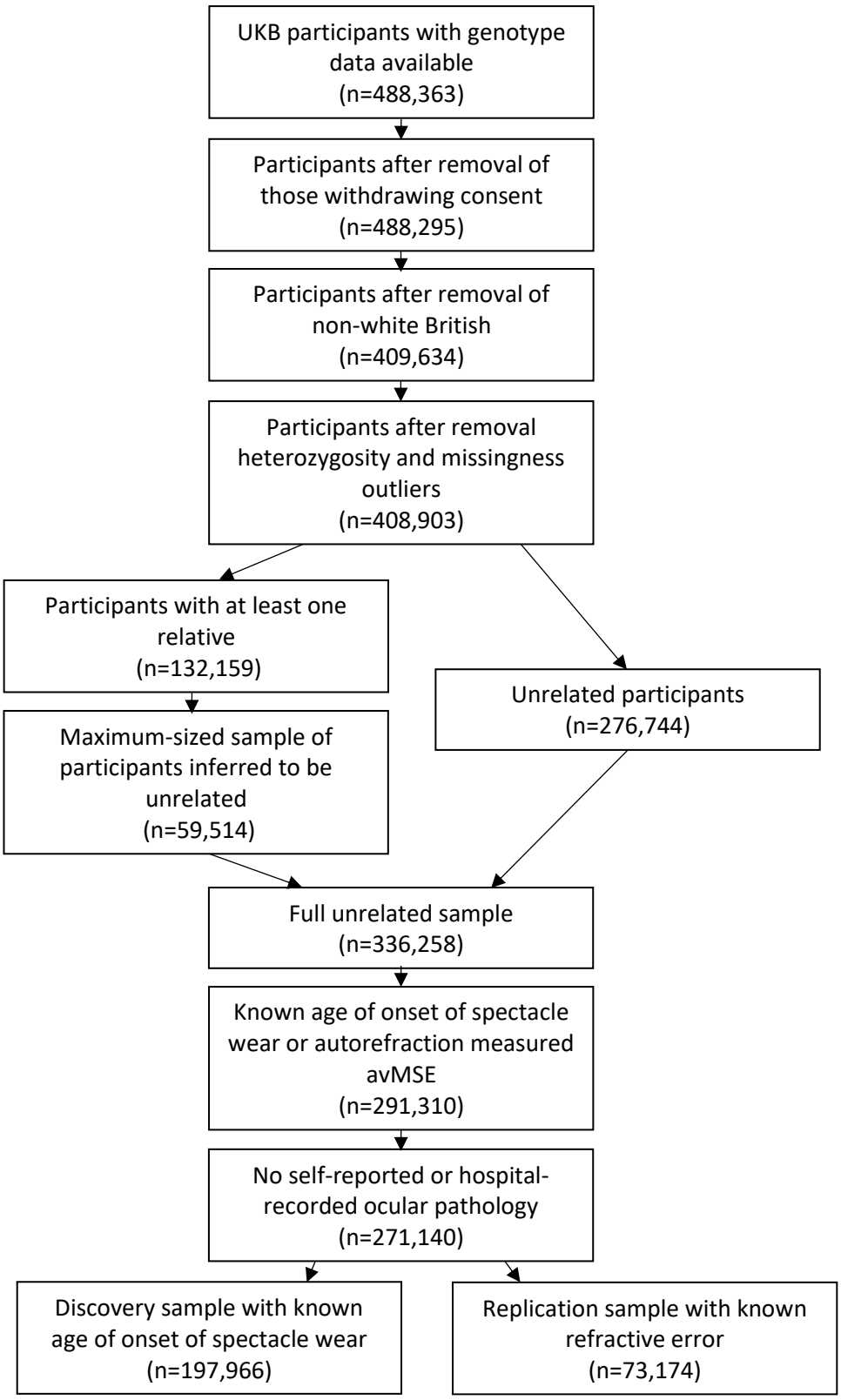


Figure 1.7. Diagram summarising quality control steps performed to obtain the discovery and replication samples. These samples were used in the comparison of risk scales (Chapter 2), vQTL analysis (Chapter 3) and MDR analysis (Chapter 6).

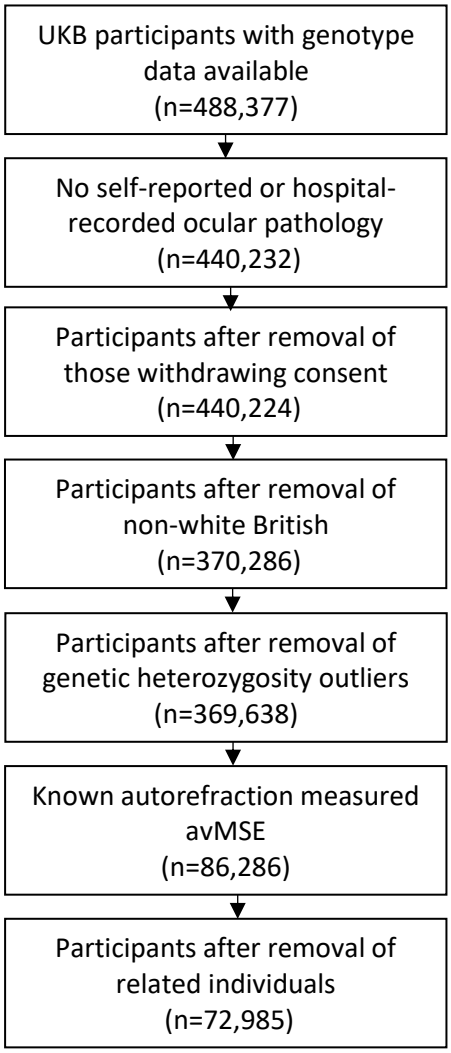


Figure 1.8. Diagram summarising quality control steps used to obtain a sample of unrelated white-British individuals with measured refractive error. The final sample of 72,985 individuals was used for conditional quantile regression analysis (Chapter 4) and assessment of polygenic interaction scores (Chapter 5).

Table 1.1. Demographic characteristics of the discovery and replication samples.

Characteristic	Discovery sample	Replication sample[b]	*P*-value
Sample size	197,966	73,174	-
Female (%)	108,133 (54.6%)	38,583 (52.7%)	1.6×10^{-18}
Age (95% C.I.); years	57.73 (57.7 to 57.8)	57.84 (57.8 to 57.9)	6.26×10^{-11}
BMI (95% C.I.)	27.26 (27.22 to 27.29)	27.42 (27.4 to 27.44)	2.04×10^{-16}
TDI (95% C.I.)	7.27 (7.26 to 7.28)	7.56 (7.54 to 7.58)	3.46×10^{-210}
Autorefraction measured refractive error (95% C.I.); D	-	-0.26 (-0.27 to -0.23)	1.55×10^{-187}
AOSW-inferred refractive error (95% C.I.); D	-0.36 (-0.37 to -0.35)	-	
AOSW (95% C.I.); years	32.2 (32.2 to 32.3)	32.4 (32.3 to 32.5)	4×10^{-3}
Myopic[a] (%)	69,296 (35.0%)	21,817 (29.8 %)	3.6×10^{-142}
University education (%)	62,814 (31.7 %)	25,989 (35.5 %)	1.33×10^{-77}
Age completed full-time education	13-15 years, *N* = 45,468 0.17 (0.16 to 0.18)	13-15 years, *N* = 13,499 0.7 (0.67 to 0.74)	2.57×10^{-218}
	16 years, *N* = 43,227 -0.28 (-0.29 to 0.02)	16 years, *N* = 16,599 -0.02 (-0.05 to 0.02)	4.96×10^{-84}
	17-20 years, *N* = 41,206 -0.49 (-0.51 to -0.48)	17-20 years, *N* = 16,357 -0.43 (-0.48 to -0.38)	3.7×10^{-34}
	21-25 years, *N* = 68,065 -0.69 (-0.70 to -0.67)	21-25 years, *N* = 26,719 -0.76 (-0.81 to -0.70)	5.82×10^{-25}

[a] Myopia was defined as a refractive error ≤ -0.75 D.
[b] Descriptive characteristics of the sample used in Chapter 4 and 5 were very similar.
Abbreviations: AOSW = age-of-onset of spectacle wear; D = dioptre; BMI - body mass index (kg/m2); TDI - Townsend deprivation index.
For continuous variables, comparisons between discovery and replication samples were made using non-parametric Mann-Whitney U test, while binary variables were assessed using two-proportion z-test.

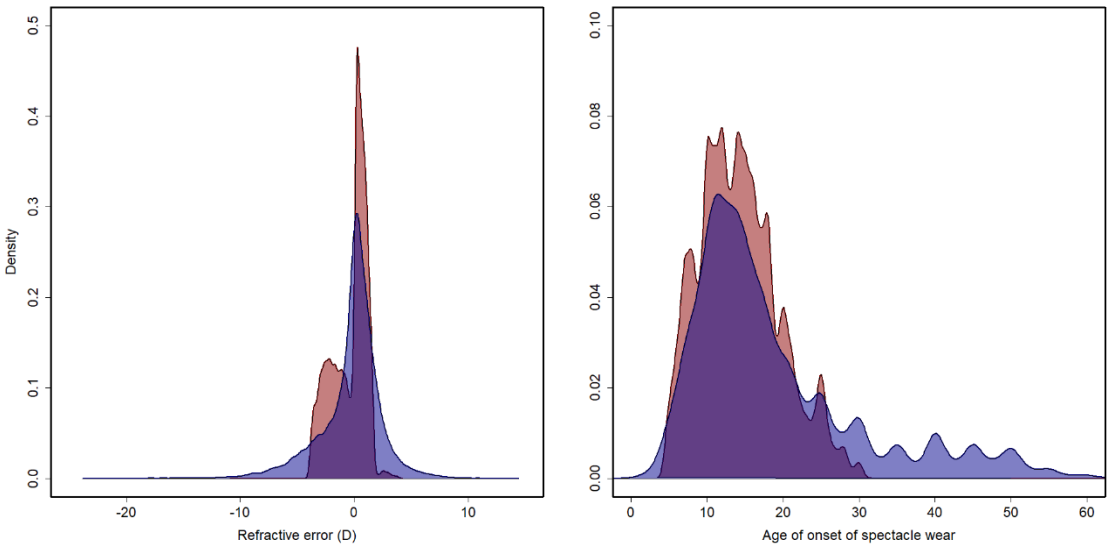


Figure 1.9. The left panel shows the distribution of AOSW-inferred refractive error for all individuals in the discovery sample (red) and the distribution of autorefraction-measured refractive error for all individuals in the replication sample (blue). The right panel shows the distribution of age of onset of spectacle wear in those classified as myopic in the discovery sample (AOSW-inferred refractive error blue ≤ -0.75 D; red) and replication sample (autorefraction-measured refractive error ≤ -0.75 D; blue).

Chapter 2

Comparison of multiplicative and additive risk scales for discovering gene-environment interactions in refractive error

2.1. Introduction

In the early years of genome-wide association studies (GWAS), sample sizes were far too small to detect genetic variants with evidence of gene-environment interaction (GxE) effects. Continued effort to genotype more participants led to larger studies, focusing on datasets comprising approximately one million individuals (Jansen et al., 2018; Yengo et al., 2018). Given that the sample sizes used in GWAS are likely to increase further still in the future, it is reasonable to anticipate that the interest in studying GxE will only increase. To date, susceptibility to myopia has been shown to be determined by both genetic and environmental factors (section 1.1.3.). Despite this, the interplay between the two determinants has not been exhaustively studied previously, and only a handful of GxE for refractive error has been documented (Chen et al., 2011a; Fan et al., 2013; Tkatchenko et al., 2015; Fan et al., 2016).

The vast majority of epidemiological studies that conduct genome-wide environment interaction study (GWEIS) analyses consider a multiplicative scale of interaction (Beaty et al., 2011; Rask-Andersen et al., 2017; Arnau-Soler et al., 2019). An explanation for this is that adding a multiplicative interaction term in the linear or logistic regression is a simple extension to the standard GWAS approach that tests for association between a phenotype and a main or marginal effect of a genetic marker. This facilitates a straightforward interpretation of an interaction effect. For example, in the case of a binary phenotype denoting disease status, GxE can be measured using odds ratios:

$$\frac{OR_{11}}{OR_{01}OR_{10}}$$

Where OR_{11}represents the joint effect of genetic and environmental factors, $OR_{01}$is the effect of the genetic factor and OR_{10} is the effect of the environmental factor. If $OR_{11}/(OR_{01}OR_{10})$ > 1, the multiplicative interaction is said to be positive (i.e. both genetics and environment act as risk factors). If $OR_{11}/(OR_{01}OR_{10})$ < 1, the multiplicative interaction is said to be negative (i.e. both genetics and environment act as preventative factors). If the ratio is 1, the product of a genetic and an environmental factor equals the joint effect of two factors (VanderWeele and Knol, 2014). An interaction measured on the additive scale is less easily understood intuitively. Assuming that the risk ratio approximates the odds ratio, as is the case when the phenotype is rare, an interaction on the additive scale can be estimated as follows:

$$OR_{11} - OR_{01} - OR_{10} - OR_{00}$$

Where OR_{00} is the effect of not being exposed to either the genetic or the environmental factor and all the other estimates are the same as before. This quantity is known as a relative excess risk due to interaction (RERI). Other measures of the additive interaction are derived from RERI and include the synergy index (Rothman, 1974, 1976), which measures the extent to which the effect of both exposures exceeds 1, and the attributable proportion (VanderWeele, 2013), which measures the proportion of the risk in the doubly exposed group of individuals that is due to the interaction itself (VanderWeele and Knol, 2014). Some additional considerations have to be taken into account when studying interactions on the additive risk scale and are discussed in the following sections.

The goal of this chapter was to comprehensively assess the extent to which GxE interactions influence refractive error development from the perspective of both risk scales. Post-GWAS analyses were performed in order to gain insight into the functional effects of GxE interactions.

2.2. Methods

2.2.1. Sample and SNP quality control

A discovery sample (N = 197,966) and a non-overlapping replication sample (N = 73,174) of UK Biobank participants were studied. The selection of these two samples is outlined in Section 1.4.2. The optimized model described in Section 1.4.1. was used

to derive the "AOSW-inferred refractive error" phenotype for participants in the discovery sample. Autorefraction-measured refractive error was available for participants in the replication sample.

Genetic variants with minor allele frequency (MAF) > 0.05, missing genotype rate ≤ 0.01 for directly genotyped markers and INFO > 0.7 for imputed markers and Hardy-Weinberg equilibrium < 1 x 10^{-6} were selected. In total, 5.4 million SNPs were retained.

2.2.2. Assessment of gene-environment interactions on the multiplicative scale

GWEIS on the multiplicative scale was performed using *PLINK* (Purcell et al., 2007), fitting the following linear regression model:

$$Refractive\ error = SNP + UniEdu + SNP \times UniEdu + \\ + Age + Sex + Array + PC1 + PC2 + \cdots PC10 \quad (Eq.\ 1)$$

Where, SNP corresponds to the numeric count of minor alleles carried by a participant (0, 1 or 2) , $Array$ is a binary variable indicating if a participant was genotyped on the UK BiLEVE Axiom array or the UK Biobank Axiom Array (Bycroft et al., 2018) and $PC1 - PC10$ are the first ten principal components, $UniEdu$ corresponds to education coded as *0* if an individual did not have a univ*ers*ity degree and *1* otherwise, and $SNP \times UniEdu$ corresponds to the gene-environment interaction parameter.

2.2.3. Assessment of gene-environment interactions on the additive scale

GWEIS on the additive scale was performed using the *additive_interaction* function in the standalone *R* package of the same name (Mathur and VanderWeele, 2018). Currently, this analysis is only possible for a case-control phenotype. Therefore, before the analysis, the two phenotypes (AOSW-inferred refractive error and autorefraction-measured refractive error) were recoded to represent myopic vs non-myopic status. A threshold of < -0.75 dioptres (D) was used to categorise individuals as either cases or controls.

After performing the GWEIS analyses, I checked whether at least one exposure (genetic or environmental) had a negative association with the phenotype. It has been previously shown that in cases where at least one exposure has a preventive effect (i.e.

OR_{10} < 1 or OR_{01} < 1), factors have to be recoded to get valid measures of interaction on the additive scale (Knol and VanderWeele, 2011; Knol et al., 2011). An explanation for this is that an odds ratio is restricted to lie between 0 and 1 for a preventive factor, while it can go from 1 to infinity for a risk factor (Knol and VanderWeele, 2011). A simple solution to the problem is recoding the variables in such a way that the stratum with the lowest risk when both factors are considered jointly becomes the reference category (Knol and VanderWeele, 2011). In my analysis, all exposures that showed evidence of association with myopic vs non-myopic status were recoded where necessary.

2.2.4. Meta-analysis

To strengthen the evidence of association, the discovery and replication samples were meta-analysed by using *p*-value based fixed effects meta-analysis in *METAL* software (Willer et al., 2010).

2.2.5. *Functional mapping and annotation*

For the GWEIS performed in the discovery and meta-analysed samples, gene-environment interaction summary statistics were analysed using *FUMA* (Watanabe et al., 2017). Gene-based tests were conducted using *MAGMA* (de Leeuw et al., 2015) through the *FUMA* platform using a 50kb buffer for each gene and setting the Bonferroni-corrected significance threshold of $p = 0.05/19{,}061 = 2.62 \times 10^{-6}$ (where 19,061 corresponds to protein coding genes obtained from Ensembl build 85 (Zerbino et al., 2015)). Note that all genes were included in the functional annotation analysis, rather than those remaining after gene-based clumping (described below). In addition, *FUMA* was used to assess functional annotations and pathway enrichment using the sentinel SNPs from associated genomic loci. Specifically, SNPs were clumped according to linkage disequilibrium ($r^2 = 0.01$) using default settings to identify independent lead SNPs with $p < 1 \times 10^{-5}$. Gene mapping was performed in protein-coding genes using positional mapping and expression quantitative trait loci (eQTL) mapping in 53 tissues from the Genotype-Tissue Expression project (GTEx v7) (Ongen et al., 2017). Gene-set enrichment was assessed using differentially expressed genes in the tissue types available from GTEx.

2.2.6. Gene-based clumping

After the gene-based test using the meta-analysis summary statistics estimated for a SNP x University education interaction on an additive scale, several genes were found to be associated with the phenotype. However, an inspection of the region of association highlighted one locus on chromosome 2 that harboured multiple closely located genes. For the purpose of potentially discovering new genes associated with refractive error, gene-based clumping (similar in spirit to the approach implemented in *PLINK* (Purcell et al., 2007) for SNP-based clumping) was performed. I used a sliding window of 5 Mb and selected the gene with the lowest *p*-value in the region as the lead, independently-associated gene. Other genes in the region were excluded from the dataset and the sliding window was advanced by 5 Mb.

2.3. Results

2.3.1. Gene-environment interaction on the multiplicative scale

After testing approximately 5.4 million SNPs in the discovery sample (N = 197,966) for a gene-environment interaction on the multiplicative scale there was only a single SNP with genome-wide significant evidence of an interaction (rs12193446, $p = 2.17 \times 10^{-9}$) (Figure 2.1. top panel and Table 2.1.). rs12193446 is situated on chromosome 6 within an intron of the *LAMA2* gene. This variant showed evidence of replication in the independent replication sample for the autorefraction measured refractive error phenotype ($p = 8 \times 10^{-3}$). The magnitude and the direction of the effect were almost identical for the discovery and replication samples ($\beta = 0.09$ and $\beta = 0.12$ dioptres, respectively). The association between rs12193446 and refractive error was further enhanced when a meta-analysis was performed ($p = 8.2 \times 10^{-11}$, Figure 2.1. bottom panel). A second variant, rs117771785, located on chromosome 12 in the intronic region of the *CCDC38* gene, showed borderline evidence of association in the meta-analysis of the discovery and replication sample GWEIS summary statistics ($p = 8.8 \times 10^{-8}$).

No gene showed evidence of association after correction for multiple testing when GxE summary statistics from the AOSW-inferred refractive error discovery sample were used in a *MAGMA* gene-based analysis. The top gene was *IGF2R* ($p = 1.1 \times 10^{-5}$ in the discovery sample and $p = 0.25$ in the replication sample). The association of *IGF2R* was

further enhanced in the meta-analysed sample (p = 6.5 x 10^{-6}), albeit remaining non-significant.

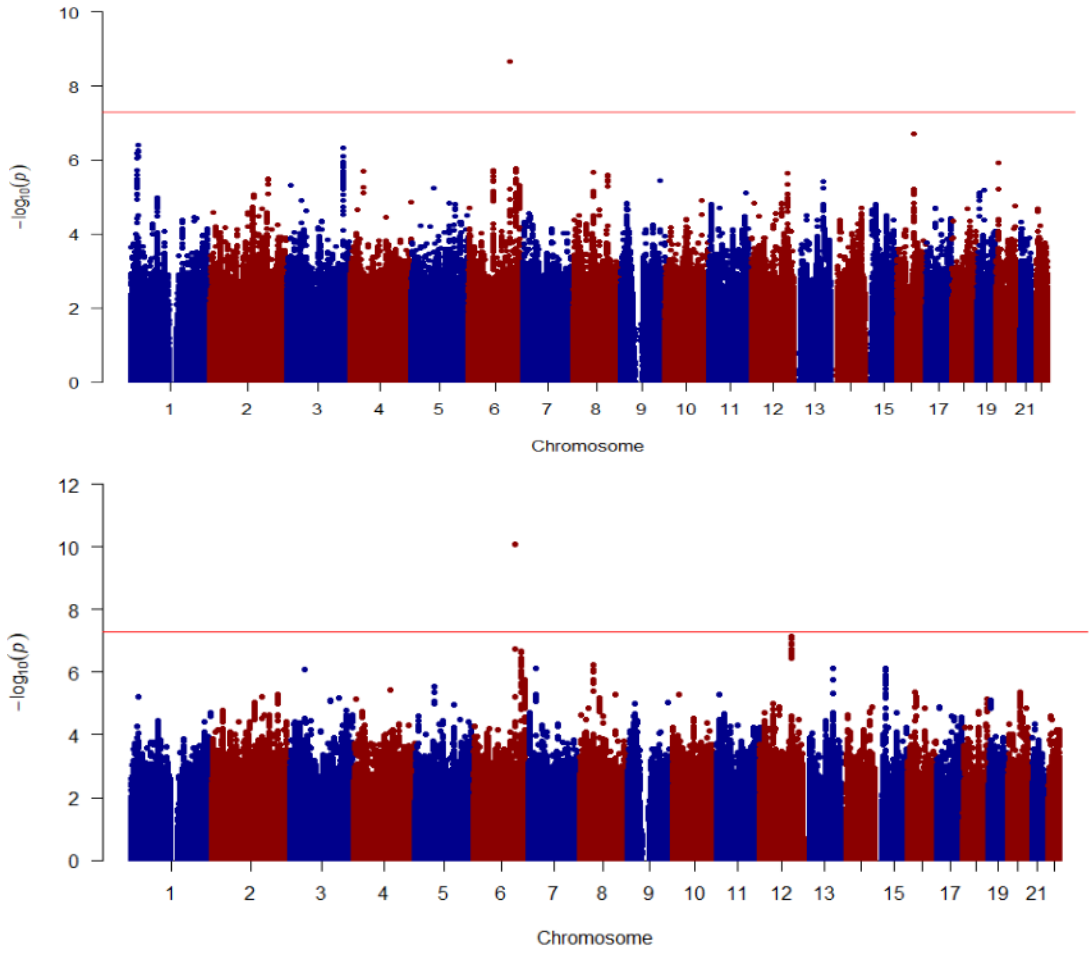


Figure 2.1. Manhattan plots showing genome-wide association for SNP x University education interaction measured on a multiplicative risk scale. Top panel shows the distribution of p-values for the discovery sample. The bottom panel shows genome-wide associations for the meta-analysed discovery and replication samples. The horizontal red line represents genome-wide significance threshold (5 x 10^{-8}).

Enrichment in brain tissues was assessed by prioritising genes based on physical position and eQTL. No genes were found to be differentially up-regulated across GTEx tissues in either the discovery or the meta-analysed samples after Bonferroni correction; two enriched tissues with down-regulated differentially expressed genes were brain cerebellar hemisphere (adjusted-p = 0.014) and brain cerebellum (adjusted-p = 0.018). No tissue showed evidence of enrichment in both directions (i.e. two-sided differential expression).

2.3.2. *Gene-environment interaction on the additive scale*

After dichotomizing individuals into myopes (cases) and non-myopes (controls), approximately 35% of participants were defined as cases in the discovery sample, while 29% were assigned as cases in the replication sample. A GWEIS in the discovery sample that tested for gene-environment interactions on the additive risk scale using this phenotype identified 19 independent genome-wide significant regions (Figure 2.2. top panel and Table 2.2.). Of these 19 loci, 11 (58%) were located in intergenic regions. Among the 19 independent loci, rs12193446 (*LAMA2*), rs634990 (*GJD2*) and rs4738756 (*TOX*) showed evidence of an interaction with University education in the replication sample ($p < 0.05/19 = 2.6 \times 10^{-3}$). All three regions have previously been shown to display marginal SNP effects for refractive error (Tedja et al., 2018). Two additional genomic regions, rs4581716 (*RBFOX1*) and rs913199 (*DNAJC6*) replicated with nominal significance ($p = 0.01$ and $p = 0.05$, respectively). The *RBFOX1* gene was previously identified as being associated with refractive error (Tedja et al., 2018), while *DNAJC6* gene has not previously been implicated. Sign concordance for the RERI statistic was poor (i.e. 10 out of 19 variants showed the opposite direction of effect in the discovery vs. replication sample); however, genetic variants that did replicate with at least nominal significance showed strong agreement (5 out of 5 variants with concordant direction of interaction effect). After recoding of risk factors due to either having a preventative association with the phenotype (i.e. $OR_{10} < 1$ or $OR_{01} < 1$), RERI estimates for each interaction were negative, implying a reduction in risk when both the genetic and environmental (University education) risk factor were present compared to their additive sum.

A meta-analysis of the GWEIS additive scale interaction summary statistics for the case-control phenotype in the discovery and replication samples identified 18 independent genomic loci with evidence of interaction effects (Table 2.3.). Of these 18 loci, 10 were not found in the discovery sample GWEIS alone. A total of 5 of the 10 genomic regions have been previously observed to be associated with myopia development in standard GWAS analyses (Kiefer et al., 2013; Verhoeven et al., 2013; Tedja et al., 2018): rs55634267 (*B4GALNT2*), rs869422 (*ZMAT4*), rs11606250 (*LRRC4C*), rs2853441 (*PRSS56*) and rs779699 (*GRM7*).

In the discovery sample, a gene-based analysis of GxE interaction summary statistics on the additive scale obtained using the dichotomised AOSW-inferred refractive error phenotype suggested the *PDE11A* gene was significantly associated after Bonferroni correction (p = 1.36 x 10^{-6}). However, there was no evidence of an interaction in the replication sample (p = 0.2). *PDE11A* is known to be implicated in myopia development from prior GWAS analyses (Kiefer et al., 2013; Tedja et al., 2018).

A gene-based analysis of meta-analysed additive scale GxE interaction summary statistics from the discovery and replication samples was also performed. An initial analysis before a gene-based clumping was implemented, identified 10 genes that surpassed the *Bonferroni* corrected threshold for multiple comparisons (p < 0.05/19,061 = 2.62 x 10^{-6}). However, a visual inspection of the regional association plots revealed 8 out 10 genes to be located in close proximity. Therefore, in order to identify the set of genes that were independently associated with the case-control myopia phenotype via an interaction with University education, gene-based clumping was performed. In total, five genes were retained after clumping. Table 2.4. shows a list of independent genes that were significant. All five were previously found to be associated with refractive error (Tedja et al., 2018).

Examining the additive scale GWEIS GxE interaction summary statistics from either the discovery sample or the meta-analysis of the discovery and replication samples, no tissue was found to be significantly enriched with eQTLs (i.e. there was no evidence of differentially expressed genes). For example, the lowest two-sided enrichment p-value was observed for small intestine terminal ileum (p = 0.03, adjusted-p = 1), when summary statistics from the meta-analysis were used.

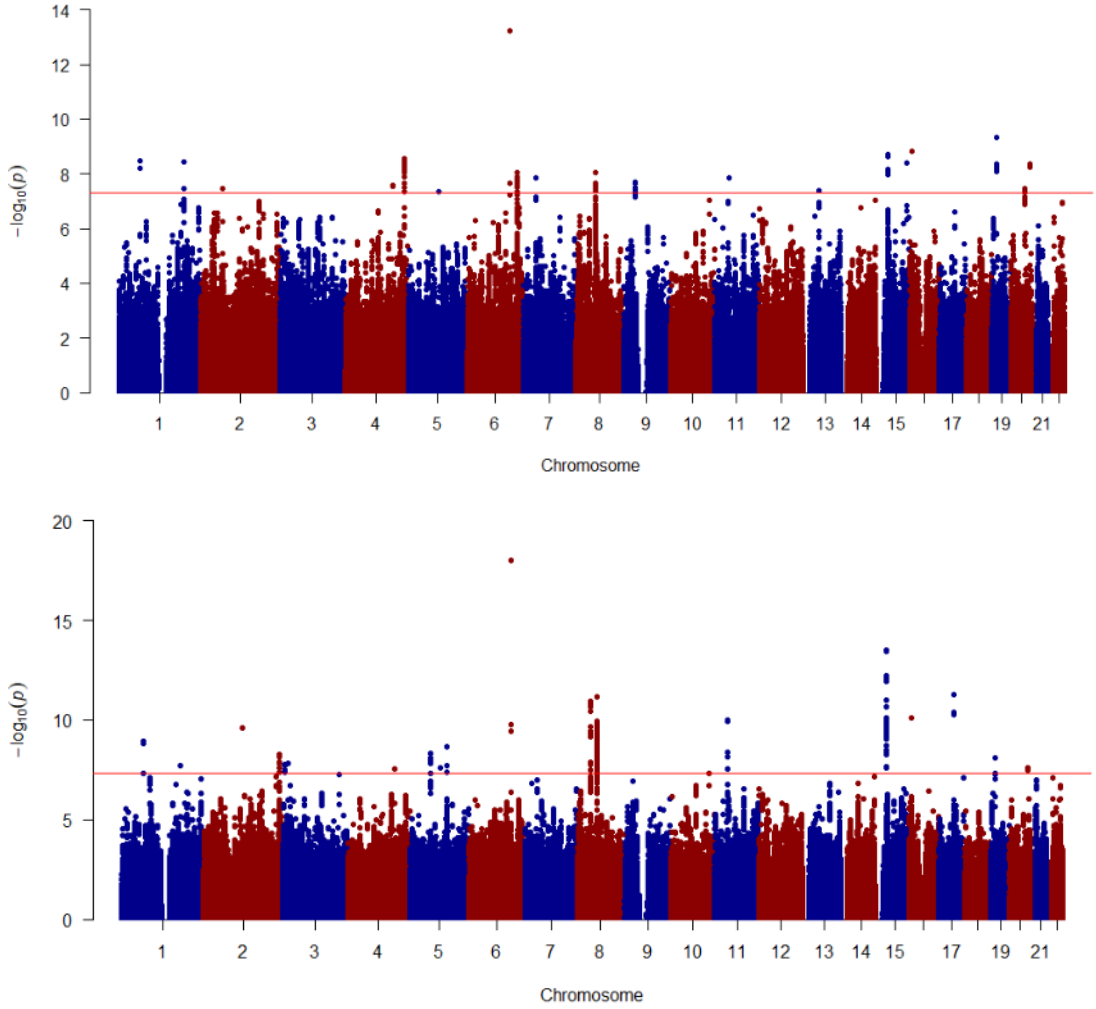


Figure 2.2. Manhattan plots showing genome-wide association for SNP x University education interaction measured on an additive risk scale. Top panel shows the distribution of *p*-values for the myopia case-control phenotype analysed in the discovery sample, while the bottom panel shows results from the meta-analysis of the discover and replication samples. The horizontal red line represents genome-wide significance threshold (5 x 10^{-8}).

Table 2.1. Loci showing genome-wide significant evidence of a SNP x University education interaction in the discovery sample for the continuous AOSW-inferred refractive error phenotype (interaction on the multiplicative risk scale).

SNP	*Nearest gene*	*CHR*	*BP*	*EA*	*N discovery*	*β discovery*	*se discovery*	*P discovery*	*N replication*	*β replication*	*se replication*	*P replication*
rs12193446	*LAMA2*	6	129820038	G	196,375	0.093	0.016	2.17×10^{-9}	72,743	0.126	0.048	8.55×10^{-3}

Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *N* - sample size, *β* - effect size in dioptres, *se* - standard error, *P* - p-value. Results are shown for loci that passed genome-wide significance threshold ($p < 5 \times 10^{-8}$).

Table 2.2. Loci showing genome-wide significant evidence of a SNP x University education interaction in the discovery sample for the myopia case-control phenotype (RERI interaction on the additive risk scale).

SNP	*Nearest gene*	*CHR*	*BP*	*EA*	*N discovery*	RERI *discovery*	RERI *se discovery*	RERI *P discovery*	*N replication*	RERI *replication*	RERI *se replication*	RERI *P replication*
rs12193446	*LAMA2*	6	129820038	G	197,144	-0.272	0.037	6.13×10^{-14}	72,978	-0.361	0.079	2.77×10^{-6}
rs2285955	*OR10H2*	19	15838967	A	193,426	-0.112	0.018	4.67×10^{-10}	71,770	-0.004	0.034	0.45
rs4581716	*RBFOX1*	16	7458135	G	197,144	-0.113	0.019	1.47×10^{-9}	72,978	-0.081	0.035	0.01
rs634990	*GJD2*	15	35006073	C	197,144	-0.107	0.018	1.93×10^{-9}	72,978	-0.156	0.034	1.76×10^{-6}
rs13106566	*LINC01098*	4	178884741	T	196,894	-0.106	0.018	2.82×10^{-9}	72,875	0.025	0.035	0.76
rs913199	*DNAJC6*	1	65871062	T	197,144	-0.105	0.018	3.46×10^{-9}	72,978	-0.055	0.034	0.05
rs12760628	*LHX9*	1	197991694	C	195,334	-0.105	0.018	3.74×10^{-9}	72,353	0.060	0.036	0.95
rs17541406	*RGMA*	15	93889459	G	197,144	-0.108	0.019	4.09×10^{-9}	72,978	0.047	0.036	0.9
rs438810	*LINC01742*	20	56527504	A	196,472	-0.110	0.019	4.47×10^{-9}	72,739	-0.016	0.036	0.32
rs6926368	*AKAP12*	6	151681431	A	195,961	-0.130	0.023	8.56×10^{-9}	72,580	0.042	0.042	0.83
rs4738756	*TOX*	8	60038104	G	194,748	-0.109	0.019	8.63×10^{-9}	72,129	-0.126	0.035	1.7×10^{-4}
rs6462657	*EEPD1*	7	36201713	A	192,795	-0.106	0.019	1.43×10^{-8}	71,548	-0.008	0.036	0.41
rs11820438	*CD82*	11	44649802	T	197,144	-0.101	0.018	1.43×10^{-8}	72,978	0.008	0.035	0.59
rs2812374	*FAM205A*	9	34740951	A	194,685	-0.108	0.020	2.03×10^{-8}	72,108	0.101	0.038	0.99
rs13107432	*INPP4B*	4	143444744	G	194,155	-0.099	0.018	2.64×10^{-8}	71,906	0.075	0.035	0.98
rs12473735	*SPRED2*	2	65896873	A	195,439	-0.098	0.018	3.38×10^{-8}	72,381	0.047	0.035	0.9
rs6130219	*PTPRT*	20	41425612	T	196,187	-0.098	0.018	3.51×10^{-8}	72,627	0.004	0.035	0.54
rs17386667	*HTR2A*	13	47980309	T	193,058	-0.105	0.020	4.06×10^{-8}	71,638	0.079	0.037	0.98
rs1154848	*MCTP1*	5	94263048	A	196,562	-0.099	0.018	4.58×10^{-8}	72,802	-0.045	0.034	0.09

Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *N* - sample size, *se* - standard error, *P* - p-value, *RERI* - relative excess due to interaction. Note that *RERI* effects were recoded because of negative exposure association. Results are shown for loci that passed genome-wide significance threshold in the discovery sample ($p < 5 \times 10^{-8}$).

Table 2.3. SNP x University education interaction summary statistics obtained from a meta-analysis of the myopia case-control phenotype using the additive risk scale.

SNP	*Nearest Gene/RNA*	*CHR*	*BP*	*EA*	*N*	*P*
rs12193446	*LAMA2*	6	129820038	G	270,122	9.24×10^{-19}
rs634990	*GJD2*	15	35006073	C	270,122	3.19×10^{-14}
rs55634267	*B4GALNT2**	17	47283815	T	270,122	5.47×10^{-12}
rs4738756	*TOX*	8	60038104	G	266,877	6.57×10^{-12}
rs869422	*ZMAT4**	8	40723970	G	269,961	1.12×10^{-11}
rs4581716	*RBFOX1*	16	7458135	G	270,122	7.61×10^{-11}
rs11606250	*LRRC4C**	11	40149300	A	267,548	1.00×10^{-10}
rs10172336	*NR_110272**	2	119342963	G	265,194	2.49×10^{-10}
rs913199	*DNAJC6*	1	65871062	T	270,122	1.13×10^{-9}
rs56347383	*MCC**	5	112397011	G	265,427	2.19×10^{-9}
rs12654457	*HTR1A**	5	63015923	C	268,117	4.85×10^{-9}
rs2853441	*PRSS56**	2	233374783	T	269,480	5.66×10^{-9}
rs2285955	*OR10H2*	19	15838967	A	265,196	8.03×10^{-9}
rs4337625	*TBC1D5**	3	17205847	T	268,539	1.41×10^{-8}
rs779699	*GRM7**	3	7519647	G	266,004	1.65×10^{-8}
rs2746325	*RASAL2**	1	178463591	T	270,122	2.00×10^{-8}
rs1154848	*MCTP1*	5	94263048	A	269,364	2.48×10^{-8}
rs438810	*LINC01742*	20	56527504	A	249,211	2.62×10^{-8}

* Denotes genetic variants that were not identified in the discovery sample of case-control myopia phenotype using additive risk scale.
Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *N* - sample size, *P* - p-value. Results are shown for loci that passed genome-wide significance threshold ($p < 5 \times 10^{-8}$).

Table 2.4. Genes with evidence of an interaction with University education implicated by an additive model analysis for the myopia case-control phenotype. Results were obtained from a gene-based *MAGMA* test using SNP x University education interaction summary statistics meta-analysed from the discover sample and replication sample. Results are presented for five genes that were retained after gene-based clumping.

GENE	*CHR*	*START BP*	*STOP BP*	*P discovery*	*P replication*
ECEL1	2	233294537	233402538	4.50×10^{-9}	$5.15 \times 10^{-5*}$
PDE11A	2	178437980	179023066	2.34×10^{-7}	2.01×10^{-1}
METAP1D	2	172814490	172997158	8.08×10^{-6}	$9.77 \times 10^{-3*}$
GJD2	15	35097732	35312040	3.45×10^{-5}	4.88×10^{-2}
TOX	8	40338109	40805352	5.66×10^{-5}	2.39×10^{-1}

* Genetic variants that showed evidence of replication with $p < 0.05/5 = 1.00 \times 10^{-2}$.
Abbreviations: *CHR* - chromosome, *BP* - base pair, *P* - p-value.

2.4. Discussion

The heterogeneous nature of causal factors influencing myopia development ranges from innate genetic variation (Tedja et al., 2018; Wan et al., 2018; Morgan and Rose, 2019) to external environmental exposures (Guggenheim et al., 2012; Read et al., 2015; Mountjoy et al., 2018). Most commonly, these factors are studied in isolation,

therefore limiting interpretability of biological mechanisms leading to changes in the phenotype. It is believed that studying genetic and environmental factors jointly will lead to improved understanding of disease aetiology (Ritchie et al., 2017). In this chapter, I sought to investigate the effect of gene-environment interactions using two measurement scales.

Using the multiplicative scale, which can be viewed as a straightforward extension of a standard genetic association approach, revealed only a single variant (rs12193446) that showed evidence of an interaction with University education (Table 2.1.). Although the variant did replicate using the carefully phenotyped independent replication sample, the nearest gene *LAMA2* has already been identified repeatedly as a susceptibility gene for refractive error (Kiefer et al., 2013; Verhoeven et al., 2013; Tedja et al., 2018). Mutations in *LAMA2* are a common cause of childhood-onset muscular dystrophy, with or without occipital cortex dysgenesis (Ding et al., 2016). The variant rs12193446 is classed as '*deleterious*' by SIFT, '*probably damaging*' by PolyPhen, but '*likely benign*' by CADD. In addition, the effect size of the interaction of 0.09 D (95% CI 0.02 to 0.16 D, $p = 2.17 \times 10^{-9}$) in the discovery sample and 0.13 (95% CI 0.03 to 0.22 D, $p = 8.55 \times 10^{-3}$) in the replication sample was comparatively large: for example, the marginal effect of this variant is 0.09 D (95% CI 0.08 to 0.12 D, $p = 6.24 \times 10^{-29}$) in the discovery sample and 0.38 D (95% CI 0.33 to 0.44 D, $p = 1.47 \times 10^{-41}$) in the replication sample). This raises the question whether the prediction of refractive error could be improved by taking into account the effect of gene-environment interaction when creating polygenic risk scores (this approach is evaluated in Chapter 5). Furthermore, Figure 2.1. demonstrates several other regions that attained a suggestive level of significance ($p < 1 \times 10^{-6}$). This suggests that future studies analysing much larger sample sizes may identify a larger number of interacting loci using the multiplicative scale. For example, the variant with the second strongest association signal was rs1079927 ($p = 2 \times 10^{-7}$); this variant was also nominally significant in the replication sample ($p = 0.04$). rs1079927 is located between the genes *CASC22* and *TOX3*, neither of which has been previously mentioned in the myopia literature; however the related genes *TOX* and *CASC15* (also known as *LINC00340*) have both been associated with refractive error (note that *CASC22* encodes a long non-coding RNA associated with the risk of cancer, as does *CASC15*). Testing for interactions using the

multiplicative scale, *MAGMA* did not identify any genes associated with AOSW-inferred refractive error in the discovery sample. However, insulin-like growth factor 2 (*IGF2R*) - a known refractive error-associated gene - showed a suggestive degree of association ($p = 1.1 \times 10^{-5}$). One of the functions of this gene includes interaction with retinoic acid, which has been shown to be involved in the signalling cascade leading to scleral remodelling in response to myopia-inducing visual stimuli in animal models (Mertz and Wallman, 2000; Troilo et al., 2006).

Using SNP x University education interaction summary statistics obtained from either the discovery or meta-analysis of AOSW-inferred refractive error phenotype on the multiplicative risk scale**,** I found two brain tissue types, brain cerebellar hemisphere (adjusted-p = 0.014) and brain cerebellum (adjusted-p = 0.018), to be enriched for down-regulated differential expression of the candidate genes implicated by the interaction analysis. Given that the connection between the visual system and the central nervous system is well established (Goebel et al., 2012), this could suggest an interconnected relationship between the two in responding to light stimulus.

The number of GxE interaction loci identified as being associated with AOSW-inferred refractive error increased to 19 when the additive risk scale was used (Table 2.2.). Several of these genomic regions, such as *GJD2* and *TOX,* were previously known to be associated with refractive error (Tedja et al., 2018). However, only 5 of the 19 variants showed at least nominal evidence of replication (nevertheless, this is more than would be expected by chance, i.e. 0.05 × 19 ≈ 1). The list of genes with at least nominal replication included *DNAJC6*, a regulator of clathrin-mediated endocytosis in neurons. Among the disorders associated with this gene is the Volkmann type of cataract, which is characterised by a congenital lens opacity (Morgan et al., 2001; Gall et al., 2002; Hirst et al., 2008; Borner et al., 2012). After recoding the RERI statistic, all interaction regression coefficients were found to be negative (RERI < 0). This suggests that the public health consequences of an intervention, by regulating the amount of education received, would be larger in the group with 0 risk alleles for the identified genetic variants (VanderWeele and Knol, 2014). A meta-analysis revealed an additional 10 loci with evidence of an interaction on the additive scale, of which 5 were previously known to be relevant to myopia (Table 2.3.). The novel regions included *MCC, HTR1A, TBC1D5* and *RASAL2* genes, and one small interfering RNA (siRNA) NR_110272. One of the

functions of the *MCC* gene is indirect negative regulation of beta-catenin (*CTNNB1*) transcriptional activity. *CTNNB1* is known to be associated with neurodevelopmental disorders such as spastic diplegia and visual defects, which can lead to strabismus, optic nerve atrophy and retinal abnormalities among a range of other symptoms (Tucci et al., 2014). The connection with remaining regions is less obvious, since a relationship with ocular disorders has not been established. However, taking into account the molecular function of *HTR1A, TBC1D5* and *RASAL2* genes, I speculate that the connection with refractive error could lie in regulation of G-nucleotide-binding protein activity. A gene-based analysis of the additive scale interaction meta-analysis results only identified five genes known to be associated with refractive error (Table 2.4.) (Tedja et al., 2018).

Collectively, the results from both the discovery sample and the meta-analysis suggest that genetic variants implicated in gene-environment interactions are likely to be enriched among the variants that display marginal SNP associations with refractive error. In the meta-analysis of additive scale summary statistics, one variant was located near a small interfering RNA NR_110272 about which not much information is available. siRNA's alter the expression of specific genes by binding and degrading mRNA after transcription has occurred (Carthew and Sontheimer, 2009). Furthermore, given that more than 50% of identified genetic variants clustered in intergenic regions, rather than protein-coding regions, this suggests that interacting variants may exert their effect by regulating gene-expression, an observation that is consistent with that made by studying SNPs with marginal effects.

It was notable that the difference in the number of interacting loci identified in the additive vs. multiplicative scale analyses was not simply due to the way the phenotype was coded. For example, using a case-control myopia phenotype and testing for GxE using a multiplicative interaction scale in the discovery sample identifies rs11591075 variant near *PLD5* gene as one of the most strongly associated (OR 1.07, 95% CI 1.04 to 1.11, $p = 2.24 \times 10^{-5}$) and no evidence of association for this variant is found in the replication sample (OR 1.00, 95% CI 0.95 to 1.06, $p = 0.95$). Note that the strength of association is several magnitudes below the genome-wide significance-threshold.

As gene-environment interactions can be studied with reference to either an additive or a multiplicative scale, this raises the question of which scale is more appropriate, or in other words, which scale is better suited for examining biologically important mechanisms leading to changes in the phenotype of interest. On the one hand, the ease of fitting models exploring multiplicative scale interactions in all statistical software allows for easy interpretation of interaction coefficients (VanderWeele and Knol, 2014). Conversely, it has been argued that the additive scale is more relevant if the goal is to assess the public health importance of an intervention (for example, an intervention for a modifiable environmental risk factor) (VanderWeele and Knol, 2014). However, for quantitative traits such as refractive error this argument may not hold. Given the arguments in favour of each of the two scales, several authors have recommended reporting findings using both scales (VanderWeele and Knol, 2014; Gauderman et al., 2017).

One limitation of the current study was the use of AOSW-inferred refractive error as the phenotype in the discovery sample. This phenotype has a correlation of approximately 0.55 with autorefraction-measured refractive error. Furthermore, as reported in section 3.3.2, the genetic correlation between AOSW-inferred refractive error and autorefraction measured refractive error was 0.93. The phenotypic correlation of 0.55 indicates that the AOSW-inferred phenotype was informative yet imprecise. For certain participants, their approximate refractive error could be inferred with high confidence. For example, individuals with an AOSW between 0-6 years of age were predominantly hyperopic; those with an AOSW between 6-27 years of age were predominantly myopic; those with an AOSW >40 years were mostly emmetropic or low hypermetropes. However, there were exceptions to these general trends, such as myopes with a very early AOSW. Importantly, the exact *level* of AOSW-inferred myopia and hyperopia would also have been imprecise, even if participants were correctly identified as myopic or hyperopic. This was especially evident from the truncated distribution of the AOSW-inferred refractive error phenotype compared to the autorefraction-measured phenotype (Figure 1.9). Nevertheless, the very large size of the discovery sample (approximately three times larger than the replication sample) partly compensated for the imprecision of the AOSW-inferred phenotype. This was evidence from the genetic correlation of 0.93 between the two traits, which

demonstrated that despite the limited accuracy in gauging refractive error using AOSW, the GWAS for AOSW-inferred refractive error identified a highly overlapping set of associated variants to the GWAS for autorefraction measured refractive error and correctly estimated the relative degree of association of these variants with refractive error (note that the two GWAS analyses also identified an equivalent number of genome-wide significant variants).

The rationale for using the AOSW-inferred refractive error in the discovery sample was that it provided a strategy for selecting candidate variants that could then be tested using the gold-standard autorefraction-measured phenotype. A second limitation of my study was the necessity to dichotomize the phenotype in order to permit an analysis of interactions on the additive scale, which required the use of an arbitrary classification threshold for assigning case/control status. It is not clear if using < -0.75 D as a threshold to classify individuals as myopes vs. non-myopes was the optimal threshold. A third limitation was the fact that the calculation of the *RERI* statistic makes the assumption that the phenotype is rare, such that odds ratios approximate risks ratios. Strictly speaking, this was not true in my analyses given that the prevalence of myopia in the discovery sample was 35%. I did consider stratifying individuals as cases vs. controls using a more stringent threshold (e.g. a threshold of < -3.00 D would yield a prevalence of 7%, making approximation between odds ratios and risk ratios more appropriate). However, a decision was made to use < -0.75 D because it is a more commonly used threshold in the myopia literature (Holden et al., 2016). The findings from the analysis of AOSW-inferred refractive error using the additive interaction scale provided some empirical support for this choice: the analysis in the discovery sample successfully identified 4 loci that were already known to be implicated in refractive error development, and all of these loci were at least nominally significant in the replication sample.

In summary, A GWEIS for SNP x University education interactions on the additive scale led to the identification of 5 potentially interacting loci. Crucially, most of the putative interactions that were identified using the discovery sample did not replicate, suggesting they were likely to be false positives. A large proportion of the loci identified as having interaction effects involving education were previously known to be associated with refractive error through standard GWAS analyses. This suggests the

enrichment of interacting loci among those showing a marginal association. Although it might be easier to study GxE using the multiplicative scale of interaction, this work suggests that future studies may benefit from using the additive scale.

Chapter 3

Genome-wide association study for loci controlling phenotypic variability in refractive error

3.1. Introduction

The standard approach adopted in genome-wide association studies (GWAS) for a quantitative trait is to estimate the difference in the *mean* level of the phenotype for individuals who differ in genotype at a given locus. However, the alternative approach of testing for a difference in phenotypic *variance* across genotype classes has also been proposed (Hill and Mulder, 2010; Pare et al., 2010). A visual representation of a hypothetical example of a locus where variance differs depending on the genotype is given in Figure 3.1. Extensive simulations and theoretical work have shown that such variance heterogeneity can arise at loci involved in either gene-environment or gene-gene interactions (Struchalin et al., 2010; Forsberg and Carlborg, 2017; Al Kawam et al., 2018). For example, Yang et al. reported variance heterogeneity for BMI at the *FTO* locus that could not be accounted for by scale effects (Yang et al., 2012b), and subsequent work (Young et al., 2016) found evidence of interactions between *FTO* locus variant rs1421085 and diet, physical activity and sleep.

Given that refractive error has a strong genetic and environmental basis, I aimed to discover genetic variants associated with variance heterogeneity of this trait, hypothesizing that this could have arisen as a result of gene-environment or gene-gene interactions. The extent of variance heterogeneity across different genotype classes was estimated using Levene's median test (Brown and Forsythe, 1974), which is based

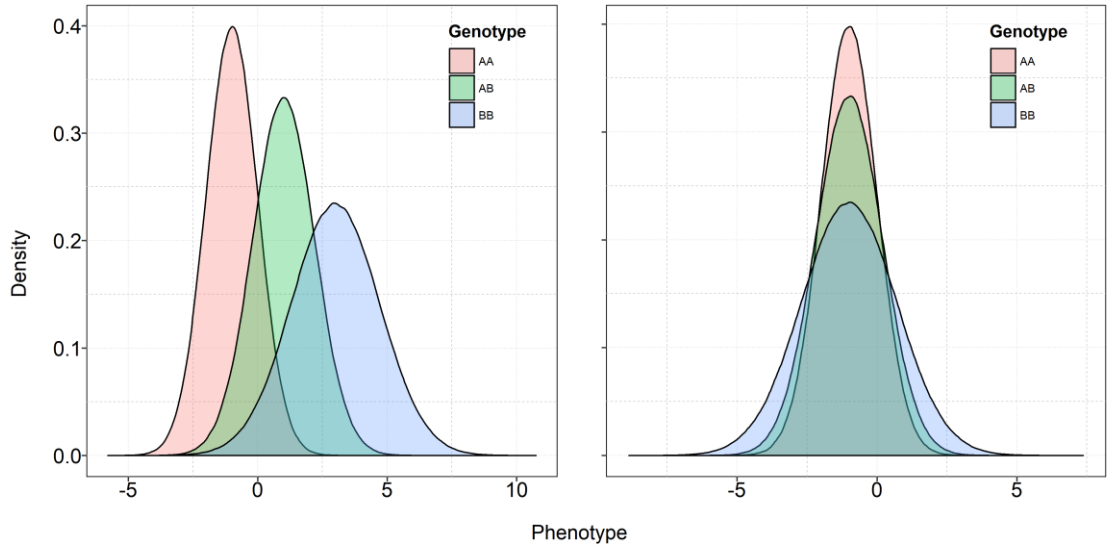


Figure 3.1. Detection of genetic variants that cause variance heterogeneity. The plot in the left panel illustrates a variant that increases both the phenotype mean and phenotype variability with each additional copy of the "B" allele. Such variance heterogeneity could be detected using Levene's median test, while the effect on the phenotype mean could be detected using a standard GWAS. The plot in the right panel illustrates a locus with no effect on the phenotype mean, but an increase in phenotype variance with each additional copy of the "B" allele. This locus could be discovered via Levene's median test but could not using a standard GWAS approach.

on a one-way analysis of variance (ANOVA). It has been noted that the existence of a mean-variance relationship across genotype classes can introduce bias when testing for variance heterogeneity (Young et al., 2018). Therefore, a sensitivity analysis was carried out using a genome-wide analysis that accounted for this potential source of bias (Lee and Nelder, 2006; Rönnegård et al., 2010; Young et al., 2018).

3.2. Methods

3.2.1. Study participants and sample quality control

A discovery sample (N = 197,966) and a non-overlapping replication sample (N = 73,174) of UK Biobank participants were studied. The selection of these two samples is outlined in Section 1.4.2. The optimized model described in Section 1.4.1. was used to derive the "AOSW-inferred refractive error" phenotype for participants in the discovery sample. Autorefraction-measured refractive error was available for participants in the replication sample. SNP quality control was performed in *PLINK* (Purcell et al., 2007) after converting genotypes from BGEN format to hard calls. The

genome-wide analysis included 5.4 million variants with minor allele frequency (MAF) 0.05 or greater, missing rate per SNP 0.02 or less and $p > 1 \times 10^{-6}$ for a test of Hardy-Weinberg equilibrium.

3.2.2. Levene's median test: genome-wide analysis in the discovery sample

Variance heterogeneity across genotype categories was assessed for 5.4 million variants in the discovery sample using Levene's median test, implemented in the *OSCA* software (Zhang et al., 2019). The extent of variance heterogeneity was investigated using the following formula:

$$W = \frac{(N-k)}{k-1} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} N_i (Z_{ij} - Z_{i\cdot})^2} \qquad (Eq. 1)$$

where k is the number of different genotype groups (for a bi-allelic SNP k = 3), N_i is the sample size of the *i*th group, N is total sample size across all genotype groups, Z_{ij} = $|Y_{ij} - \tilde{Y}_{i\cdot}|$, where Y_{ij} is the phenotype of the measured variable for the *j*th sample from the *i*th group and $\tilde{Y}_{i\cdot}$ is a median of the *i*th group, $Z_{i\cdot}$ is the mean of the Z_{ij} for *i*th group and $Z_{\cdot\cdot}$ is the mean of all Z_{ij}. A similar model includes Levene's mean test, where $\tilde{Y}_{i\cdot}$ is replaced by the mean of the *i*th group. This model was assessed when simulating type 1 error along with Levene's median test. When performing variance quantitative trait locus (vQTL) analysis using Levene's test, age, sex, genotyping array and the first 10 principal components were included in the model as covariates.

3.2.3. Sensitivity analysis: Heteroskedastic linear model

As a sensitivity analysis, a variance heterogeneity test was also performed in the discovery sample using a heteroskedastic linear model (*HLMM* software (Young et al., 2018)). This model estimates an additive component that corresponds to the *mean* effect of a genotype, a log-linear variance component that reflects the *variance* effect of a genotype, and a dispersion component that corrects for a generalised mean-variance relationship (r_{av}). The latter component is of interest in this chapter, since it can indicate a locus involved in a gene-environment of gene-gene interaction, while accounting for the statistical noise. Calculation of dispersion effects involves two steps. First, the correlation between the additive and log-linear variance effect of all available variants is calculated. Second, variants are identified that show evidence of residual

variance heterogeneity after accounting for the mean-variance relationship. Tests were performed for the same set of 5.4 million variants described above. Age, sex, genotyping array and first 10 principal components were included as covariates, and analyses were performed while correcting for the *mean* and the *variance* effects of these confounders.

3.2.4. Assessment of type-1 error rate due to non-normal trait distribution

The AOSW-inferred refractive error phenotype was not normally distributed, hence I performed simulations to investigate if this non-normality increased the false positive rate of either Levene's test (mean or median), or the heteroskedastic linear model implemented in *HLMM* (false positive rate was estimated for the additive, log-linear variance and dispersion components). One thousand biallelic genetic variants with MAF > 0.05 and no effect on the phenotype were simulated using the *PLINK* --simulate-qt command in the discovery sample (n=197,966), while retaining the relationship between the phenotype and all of the covariates. The type-1 error rate was calculated as the proportion of variants that passed α = (0.01,0.05) thresholds. Simulations were performed twice: first using untransformed AOSW-inferred refractive error, followed by the rank-based inverse normally transformed (RINT) AOSW phenotype. The RINT transformation leads to a phenotype with a standard normal distribution.

3.2.5. Independent replication of variants identified in the discovery sample

Variants associated with variance heterogeneity in the discovery sample (Levene's test $p < 5 \times 10^{-8}$) were clumped in order to identify the lead variant in each region. Clumping was performed with *PLINK* using a physical distance threshold of one megabase and an LD threshold of $r^2 < 0.01$. These lead variants were evaluated using Levene's test in the replication sample (n=73,174) for the autorefraction measured refractive error phenotype. Independent replication using *HLMM* was performed similarly for variants with *HLMM* $p < 5 \times 10^{-8}$ in the sensitivity analysis.

3.2.6. Definition of novel variants

I sought to discover whether any of the genomic regions exhibiting variance heterogeneity overlapped with regions known to have marginal associations with refractive error. I restricted the comparison to the 149 genomic regions identified in the CREAM + 23andMe GWAS meta-analysis reported by Tedja et al., that showed at

least nominal ($p < 0.05$) replication in the UK Biobank sample. All variants (i.e. variants identified by vQTL analysis and variants identified by *CREAM*) with LD threshold of $r^2 > 0.1$ of the lead vQTL were considered to belong to the same region.

3.2.7. Effect size similarity between AOSW-inferred and autorefraction measured refractive error

To evaluate the concordance of the Levene's test results obtained when analysing either the AOSW-inferred refractive error phenotype or the autorefraction-measured refractive error phenotype on a genome-wide basis, I applied Levene's median test genome-wide in the replication sample and calculated the genetic correlation of effects sizes at each locus for the two traits using LD score regression (Bulik-Sullivan et al., 2015). LD score regression was discussed in Chapter 1 section 1.4.1.

3.2.8. Gene-based association and gene-set enrichment

Using *MAGMA* software (de Leeuw et al., 2015), I performed a gene-level association analysis based on summary statistics obtained from Levene's median test in the discovery sample. Variants were annotated to genes based on NCBI (Build 37.3) gene definitions (NCBI Resource Coordinators, 2016). In order to include possible transcription regulators, I included 50 Kb regions upstream and downstream of the transcribed region. A total of 18,452 genes had at least one variant mapped to them, and 62.3% of the variants were mapped to at least one gene. Variants that did not map to any gene were excluded from the analysis. For the purpose of potentially discovering new genes associated with refractive error, I performed gene-based clumping. I used a sliding window of 5 Mb and selected the gene with the lowest *p*-value in the region as the lead, independently-associated gene. Other genes in the region were excluded from the dataset and the sliding window was advanced by 5 Mb. Genes that showed evidence of association with AOSW-inferred refractive error after clumping, were selected for replication. In addition, for the full list of 18,452 genes that had at least one variant mapped to them, *biological pathway*, *cellular component* and *molecular function* Gene Ontology domains were used to perform gene-set enrichment analysis in *MAGMA* (de Leeuw et al., 2015). I used $p < 0.05/(3 \times$ No of categories in each domain) as the threshold for selecting true positive enrichment.

3.2.9. Testing vQTL loci for direct evidence of SNP x education interaction

For the lead vQTL identified using Levene's median test in the discovery sample, I tested for direct evidence of gene-environment interaction using a binary variable (*UniEdu*) indicating whether participants self-reported having a university degree. The variable was coded *1* if an individual reported obtaining a university degree and *0* otherwise. A genotype x *UniEdu* interaction was tested in the replication sample for the autorefraction measured refractive error phenotype using the following equation:

$$Refractive\ error = SNP + UniEdu + SNP \times UniEdu + \\ + Age + Sex + Array + PC1 + PC2 + \cdots PC10 \qquad (Eq.\ 2)$$

Where, SNP corresponds to the numeric count of minor alleles carried by a participant (0, 1 or 2) , $Array$ is a binary variable indicating if a participant was genotyped on the UK BiLEVE Axiom array or the UK Biobank Axiom Array (Bycroft et al., 2018) and $PC1 - PC10$ are the first ten principal components.

In addition, 'Age completed full-time education' (*EduYears*) was selected as another environmental exposure variable. UK Biobank participants with a university degree were not asked the age they completed full-time education; hence these individuals were assumed to have completed their education at the age of 21 years. Age completed education categories with low counts were merged with adjacent categories, resulting in four final *EduYears* categories: 13–15, 16, 17–20, and 21–26 years, each of which comprised at least 40,000 individuals. A linear regression model that included a term for a vQTL x *EduYears* interaction, similar to the one in (*Eq*. 2), was performed including the same set of covariates as in the original analysis. For the analysis of the categorical variable *EduYears*, interaction effects were estimated using the 13-15 years group as the reference group.

3.3. Results

3.3.1. Simulations to determine type-1 error rate of variance heterogeneity tests

It is known that some tests for variance heterogeneity, such as Bartlett's test, can be sensitive to the underlying distribution of the trait of interest (Wang et al., 2019). Given the non-normal distribution of the AOSW-inferred refractive error phenotype, simulations were performed to find out whether the appropriate type 1 error rate was

maintained when tests for variance heterogeneity were carried out. For the Levene's tests implemented in the *OSCA* software (Zhang et al., 2019), Levene's mean test showed massive inflation of the type-1 error rate when analysing the untransformed phenotype (Table 3.1.). Levene's median test attained correct type-1 error rates. Both tests had appropriate type 1 error rates at α = 0.05 threshold when the phenotype was rank-based inverse normally transformed. These results suggested that the use of Levene's median test for analysis of the untransformed phenotype would perform well for a genome-wide vQTL analysis of refractive error.

For the heteroskedastic linear model implemented in the *HLMM* software (Young et al., 2018), I observed very good control of the type-1 error rate for the additive and dispersion effects when analysing either the untransformed or inverse-normal transformed AOSW-inferred refractive error phenotype (Table 3.2.). However, there was a lower than-expected proportion of false positives for the log-linear variance component; for example, the proportion of false positive findings was 3.6 times smaller than expected at α = 0.05 using the untransformed phenotype. Analyses using the transformed refractive error phenotype partially corrected the lower than expected false positive rate. Given that I did not observe an inflation of false positive findings, the genome-wide vQTL analysis using heteroskedastic linear model was performed using untransformed AOSW-inferred refractive error as the phenotype.

3.3.2. Genome-wide vQTL analysis using Levene's median test

Genome-wide analyses were performed for 5.4 million genetic variants in a discovery sample of 197,966 unrelated participants for the untransformed AOSW-inferred refractive error phenotype. Validation was assessed in a replication sample of 73,174 unrelated participants for the untransformed autorefraction-measured refractive error.

Table 3.1. Type-1 error summary for different Levene's vQTL tests implemented in *OSCA* software.

	Levene's mean test		Levene's median test	
	(α) = 0.01	(α) = 0.05	(α) = 0.01	(α) = 0.05
Untransformed phenotype	0.049	0.144	0.01	0.045
Transformed* phenotype	0.008	0.039	0.009	0.039

The rank-based inverse normal transformation of AOSW-inferred refractive error was performed, resulting in a normal distribution with mean zero dioptres and standard deviation of one (N(0,1)). Type-1 error threshold is represented by α.

Table 3.2. Summary of simulation results for additive, log-linear variance and dispersion components estimated by using the heteroskedastic linear model (*HLM*).

		Type 1 error rate threshold (α)	
	Component	(α) = 0.01	(α) = 0.05
Untransformed phenotype	Additive	0.010	0.049
	Log-linear variance	0.001	0.014
	Dispersion	0.010	0.049
Transformed phenotype	Additive	0.010	0.050
	Log-linear variance	0.006	0.030
	Dispersion	0.010	0.050

The rank-based inverse normal transformation of AOSW-inferred refractive error was performed, resulting in a normal distribution with mean of zero dioptres and standard deviation of one (N(0,1)). Type-1 error threshold is represented by α. Note that *HLMM* estimates log-linear variance instead of variance effects because the location measure of distribution cannot be negative.

Considering all 5.4 million variants, the estimated genetic correlation between AOSW-inferred refractive error and autorefraction-measured refractive error was 0.78 (95% CI 0.65 to 0.91, $p = 1.6 \times 10^{-31}$) for the effect estimates obtained from Levene's median test. In comparison, a genetic correlation between the two traits of 0.93 (95% CI 0.88 to 0.97, $p < 2.2 \times 10^{-16}$) was observed for marginal SNP effect estimates obtained using a standard additive GWAS model. This demonstrates a strong genetic overlap between the discovery and replication phenotypes based on the variants with marginal effects and/or effects on variance heterogeneity. The variance heterogeneity analysis using Levene's median test in the discovery sample identified 48 independent loci that showed evidence of unequal variance across genotypes (Figure 3.2., Table 3.3.). Approximately one third (n = 14) of these vQTLs passed the Bonferroni corrected threshold for multiple comparisons ($p < 0.05/48 = 1 \times 10^{-3}$) in the replication sample and 34 (71%) showed at least nominal replication ($p < 0.05$).

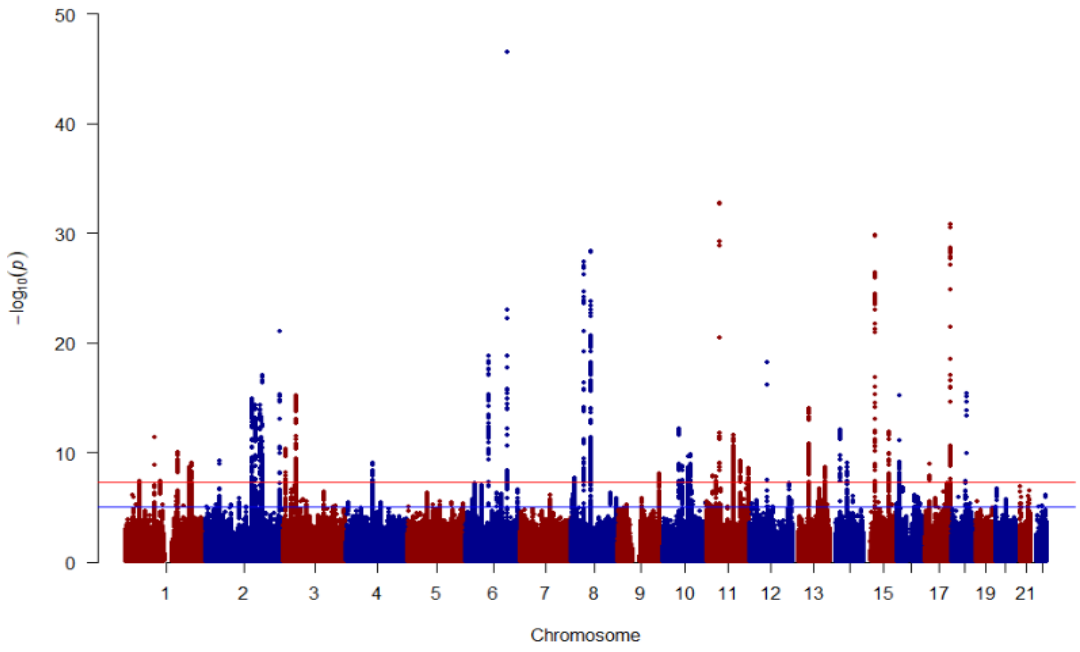


Figure 3.2. Manhattan plot showing genome-wide associations based on Levene's median vQTL analysis in the discovery sample. The red horizontal line shows genome-wide significant threshold ($p < 5 \times 10^{-8}$), while the blue horizontal line depicts suggestive genomic regions ($p < 1 \times 10^{-6}$).

Directionality concordance between effects in the discovery and replication cohorts was observed for 45 out of 48 vQTLs. Variant rs12193446 located near the *LAMA2* gene displayed the highest level of variance heterogeneity and the strongest evidence of association with the phenotype in the discovery sample. Each copy of the risk allele was associated with a 0.077 D decrease in refractive error variance (95% CI 0.066 to 0.088, $p = 2.8 \times 10^{-47}$).

3.3.3. Sensitivity analysis: genome-wide vQTL analysis using the heteroskedastic model

In a genome-wide analysis using the heteroskedastic linear model, 14 vQTLs showed evidence of dispersion effects in the discovery sample (Table 3.4. and Figure 3.3.). However, in the replication sample, none of these 14 variants showed evidence of a dispersion effect after adjustment for multiple comparisons ($p < 0.05/14 = 3.6 \times 10^{-3}$). Nevertheless, 2 variants rs3138142 (*RDH5*) and rs7775087 (*ZMAT4*) showed nominal evidence of replication ($p = 0.010$ and $p = 0.048$, respectively). All 14 of the variants (or variants with LD threshold of $r^2 < 0.01$ within a one megabase physical distance window) identified by the heteroskedastic linear model were also identified using

Levene's median test (Table 3.4.). Thus, this suggested that approximately 30% of the genome-wide significant variants identified by Levene's test displayed variance heterogeneity across genotypes that could not be explained purely by confounding due to a mean-variance relationship. In addition, all 14 of these vQTL were in LD (r^2 > 0.1) with variants known to have marginal associations with refractive error as identified by the *CREAM* consortium (Tedja et al., 2018). Similarly to the GWAS using Levene's test, rs12193446 near *LAMA2* displayed the strongest evidence of a dispersion effect in the discovery sample (β = -0.139, 95% CI -0.166 to -0.112, p = 6.3 x 10^{-23}).

Table 3.3. Summary of 48 independent vQTLs with $p < 5 \times 10^{-8}$ identified using Levene's median test in the discovery sample. Genetic variants are ranked based on their *p*-value in the discovery sample. The phenotype analysed in the discovery sample was AOSW-inferred refractive error, and the phenotype analysed in the replication sample was autorefraction-measured refractive error. All results are for Levene's median test.

SNP	*Nearest Gene*	*Chr*	*BP*	$\beta_{discovery}$	$se_{discovery}$	$P_{discovery}$	$N_{discovery}$	$\beta_{replication}$	$se_{replication}$	$P_{replication}$	$N_{replication}$
rs12193446	*LAMA2*	6	129820038	-0.078	0.005	2.88×10^{-47}	197,966	-0.062	0.009	2.52×10^{-12}*	73,174
rs11606250	*LRRC4C*	11	40149300	0.051	0.004	1.75×10^{-33}	196,204	0.051	0.007	3.90×10^{-13}*	72,514
rs7405453	*TSPAN10*	17	79615572	0.039	0.003	1.52×10^{-31}	197,002	0.037	0.005	6.47×10^{-12}*	72,881
rs524952	*GJD2*	15	35005886	0.037	0.003	1.51×10^{-30}	197,854	0.042	0.005	1.48×10^{-15}*	73,131
rs10089517	*TOX*	8	60178721	-0.037	0.003	4.05×10^{-29}	197,966	-0.022	0.005	6.75×10^{-5}*	73,174
rs16890057	*ZMAT4*	8	40726582	-0.043	0.004	3.81×10^{-28}	197,799	-0.029	0.006	8.82×10^{-6}*	73,113
rs1550094	*PRSS56*	2	233385396	0.033	0.003	9.61×10^{-22}	197,966	0.033	0.006	3.69×10^{-9}*	73,174
rs6929347	*KCNQ5*	6	73629566	-0.029	0.003	1.52×10^{-19}	195,623	-0.023	0.005	2.02×10^{-5}*	72,329
rs3138142	*RDH5*	12	56115585	-0.033	0.004	6.28×10^{-19}	196,730	-0.024	0.006	6.75×10^{-5}*	72,709
2:178827571_GA_G	*PDE11A*	2	178827571	0.027	0.003	9.45×10^{-18}	197,243	0.014	0.005	6.53×10^{-3}	72,923
rs17713847	*MYO5B*	18	47376162	0.037	0.004	3.87×10^{-16}	197,076	0.021	0.007	5.06×10^{-3}	72,870
rs9872571‡	*CTNNB1*	3	41232162	0.026	0.003	6.35×10^{-16}	197,070	0.017	0.005	1.57×10^{-3}	72,835
rs4581716	*RBFOX1*	16	7458135	0.026	0.003	6.51×10^{-16}	197,966	0.018	0.005	6.00×10^{-4}*	73,174
rs62169487	*PABPC1P2*	2	146862268	-0.026	0.003	1.30×10^{-15}	195,662	-0.008	0.005	1.29×10^{-1}	72,285
rs17499593	*SLC25A12*	2	172649755	-0.031	0.004	5.12×10^{-15}	197,792	0.002	0.007	7.82×10^{-1}	73,105
rs3769359	*GPD2*	2	157394443	0.028	0.004	5.54×10^{-15}	197,648	0.012	0.006	3.59×10^{-2}	73,033
rs4942848	*RCBTB1*	13	50141345	-0.027	0.003	8.77×10^{-15}	197,966	-0.015	0.006	9.33×10^{-3}	73,174
rs17010513	*FRMPD2*	10	49403140	0.026	0.004	7.01×10^{-13}	197,966	0.016	0.006	5.74×10^{-3}	73,174
rs2300861	*AKAP6*	14	33294781	-0.023	0.003	8.86×10^{-13}	194,490	-0.010	0.005	5.04×10^{-2}	71,865
rs13380109	*RASGRF1*	15	79378775	0.023	0.003	1.34×10^{-12}	197,774	0.015	0.005	4.88×10^{-3}	73,112
11:84736896_TA_T	*DLG2*	11	84736896	0.023	0.003	2.57×10^{-12}	196,452	0.018	0.005	1.27×10^{-3}	72,644
rs6428600‡	*SNORD3G*	1	91192297	0.022	0.003	4.22×10^{-12}	195,962	0.014	0.005	6.31×10^{-3}	72,367
rs466700	*LMCD1-AS1*	3	8178889	0.021	0.003	5.13×10^{-11}	197,656	0.016	0.005	2.32×10^{-3}	73,078
rs1361062	*PBX1*	1	164175828	0.022	0.003	9.39×10^{-11}	197,289	0.022	0.006	1.54×10^{-4}*	72,927
rs10887262	*RGR*	10	86009171	0.022	0.004	1.51×10^{-10}	196,977	0.009	0.006	1.06×10^{-1}	72,821
rs10740465	*KCNMA1*	10	79101195	0.021	0.003	2.35×10^{-10}	197,177	0.003	0.005	5.50×10^{-1}	72,867
rs338076	*SIX3*	2	45171046	-0.022	0.003	5.77×10^{-10}	195,071	-0.006	0.006	2.85×10^{-1}	72,087
rs11226861	*GRIA4*	11	105705843	0.021	0.003	6.33×10^{-10}	194,245	0.021	0.005	9.65×10^{-5}*	71,822
rs1837645	*LINC00989*	4	80481235	0.024	0.004	8.52×10^{-10}	196,481	0.014	0.006	2.88×10^{-2}	72,619
rs12893484	*BMP4*	14	54414738	0.020	0.003	8.88×10^{-10}	197,340	0.010	0.005	4.88×10^{-2}	72,935
rs2796260	*CD46*	1	207914597	0.021	0.003	9.55×10^{-10}	195,020	0.006	0.006	3.13×10^{-1}	72,045

‡ denotes vQTL that has not been previously identified to be associated with refractive error by *CREAM*.

*denotes vQTL that showed significant replication after accounting for multiple comparisons ($p < 0.05/48 = 1 \times 10^{-3}$).

Abbreviations: *SNP* - single nucleotide polymorphism; *Chr* - chromosome; *BP* - base pair; β - effect size (defined in terms of the change in the variance); *se* - standard error; *P* - p-value; *N* - sample size.

Table 3.3. Summary of 48 independent vQTLs with $p < 5 \times 10^{-8}$ identified using Levene's median test in the discovery sample. Continued.

SNP	*Nearest Gene*	*Chr*	*BP*	$\beta_{discovery}$	$se_{discovery}$	$P_{discovery}$	$N_{discovery}$	$\beta_{replication}$	$se_{replication}$	$P_{replication}$	$N_{replication}$
rs2969185	*SHISA6*	17	11406081	0.020	0.003	1.22×10^{-9}	194,706	0.011	0.005	4.13×10^{-2}	72,078
rs67362351	*BICC1*	10	60306548	0.020	0.003	1.77×10^{-9}	197,384	0.013	0.006	2.04×10^{-2}	72,954
rs2258280‡	*NR5A2*	1	200097648	-0.019	0.003	2.06×10^{-9}	197,422	-0.009	0.005	7.64×10^{-2}	72,974
rs71433443	*ZIC2*	13	100651350	0.019	0.003	2.40×10^{-9}	195,468	0.025	0.005	2.88×10^{-6}*	72,246
rs35654095‡	*LRMDA*	10	78472016	-0.019	0.003	2.82×10^{-9}	196,765	-0.017	0.005	1.06×10^{-3}	72,726
rs1790165	*NTM*	11	131928971	-0.019	0.003	2.98×10^{-9}	197,237	-0.006	0.005	2.23×10^{-1}	72,905
rs2808514	*LINC00862*	1	200343081	0.019	0.003	7.08×10^{-9}	196,889	0.007	0.005	2.16×10^{-1}	72,748
rs985631	*OPCML*	11	132982714	-0.019	0.003	8.79×10^{-9}	194,335	-0.005	0.005	3.16×10^{-1}	71,831
rs4837011‡	*PPP6C*	9	127923014	0.020	0.003	9.63×10^{-9}	196,528	0.014	0.006	1.21×10^{-2}	72,616
rs1550871‡	*PTPN5*	11	18750886	-0.018	0.003	1.50×10^{-8}	197,839	-0.014	0.005	5.81×10^{-3}	73,112
rs1806153‡	*PAX6-AS1*	11	31850105	0.021	0.004	1.53×10^{-8}	197,385	0.015	0.006	1.74×10^{-2}	72,950
rs28570522	*PINX1*	8	10630568	-0.018	0.003	2.22×10^{-8}	194,516	-0.018	0.005	1.05×10^{-3}	71,866
rs9518260	*NALCN-AS1*	13	101632306	0.020	0.004	2.37×10^{-8}	197,171	0.003	0.006	6.14×10^{-1}	72,893
rs75570	*HIVEP3*	1	42248979	-0.018	0.003	3.93×10^{-8}	195,335	-0.003	0.005	5.14×10^{-1}	72,195
rs549903	*NTNG1*	1	108089900	-0.017	0.003	4.17×10^{-8}	197,489	-0.016	0.005	2.25×10^{-3}	73,008
rs524373	*KCNA4*	11	30027074	-0.020	0.004	4.31×10^{-8}	197,651	-0.026	0.006	1.56×10^{-5}*	73,059
rs58004513	*SLC14A2*	18	42887885	-0.024	0.004	4.53×10^{-8}	195,272	0.013	0.007	8.14×10^{-2}	72,137

‡ denotes vQTL that has not been previously identified to be associated with refractive error by *CREAM*.

*denotes vQTL that showed significant replication after accounting for multiple comparisons ($p < 0.05/48 = 1 \times 10^{-3}$).

Abbreviations: *SNP* - single nucleotide polymorphism; *Chr* - chromosome; *BP* - base pair; β - effect size (defined in terms of the change in the variance); *se* - standard error; *P* - p-value; *N* - sample size.

Table 3.4. Summary of 14 independent vQTLs with evidence of dispersion effects with $p < 5 \times 10^{-8}$ identified using *HLM*. Genetic variants are ranked based on their dispersion test *p*-value in the discovery sample. The phenotype analysed in the discovery sample was transformed AOSW-inferred refractive error, and the phenotype analysed in the replication sample was autorefraction-measured refractive error. All results are for the *heteroskedastic linear model* dispersion test.

							Discovery sample			Replication sample		
SNP	*Nearest Gene*	*Chr*	*BP*	*MAF*	*EA*	*NEA*	β *(D)*	*se*	*P*	β *(D)*	*se*	*P*
rs12193446	*LAMA2*	6	129820038	0.1	G	A	-0.139	0.014	6.31×10^{-23}	-0.020	0.013	0.122
rs11602008	*LRRC4C*	11	40149305	0.17	T	A	0.097	0.011	2.07×10^{-17}	0.010	0.010	0.308
rs524952	*GJD2*	15	35005886	0.49	A	T	0.070	0.008	2.25×10^{-16}	-0.008	0.008	0.286
rs3138142	*RDH5*	12	56115585	0.24	T	C	-0.077	0.010	5.19×10^{-15}	0.023	0.009	0.010*
rs10089517	*TOX*	8	60178721	0.35	A	C	-0.068	0.009	2.42×10^{-14}	0.014	0.008	0.078
rs7775087	*KCNQ5*	6	73606783	0.44	G	T	-0.061	0.009	1.34×10^{-12}	-0.015	0.008	0.048*
rs16890057	*ZMAT4*	8	40726582	0.21	A	G	-0.073	0.010	1.88×10^{-12}	-0.003	0.009	0.712
rs4581716	*RBFOX1*	16	7458135	0.38	G	A	0.058	0.009	3.09×10^{-11}	0.000	0.008	0.969
rs1550094	*PRSS56*	2	233385396	0.3	G	A	0.059	0.009	1.61×10^{-10}	0.008	0.008	0.314
rs71433443	*ZIC2*	13	100651350	0.45	T	G	0.050	0.009	5.35×10^{-9}	0.008	0.008	0.309
rs12883788	*AKAP6*	14	33303540	0.46	T	C	-0.048	0.009	2.19×10^{-8}	0.006	0.008	0.401
rs62100213	*ACAA2*	18	47302809	0.16	T	C	0.065	0.012	2.69×10^{-8}	0.002	0.010	0.854
rs467283	*LMCD1-AS1*	3	8182065	0.4	G	A	0.047	0.009	4.05×10^{-8}	0.007	0.008	0.354
rs61840038	*FRMPD2*	10	49406911	0.27	A	T	0.052	0.010	4.46×10^{-8}	-0.004	0.009	0.664

*denotes vQTL that showed nominal replication ($p < 0.05$).

Abbreviations: *SNP* - single nucleotide polymorphism; MAF -minor allele frequency; EA - effect allele; NEA - non-effect allele; β - effect size (defined in terms of the change in dioptres for every additional effect allele); *se* - standard error; *P* - p-value.

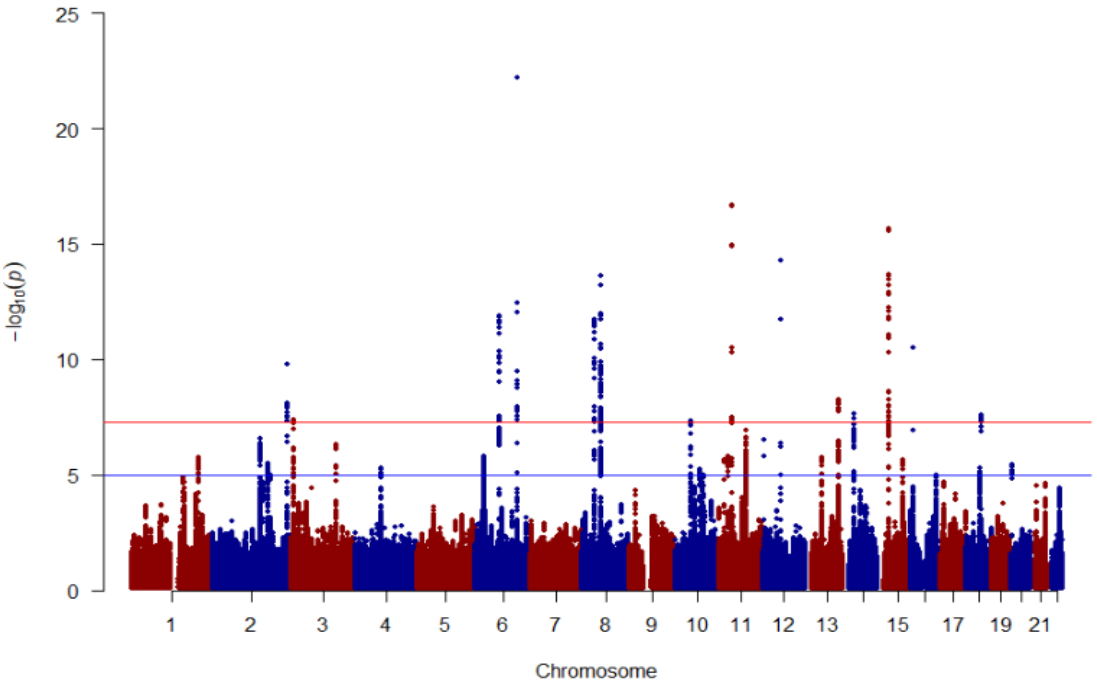


Figure 3.3. Manhattan plot for dispersion effects estimated using the *heteroskedastic linear model* in the discovery sample. The red horizontal line shows genome-wide significant threshold ($p < 5 \times 10^{-8}$), while the blue horizontal line depicts suggestive genomic regions ($p < 1 \times 10^{-6}$).

As discussed above (section 3.2.3.), the mean-variance relationship (r_{av}) could lead to the discovery of variants that are not involved in interactions. Therefore, I compared the extent of overlap between the additive and log-linear variance effects. The scatterplot in Figure 3.4. demonstrates a strong relationship ($r = -0.66$, $p < 2.2 \times 10^{-16}$) between the two components in the discovery sample. However, only a weak relationship was observed for autorefraction-measured refractive error in the replication sample ($r = -0.07$, $p < 2.2 \times 10^{-16}$).

A comparison of genomic regions showing variance heterogeneity with those known from prior work (Tedja et al., 2018) to display marginal association with refractive error revealed a high degree of overlap (Figure 3.5.). Nevertheless, my analysis focusing on variance heterogeneity across SNP genotypes rather than the mean differences identified 7 novel regions not previously associated with refractive error by *CREAM* (Table 3.3.). Of the 7 lead variants at these loci, nominal evidence of replication ($p < 0.05$) was observed for 6, while only one variant had no evidence of replication ($p > 0.05$).

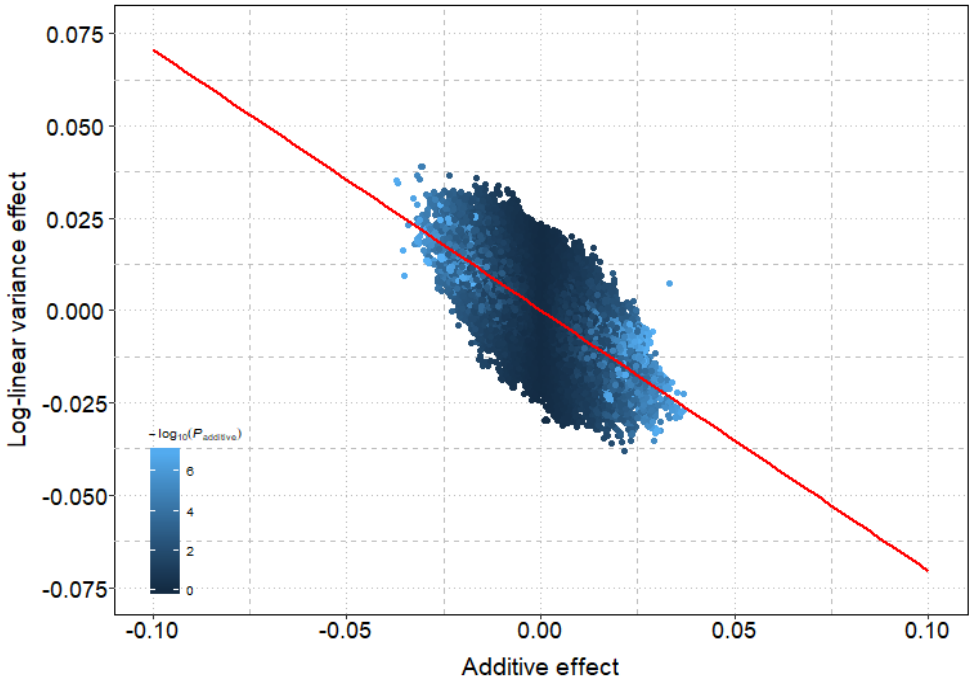


Figure 3.4. Mean-variance relationship between additive and log-linear variance effects estimated by heteroskedastic linear model for the AOSW-inferred refractive error phenotype in the discovery sample. Estimated additive and log-linear variance effects on AOSW-inferred refractive error are plotted for all genome-wide vQTLs, shaded according to the $-\log_{10}(p)$ for the additive effect, up to a maximum of $-\log_{10}(5 \times 10^{-8})$. The red line represents the expected log-linear variance effect given a specific additive effect on AOSW-inferred refractive error.

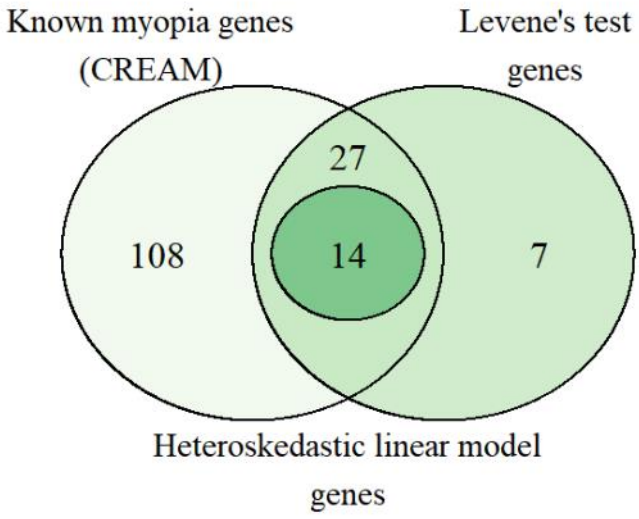


Figure 3.5. Venn diagram illustrating the overlap of refractive error associated genomic regions using different methods. The figure shows common genomic regions discovered by *CREAM* testing for the mean difference across genotypes and Levene's median test/heteroskedastic linear model testing for the variance differences across genotypes.

3.3.4. Gene-based association and gene-set enrichment

Gene-level analysis based on summary statistics from Levene's median test in the discovery sample identified 82 independent genes associated with AOSW-inferred refractive error, of which 8 showed strong evidence of replication (Table 3.5.; Bonferroni corrected threshold $p < 6.1 \times 10^{-4}$), while a further 18 (32% in total) showed at least nominal evidence of replication ($p < 0.05$). The full results for 82 independent genes are presented in Appendix A. The Spearman correlation between gene-based analysis *p*-values in the discovery vs. replication sample for all 18,452 genes examined was 0.08 ($p < 2.2 \times 10^{-6}$). All 8 vQTL genes that showed evidence of replication have been previously found to be associated with refractive error or other related ocular traits such as axial length.

Gene-set enrichment analysis identified '*ionotropic glutamate receptor complex*' as the only *cellular component* gene set to be enriched for genes highlighted by the variance heterogeneity analysis ($p = 1.4 \times 10^{-5}$). Among the 44 genes in the ionotropic glutamate receptor complex gene set, 4 genes were associated with this Gene Ontology domain (*DLG2* $p = 1.1 \times 10^{-13}$; *GRIA4* $p = 1.06 \times 10^{-9}$; *SHISA6* $p = 6 \times 10^{-8}$; *GRIN2A* $p = 3.06 \times 10^{-7}$). In the replication sample, there was minimal evidence for an enrichment of this set ($p = 0.14$), however there was evidence of replication for the top 4 genes (*DLG2* $p = 2.7 \times 10^{-4}$; *GRIA4* $p = 4.2 \times 10^{-4}$; *SHISA6* $p = 0.014$; *GRIN2A* $p = 1.6 \times 10^{-3}$) associated with '*ionotropic glutamate receptor complex*'.

Table 3.5. Genes that showed significant evidence of variance heterogeneity in refractive error in the replication sample after adjustment for multiple comparisons.

GENE	*CHR*	*START BP*	*STOP BP*	$P_{discovery}$	$P_{replication}$	*Associated trait*
LAMA2	6	129154286	129887711	1.02×10^{-36}	9.09×10^{-7}	Refractive error (Tedja et al., 2018)
NPLOC4	17	79473913	79654138	6.58×10^{-29}	1.44×10^{-9}	Strabismus (Plotnikov et al., 2019)
ZMAT4	8	40338109	40805345	4.42×10^{-26}	8.34×10^{-5}	Refractive error (Tedja et al., 2018)
*CHRND**	2	233340870	233451375	1.29×10^{-22}	1.29×10^{-9}	Refractive error (Kiefer et al., 2013)
LRRC4C	11	400085524	41531186	1.89×10^{-15}	2.05×10^{-6}	Refractive error (Tedja et al., 2018)
LRIT1	10	85941276	86051217	2.44×10^{-12}	1.92×10^{-7}	Refractive error (Tedja et al., 2018)
GRIA4	11	105430800	105902819	1.06×10^{-9}	4.25×10^{-4}	Refractive error (Tedja et al., 2018)
*ZIC5**	13	100565275	100674178	1.49×10^{-9}	1.62×10^{-5}	Hypermetropia (Kiefer et al., 2013)

START and *STOP* base-pair positions include 50kb region on either side of the transcribed gene.
**CHRND* and *ZIC5* are located near *PRSS56* and *ZIC2* genes which were identified to be associated with refractive error by Kiefer et al.

The only *molecular function* domain gene set identified in the vQTL gene-based analysis was '*transcriptional activator of RNA polymerase II distal enhancer sequence specific binding*' ($p = 7.2 \times 10^{-6}$). Among the 25 genes included in this gene set, 5 drove the enrichment (*SIX3* $p = 2.8 \times 10^{-8}$; *ZEB2* $p = 1.85 \times 10^{-6}$; *NFE2L2* $p = 5.52 \times 10^{-6}$; *MEF2C* $p = 4.5 \times 10^{-5}$; *SIX6* $p = 7.4 \times 10^{-4}$). As with the *ionotropic glutamate receptor complex* gene set, this set did not show an enrichment in the replication sample ($p = 0.31$), although 3 of the 5 implicated genes displayed suggestive evidence of replication (*SIX3* $p = 0.68$; *ZEB2* $p = 0.07$; *NFE2L2* $p = 0.34$; *MEF2C* $p = 4.5 \times 10^{-5}$; *SIX6* $p = 0.07$). None of the gene sets in the *biological pathway* Gene Ontology domain showed enrichment after correction for multiple testing. In summary, while ionotropic glutamate signalling and RNA polymerase II enhancer binding were highlighted as candidate biological processes enriched for vQTL - and therefore potentially engaged in gene-environment interactions contributing to myopia development - the level of statistical support was far from definitive.

3.3.5. vQTL x education interaction

For the 34 genomic loci identified using Levene's median test as contributing to phenotypic variability in refractive error (i.e. variants with at least nominal evidence of replication), I looked for direct evidence of gene-environment interactions. I used the two variables, *UniEdu* and *EduYears*, as markers of exposure to education. *UniEdu* is a binary variable that indicated whether individuals did or did not hold a university degree, while *EduYears* is a 4-level categorical variable classifying the age at which formal education was completed.

In the discovery sample, 8 of the 34 variants tested (23%) showed evidence of an interaction with *UniEdu* after accounting for multiple comparisons ($p < 1 \times 10^{-3}$), and another 8 variants (46% in total) showed at least nominal evidence of an interaction ($p < 0.05$). Of these, rs12193446 (*LAMA2*) had the strongest influence on the AOSW-inferred refractive error phenotype ($\beta = 0.09$ D per copy of the risk allele). Three of the 8 Bonferroni-significant variants also showed nominal evidence of an interaction with *UniEdu* in the replication sample for the autorefraction-measured phenotype: rs12193446 ($p = 0.008$), rs16890057 ($p = 0.003$) and rs524952 ($p = 0.043$). An additional variant (rs35654095) had strong evidence for an interaction with *UniEdu* in the

replication sample, but not in the discovery sample (p = 2.27 x 10^{-5} and p = 0.464, respectively). Full results are displayed in Table 3.6.

To avoid false positives in the analysis of interaction with *EduYears* exposure, I set the multiple comparison threshold to 4.9 x 10^{-4} (p = 0.05/[34 x 3] in order to take into account the number of vQTLs tested and the 3 *EduYears* categories). In the discovery sample, the number of vQTL x *EduYears* interactions above the level expected by chance was 8 in the group of individuals who completed their education at 21-25 years (Table 3.7.). At all of these loci there was a systematic increase in the effect associated with the myopia-associated risk variant with additional years spent in education. For example, compared to individuals in the lowest *EduYears* category, rs11606250 located near *LRRC4C* gene had an effect size of -0.05 (95% CI -0.08 to -0.02, p = 0.002) for those completing education at age 16 years, -0.06 D (95% CI -0.1 to -0.03, p = 1.9 x 10^{-4}) for those completing education at age 17-20 years, and -0.08 D (95% CI -0.11 to -0.05, p = 3.7 x 10^{-7}) for those completing education at age 21-26. I observed the largest effect size difference across *EduYears* groups to be 0.07 D for rs16890057 (p = 2.27 x 10^{-6}), followed by rs12193446 (p = 2.08 x 10^{-12}) with an effect size difference of 0.06 D (Table 3.7.). Two variants that showed vQTL x *EduYears* interaction, rs28570522 and 2:178827571_GA_G, were not identified when *UniEdu* was used as the environmental exposure.

3.4. Discussion

Testing for variance heterogeneity has been proposed as an indirect strategy for discovering genetic variants that influence quantitative traits via interaction effects. Numerous methods to assess unequal variance have been proposed (Pare et al., 2010; Rönnegård et al., 2010; Shen et al., 2012; Deng et al., 2013; Cao et al., 2014; Cao et al., 2015; Corty et al., 2018; Corty and Valdar, 2018; Young et al., 2018) in the context of both gene-environment and gene-gene interactions (Struchalin et al., 2010; Forsberg and Carlborg, 2017). Here, my focus was the discovery of genetic variants involved in shaping refractive error development that display evidence of variance heterogeneity.

I identified 48 genomic regions contributing to phenotypic variability in refractive error, of which 34 showed at least nominal replication using an independent sample.

A separate genome-wide analysis using the heteroskedastic linear model, performed as a sensitivity analysis, suggested that 14 of these variants were likely to be involved in either gene-environment or gene-gene interaction or both, while the remainder may potentially have been false positives detected by virtue of their mean-variance relationship (Young et al., 2018). A comparison of loci identified by *CREAM* using a standard GWAS model testing for marginal effects and the current Levene's test vQTL analysis showed a remarkably high degree of overlap (Figure 3.5.).

Analysis of variance heterogeneity using Levene's median test identified 7 novel loci not previously identified by *CREAM*, 6 of which showed nominal replication ($p < 0.05$) using autorefraction-measured refractive error. Notably, 2 of these 7 loci have been reported to be associated with other ocular traits. Firstly, after exome sequencing of 20 patients with high myopia, Wan et al. identified *CTNNB1* gene as a key linker node in a functional interaction network (Wan et al., 2018). Furthermore, a *de novo* mutation in the *CTNNB1* gene was identified in a patient presenting with retinal detachment, and lens and vitreous opacities (Li et al., 2017). A form of exudative vitreopathy characterised by abrupt cessation of growth of peripheral capillaries, leading to an avascular peripheral retina has also been linked to this gene (Panagiotou et al., 2017). Secondly, lead variant rs35654095 is situated approximately 140kb downstream of the *LRMDA* gene on chromosome 10. *LRMDA* is one of the genes responsible for oculocutaneous albinism (Kamaraj and Purohit, 2014), which is characterized by reduced or absent ocular pigmentation, decreased visual acuity, macular hypoplasia, optic dysplasia, atypical choroidal vessels, and nystagmus. The connection between refractive error and remaining 5 novel loci is less obvious. For example, *PTPN5* may be responsible for the control of the synaptic plasticity and neuronal cell survival (Pelov et al., 2012). The relationship between the visual system and the central nervous system was mentioned in Chapter 2. Yang et al. suggested that reduced expression of *PTPN5* can lead to calcium

Table 3.6. Summary of SNP x *UniEdu* interaction test results for vQTL variants identified using Levene's median test. The effect (β) quantifies the change in refractive error per copy of the risk allele in those with vs. without a university degree.

					Discovery sample			Replication sample		
SNP	*Nearest Gene*	*MAF*	*EA*	*NEA*	*β (D)*	*se*	*P*	*β (D)*	*se*	*P*
rs12193446	*LAMA2*	0.1	G	A	0.094	0.016	4.89×10^{-9}*	0.128	0.048	7.60×10^{-3}
rs10089517	*TOX*	0.35	A	C	0.044	0.010	1.34×10^{-5}*	0.021	0.030	4.76×10^{-1}
rs16890057	*ZMAT4*	0.2	A	G	0.051	0.012	1.73×10^{-5}*	0.106	0.035	2.57×10^{-3}
rs524952	*GJD2*	0.49	A	T	-0.038	0.010	5.42×10^{-5}*	-0.057	0.028	4.31×10^{-2}
rs12893484	*BMP4*	0.43	G	A	-0.039	0.010	5.89×10^{-5}*	-0.005	0.029	8.68×10^{-1}
rs4581716	*RBFOX1*	0.38	G	A	-0.036	0.010	2.73×10^{-4}*	-0.037	0.029	2.04×10^{-1}
rs11606250	*LRRC4C*	0.17	A	G	-0.042	0.013	8.96×10^{-4}*	-0.047	0.038	2.15×10^{-1}
rs1550094	*PRSS56*	0.3	G	A	-0.034	0.010	9.79×10^{-4}*	-0.047	0.031	1.26×10^{-1}
rs13380109	*RASGRF1*	0.42	A	G	-0.029	0.010	2.42×10^{-3}	-0.023	0.029	4.15×10^{-1}
rs549903	*NTNG1*	0.49	G	A	0.029	0.010	2.49×10^{-3}	0.005	0.028	8.61×10^{-1}
2:178827571_GA_G	*PDE11A*	0.46	G	GA	-0.027	0.010	4.72×10^{-3}	-0.053	0.029	6.56×10^{-2}
rs1837645	*LINC00989*	0.2	C	T	-0.031	0.012	8.17×10^{-3}	-0.010	0.035	7.83×10^{-1}
rs28570522	*PINX1*	0.38	A	G	0.026	0.010	8.78×10^{-3}	0.009	0.029	7.55×10^{-1}
rs67362351	*BICC1*	0.33	A	C	-0.026	0.010	1.14×10^{-2}	-0.037	0.030	2.16×10^{-1}
rs17010513	*SLC14A2*	0.23	C	T	-0.025	0.011	1.78×10^{-2}	-0.024	0.032	4.51×10^{-1}
rs17713847	*MYO5B*	0.15	A	G	-0.029	0.013	3.08×10^{-2}	-0.009	0.040	8.19×10^{-1}
rs9872571	*CTNNB1*	0.46	T	C	-0.018	0.010	6.12×10^{-2}	-0.032	0.028	2.61×10^{-1}
rs2969185	*SHISA6*	0.47	A	C	-0.018	0.010	6.63×10^{-2}	0.009	0.029	7.58×10^{-1}
rs6929347	*KCNQ5*	0.45	A	G	0.017	0.010	8.02×10^{-2}	0.044	0.029	1.23×10^{-1}
rs4942848	*RCBTB1*	0.29	G	A	0.018	0.010	8.07×10^{-2}	-0.031	0.031	3.21×10^{-1}
rs524373	*KCNA4*	0.26	G	A	0.018	0.011	9.80×10^{-2}	0.079	0.032	1.51×10^{-2}
rs6428600	*SNORD3G*	0.45	G	C	-0.015	0.010	1.24×10^{-1}	0.014	0.029	6.36×10^{-1}
11:84736896_TA_T	*DLG2*	0.36	T	TA	-0.013	0.010	1.89×10^{-1}	-0.001	0.030	9.61×10^{-1}
rs11226861	*GRIA4*	0.37	A	T	-0.013	0.010	1.92×10^{-1}	-0.029	0.030	3.37×10^{-1}
rs3138142	*RDH5*	0.24	T	C	0.009	0.011	4.17×10^{-1}	0.098	0.033	2.95×10^{-3}
rs35654095	*LRMDA*	0.44	T	C	0.007	0.010	4.64×10^{-1}	0.121	0.029	2.27×10^{-5}*
rs71433443	*ZIC2*	0.45	T	G	-0.007	0.010	4.73×10^{-1}	0.003	0.029	9.18×10^{-1}
rs3769359	*GPD2*	0.27	T	G	-0.008	0.011	4.79×10^{-1}	-0.021	0.032	5.10×10^{-1}
rs1550871	*PTPN5*	0.47	A	G	0.006	0.009	5.53×10^{-1}	0.034	0.028	2.35×10^{-1}
rs1361062	*PBX1*	0.31	T	A	0.006	0.010	5.86×10^{-1}	0.010	0.031	7.55×10^{-1}
rs466700	*LMCD1-AS1*	0.41	T	G	0.004	0.010	6.68×10^{-1}	-0.019	0.029	5.07×10^{-1}
rs1806153	*PAX6-AS1*	0.23	T	G	0.004	0.011	7.10×10^{-1}	-0.013	0.034	7.10×10^{-1}
rs7405453	*TSPAN10*	0.36	A	G	0.003	0.010	7.67×10^{-1}	-0.043	0.030	1.48×10^{-1}
rs4837011	*PPP6C*	0.31	T	G	0.002	0.010	8.51×10^{-1}	0.024	0.031	4.35×10^{-1}

*denotes vQTL that showed evidence of a genotype x *UniEdu* interaction after accounting for multiple comparisons ($p < 0.05/34 = 1.4 \times 10^{-3}$).

Abbreviations: *SNP* - single nucleotide polymorphism; MAF -minor allele frequency; EA - effect allele; NEA - non-effect allele; β - effect size (defined in terms of the change in dioptres); *se* - standard error; *P* - p-value.

Table 3.7. Summary of *vQTL* x *EduYears* interaction based on 34 vQTLs that showed at least nominal replication using Levene's median test for variance heterogeneity. Interaction effects were estimated in each *EduYears* category with respect to the baseline group of participants who completed their education at the age of 13 to 15 years.

			***EduYears* category**								
			16 years			17-20 years			21-25 years		
SNP	***Nearest Gene***	***Sample***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***
rs6428600	*SNPRD3G*	D	-0.011	0.013	4.00×10^{-1}	-0.014	0.014	3.07×10^{-1}	-0.024	0.012	4.56×10^{-2}
		R	-0.022	0.043	6.06×10^{-1}	-0.024	0.045	5.87×10^{-1}	-0.019	0.039	6.20×10^{-1}
rs549903	*NTNG1*	D	0.009	0.013	5.14×10^{-1}	0.018	0.014	1.97×10^{-1}	0.036	0.012	2.23×10^{-3}
		R	0.010	0.043	8.19×10^{-1}	0.066	0.044	1.35×10^{-1}	0.041	0.038	2.82×10^{-1}
rs1361062	*PBX1*	D	-0.002	0.015	9.05×10^{-1}	0.007	0.015	6.60×10^{-1}	0.001	0.013	9.25×10^{-1}
		R	-0.048	0.047	3.07×10^{-1}	-0.054	0.048	2.57×10^{-1}	0.019	0.042	6.48×10^{-1}
rs3769359	*GPD2*	D	-0.030	0.015	4.36×10^{-2}	-0.029	0.015	6.20×10^{-2}	-0.031	0.013	1.97×10^{-2}
		R	-0.057	0.048	2.35×10^{-1}	0.042	0.049	3.91×10^{-1}	-0.018	0.043	6.70×10^{-1}
2:178827571_GA_G	*PDE11A*	D	-0.004	0.013	7.49×10^{-1}	-0.026	0.014	6.11×10^{-2}	-0.042	0.012	3.96×10^{-4}*
		R	-0.044	0.043	3.07×10^{-1}	-0.079	0.045	7.52×10^{-2}	-0.058	0.039	1.33×10^{-1}
rs1550094	*PRSS56*	D	-0.038	0.015	8.40×10^{-3}	-0.043	0.015	4.45×10^{-3}	-0.060	0.013	3.16×10^{-6}*
		R	0.041	0.047	3.79×10^{-1}	0.020	0.048	6.81×10^{-1}	-0.004	0.041	9.30×10^{-1}
rs466700	*LMCD1-AS1*	D	-0.012	0.014	3.69×10^{-1}	-0.024	0.014	8.25×10^{-2}	-0.016	0.012	1.79×10^{-1}
		R	-0.024	0.044	5.90×10^{-1}	0.000	0.045	9.94×10^{-1}	-0.035	0.039	3.73×10^{-1}
rs9872571	*CTNNB1*	D	-0.017	0.013	2.00×10^{-1}	-0.024	0.014	8.78×10^{-2}	-0.036	0.012	2.59×10^{-3}
		R	-0.100	0.043	1.98×10^{-2}	-0.033	0.044	4.56×10^{-1}	-0.093	0.038	1.55×10^{-2}
rs1837645	*LINC00989*	D	0.001	0.017	9.66×10^{-1}	-0.029	0.017	8.58×10^{-2}	-0.035	0.015	1.74×10^{-2}
		R	-0.056	0.053	2.89×10^{-1}	-0.005	0.054	9.25×10^{-1}	-0.064	0.047	1.73×10^{-1}
rs6929347	*KCNQ5*	D	0.023	0.014	8.73×10^{-2}	0.050	0.014	3.07×10^{-4}*	0.039	0.012	1.20×10^{-3}
		R	-0.016	0.043	7.05×10^{-1}	-0.003	0.045	9.52×10^{-1}	0.043	0.039	2.66×10^{-1}

*denotes vQTL that showed significant replication after accounting for multiple comparisons ($p < 0.05/(3 \times 34) = 4.9 \times 10^{-4}$).

Abbreviations: SNP - single nucleotide polymorphism; Sample (D - discovery; R - replication); β - effect size (defined in terms of the change in dioptres across *EduYears* categories); *se* - standard error; *P* - p-value.

Table 3.7. Summary of *vQTL* x *EduYears* interaction based on 34 vQTLs that showed at least nominal replication using Levene's median test for variance heterogeneity. Continued.

			***EduYears* category**								
			16 years			17-20 years			21-25 years		
SNP	***Nearest Gene***	***Sample***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***
rs12193446	*LAMA2*	D	0.077	0.023	6.00×10^{-4}	0.099	0.023	2.26×10^{-5}*	0.140	0.020	2.08×10^{-12}*
		R	-0.007	0.073	9.20×10^{-1}	0.027	0.074	7.12×10^{-1}	0.170	0.065	8.70×10^{-3}
rs28570522	*PINX1*	D	0.027	0.014	4.78×10^{-2}	0.023	0.014	1.08×10^{-1}	0.051	0.012	3.03×10^{-5}*
		R	0.058	0.045	1.93×10^{-1}	0.085	0.046	6.34×10^{-2}	0.073	0.040	6.50×10^{-2}
rs16890057	*ZMAT4*	D	-0.005	0.017	7.61×10^{-1}	0.042	0.017	1.33×10^{-2}	0.069	0.015	2.27×10^{-6}*
		R	0.033	0.053	5.33×10^{-1}	-0.015	0.054	7.88×10^{-1}	0.092	0.047	5.10×10^{-2}
rs10089517	*TOX*	D	0.020	0.014	1.65×10^{-1}	0.022	0.015	1.30×10^{-1}	0.056	0.012	7.51×10^{-6}*
		R	0.029	0.045	5.20×10^{-1}	0.013	0.046	7.75×10^{-1}	0.044	0.040	2.68×10^{-1}
rs4837011	*PPP6C*	D	-0.015	0.015	3.19×10^{-1}	-0.043	0.015	$4.57E \times 10^{-3}$	-0.023	0.013	7.66×10^{-2}
		R	-0.097	0.047	3.64×10^{-2}	-0.086	0.048	7.28×10^{-2}	-0.052	0.041	2.05×10^{-1}
rs17010513	*FRMPD2*	D	0.000	0.015	9.99×10^{-1}	-0.023	0.016	1.41×10^{-1}	-0.034	0.013	1.05×10^{-2}
		R	0.011	0.049	8.18×10^{-1}	-0.017	0.050	7.38×10^{-1}	-0.060	0.043	1.65×10^{-1}
rs67362351	*BICC1*	D	0.000	0.014	9.77×10^{-1}	-0.025	0.015	8.59×10^{-2}	-0.024	0.013	5.46×10^{-2}
		R	-0.025	0.045	5.90×10^{-1}	-0.009	0.047	8.41×10^{-1}	-0.050	0.040	2.20×10^{-1}
rs1550871	*PTPN5*	D	0.023	0.013	7.90×10^{-2}	0.043	0.014	2.07×10^{-3}	0.025	0.012	3.45×10^{-2}
		R	0.016	0.043	7.03×10^{-1}	0.073	0.044	9.76×10^{-2}	0.063	0.038	1.01×10^{-1}
rs524373	*KCNA4*	D	0.017	0.015	2.53×10^{-1}	0.022	0.016	1.64×10^{-1}	0.034	0.014	1.06×10^{-2}
		R	0.066	0.049	1.78×10^{-1}	0.091	0.050	6.78×10^{-2}	0.138	0.043	1.50×10^{-3}
rs1806153	*PAX6-AS1*	D	-0.007	0.016	6.42×10^{-1}	-0.010	0.016	5.63×10^{-1}	0.002	0.014	9.11×10^{-1}
		R	-0.001	0.051	9.92×10^{-1}	-0.011	0.053	8.36×10^{-1}	-0.005	0.046	9.16×10^{-1}
rs11606250	*LRRC4C*	D	-0.054	0.018	2.50×10^{-3}	-0.069	0.019	1.90×10^{-4}*	-0.080	0.016	3.75×10^{-7}*
		R	-0.036	0.058	5.37×10^{-1}	-0.075	0.059	2.09×10^{-1}	-0.110	0.051	3.15×10^{-2}

*denotes vQTL that showed significant replication after accounting for multiple comparisons ($p < 0.05/(3 \times 34) = 4.9 \times 10^{-4}$).

Abbreviations: SNP - single nucleotide polymorphism; Sample (D - discovery; R - replication); β - effect size (defined in terms of the change in dioptres across *EduYears* categories); *se* - standard error; *P* - p-value.

Table 3.7. Summary of *vQTL* x *EduYears* interaction based on 34 vQTLs that showed at least nominal replication using Levene's median test for variance heterogeneity. Continued.

			***EduYears* category**								
			16 years			17-20 years			21-25 years		
SNP	***Nearest Gene***	***Sample***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***	***β (D)***	***se***	***P***
11:84736896_TA_T	*DLG2*	D	-0.012	0.014	3.75×10^{-1}	-0.034	0.015	1.96×10^{-2}	-0.028	0.012	2.42×10^{-2}
		R	0.004	0.045	9.22×10^{-1}	-0.003	0.046	9.49×10^{-1}	-0.010	0.040	8.01×10^{-1}
rs11226861	*GRIA4*	D	0.003	0.014	8.47×10^{-1}	-0.003	0.014	8.42×10^{-1}	-0.007	0.012	5.44×10^{-1}
		R	-0.030	0.045	5.01×10^{-1}	-0.026	0.046	5.74×10^{-1}	-0.035	0.040	3.84×10^{-1}
rs3138142	*RDH5*	D	0.024	0.016	1.27×10^{-1}	0.027	0.016	8.86×10^{-2}	0.032	0.014	2.14×10^{-2}
		R	-0.005	0.050	9.27×10^{-1}	0.034	0.052	5.14×10^{-1}	0.087	0.045	4.99×10^{-2}
rs4942848	*RCBTB1*	D	0.027	0.015	6.70×10^{-2}	0.014	0.015	3.63×10^{-1}	0.032	0.013	1.37×10^{-2}
		R	-0.066	0.047	1.60×10^{-1}	-0.010	0.049	8.35×10^{-1}	-0.040	0.042	3.44×10^{-1}
rs71433443	*ZIC2*	D	0.006	0.013	6.60×10^{-1}	-0.006	0.014	6.52×10^{-1}	-0.013	0.012	2.64×10^{-1}
		R	0.053	0.043	2.18×10^{-1}	-0.033	0.045	4.59×10^{-1}	0.034	0.039	3.77×10^{-1}
rs12893484	*BMP4*	D	-0.023	0.014	8.50×10^{-2}	-0.002	0.014	9.08×10^{-1}	-0.041	0.012	6.20×10^{-4}
		R	0.019	0.043	6.57×10^{-1}	0.013	0.045	7.70×10^{-1}	-0.005	0.038	9.06×10^{-1}
rs524952	*GJD2*	D	-0.039	0.013	3.70×10^{-3}	-0.057	0.014	3.53×10^{-5}*	-0.063	0.012	1.08×10^{-7}*
		R	-0.024	0.043	5.74×10^{-1}	-0.087	0.044	4.78×10^{-2}	-0.098	0.038	9.90×10^{-3}
rs13380109	*RASGRF1*	D	-0.015	0.014	2.71×10^{-1}	-0.013	0.014	3.36×10^{-1}	-0.041	0.012	5.76×10^{-4}
		R	0.000	0.043	9.95×10^{-1}	0.005	0.045	9.16×10^{-1}	-0.030	0.039	4.42×10^{-1}
rs4581716	*RBFOX1*	D	0.007	0.014	5.98×10^{-1}	-0.020	0.014	1.64×10^{-1}	-0.034	0.012	4.88×10^{-3}
		R	0.016	0.044	7.18×10^{-1}	0.018	0.045	6.85×10^{-1}	-0.019	0.039	6.27×10^{-1}
rs2969185	*SHISA6*	D	-0.022	0.014	1.05×10^{-1}	-0.021	0.014	1.31×10^{-1}	-0.026	0.012	3.20×10^{-2}
		R	0.003	0.043	9.51×10^{-1}	0.005	0.044	9.06×10^{-1}	0.016	0.038	6.83×10^{-1}
rs7405453	*TSPAN10*	D	-0.017	0.014	2.15×10^{-1}	-0.002	0.014	8.71×10^{-1}	-0.007	0.012	5.82×10^{-1}
		R	0.024	0.045	5.99×10^{-1}	0.018	0.046	7.01×10^{-1}	-0.034	0.040	3.97×10^{-1}
rs17713847	*MYO5B*	D	-0.028	0.019	1.42×10^{-1}	-0.018	0.019	3.46×10^{-1}	-0.032	0.017	5.23×10^{-2}
		R	-0.039	0.061	5.25×10^{-1}	-0.086	0.062	1.67×10^{-1}	-0.085	0.054	1.17×10^{-1}
rs35654095	*LRMDA*	D	0.013	0.014	3.20×10^{-1}	0.030	0.014	3.35×10^{-2}	0.025	0.012	3.86×10^{-2}
		R	0.021	0.043	6.31×10^{-1}	0.042	0.044	3.49×10^{-1}	0.109	0.039	4.60×10^{-3}

*denotes vQTL that showed significant replication after accounting for multiple comparisons ($p < 0.05/(3 \times 34) = 4.9 \times 10^{-4}$).

Abbreviations: SNP - single nucleotide polymorphism; Sample (D - discovery; R - replication); β - effect size (defined in terms of the change in dioptres across *EduYears* categories); *se* - standard error; *P* - p-value.

channel expression and a recovery of potassium channels from inactivation, which in turn increases neuronal vulnerability to glutamate toxicity (Yang et al., 2012a). Both Ca^{2+} and K^{+} channels have been shown to be regulated by other refractive error genes (Tedja et al., 2018). Like *PTPN5*, *PPP6C* is a protein phosphatase. Among associated disorders of this gene causes spinocerebellar ataxia type 12*,* a very rare condition characterized by relatively mild cerebellar ataxia and the presence of action tremor (URL: https://www.orpha.net/consor/cgi-bin/index.php, accessed: 29 October 2019). *SNORD3G* is a small nucleolar RNA and *PAX6-AS1* is a long non-coding RNA. Both could be responsible for coordinated control of gene expression. For example, *PAX6-AS1* may regulate expression of *PAX6* gene, which is involved in ocular morphogenesis and is expressed in numerous ocular tissues during development (Yu et al., 2011). Interestingly, *PAX6-AS1* and *NR5A2* are both highly expressed in the pancreas (Buckle et al., 2018; Seitz et al., 2019). The relationship between the eyes and the pancreas is well established (Ilegems et al., 2013; Vessey et al., 2005). For example, pancreatic peptide hormones, such as glucagon, has been shown to play neuromodulatory roles in neurons of the chicken retina (Fischer et al., 2006) and glucagon's activity is influenced by defocus (Vessey et al., 2005). In addition, the connection with the eye problems and the pancreas has been studied in diabetics (Fledelius, 1985).

Gene-level analysis identified 82 genes to be associated with AOSW-inferred refractive error. Approximately 32% of these showed at least nominal evidence of replication ($p < 0.05$). Gene set enrichment analysis identified '*ionotropic glutamate receptor complex*' as a cellular component enriched for genes associated with the AOSW-inferred phenotype. Several recent studies have suggested the involvement of this cellular component in refractive error development (de Souza et al., 2012; Hendriks et al., 2017; Tedja et al., 2018). The molecular function '*transcriptional activator of RNA polymerase II distal enhancer sequence specific binding*' was also implicated, which is a novel finding, albeit with less support in the replication sample. Therefore, the associated vQTL genes in this gene set could be responsible for regulating transcription of one or more downstream myopia-predisposing genes, perhaps in response to visual or other environmental cues.

Refractive error is influenced by environmental factors such as educational attainment (Mountjoy et al., 2018). Therefore, following the 2-step strategy proposed by (Zhang

et al., 2016) including a screening step (e.g. testing for either marginal effects or for variance heterogeneity) followed by a test for interaction, I examined whether direct evidence of genotype x environment interaction could be found for the confirmed vQTL loci. Using university education as an environmental exposure, a total of 3 vQTLs showed convincing evidence of an interaction, with an effect size of approximately 0.1 D (the typical magnitude of effect on the phenotype mean in a standard GWAS for refractive error). A more complex pattern of results was observed when age at the completion of education was examined as an environmental exposure. At 8 loci there was convincing evidence for interaction effects, with – in all cases – a systematic increase in the effect of the myopia-predisposing risk allele with additional time spent in education. These findings provide molecular-level support for the notion that education contributes to myopia development (Mountjoy et al., 2018). However, not all variants that were identified as being vQTL in the genome-wide analysis displayed evidence of a gene-education interaction. This could be at least partly explained by the presence of other interacting factors (either gene-gene or gene-environment) that I did not explicitly consider in this study.

A limitation of using AOSW-inferred refractive error as a phenotype in the discovery sample was discussed in section 2.4. However, in support of this approach, there was a genetic correlation of 0.93 (95% CI 0.88 to 0.97, $p < 2.2 \times 10^{-16}$) between the AOSW-inferred and autorefraction-measured phenotype, when using marginal SNP effects (Chapter 1 section 1.4.1.) and 0.78 (95% CI 0.65 to 0.91, $p = 1.6 \times 10^{-31}$) when using summary statistics obtained by Levene's test. Ultimately, approximately 70% of vQTL identified in the discovery sample displayed at least nominal evidence of variance heterogeneity in the independent replication sample ($p < 2.2 \times 10^{-16}$).

As illustrated in Figure 3.4., the mean-variance relationship for AOSW-inferred refractive error in the discovery sample was strong ($r_{av} = -0.66$). Such a relationship could possibly indicate that genome-wide significant variants controlling phenotype variability are more likely to have stronger marginal SNP effects, as was the case when Levene's test was applied. Therefore, future studies might benefit from carrying out a joint analysis of additive and variance effects (Shen et al., 2012; Cao et al., 2014; Cao et al., 2015; Corty et al., 2018). However, the mean-variance relationship was weaker in the replication sample ($r_{av} = 0.07$), suggesting that $r_{av} = -0.66$ could have been

observed as a consequence of the imprecise inference of AOSW phenotype. Unobserved factors such as each individual's behaviour or lifestyle could have contributed to the poor prediction of this phenotype. For example, unwillingness to wear glasses could have delayed the age of onset of spectacles wear, thus reducing the correlation between AOSW and refractive error. Another source of bias in the prediction of AOSW phenotype could be attributed to demographic differences between discovery and replication samples. This demonstrates that care needs to be taken when comparing the two phenotypes used in this chapter.

While it is true that variance heterogeneity is a hallmark feature of gene-environment or gene-gene interactions, there are other potential causes for unequal variance across genotype classes. Shifts in variance can be caused by incomplete linkage disequilibrium between a causal polymorphism and the tested marker (Ek et al., 2018) or the presence of multiple causal alleles in a region (Forsberg et al., 2015). Here, I was unable to definitively attribute gene-gene or gene-environment interaction as the cause of variance heterogeneity at the majority of the vQTL loci discovered. Further work will be required in larger samples, or in samples with information on refractive error and a wide range of environmental exposures during childhood.

In summary, the analyses carried out in this chapter suggested that genetic variants controlling refractive error variability are likely to be common. Although many of the 34 newly-discovered loci with effects on variance heterogeneity also displayed marginal SNP effects, 7 were not previously reported to be associated with refractive error by *CREAM*. Selecting a subset of variants based on variance heterogeneity across genotypes proved to be a useful strategy to identify loci involved in genotype-education interactions. Finally, the findings imply that a joint test of additive and variance effects may provide higher statistical power to detect refractive error variants than either test in isolation.

Chapter 4

Evidence of widespread gene-environment or gene-gene interactions in myopia development

4.1. Introduction

Since an individual's genotypes are largely fixed during the lifetime, the standard approach of analysing complex traits involves the assumption that a genetic variant has the same effect in all individuals (i.e. the effects are uniform for all people in the population). However, in the case where environmental and genetic risk factors that alter the susceptibility of disease progression exist, this assumption will not be satisfied. Several recent studies have demonstrated that differences in genetic variant effect sizes across individuals can be viewed in terms of gene-environment (GxE) or gene-gene interactions (GxG) (Paré et al., 2010; Beyerlein et al., 2011; Williams, 2012; Abadi et al., 2017). The two vQTL detection methods (Levene's median test and the heteroskedastic linear model) discussed in the previous chapter were designed to detect variance heterogeneity, which is expected from theory to be a hallmark feature of GxE or GxG interactions. A major advantage of methods based on detecting such heterogeneity of effect sizes is that the identity of the environmental or genetic risk factors underlying a GxE or GxG interaction effect does not have to be pre-specified or measured directly.

Motivated by the observed enrichment of GxE among the variants known to be associated with refractive error, the goal of this chapter was to comprehensively assess the extent to which these known susceptibility variants display inter-individual variability of effect sizes across the sample distribution. The results for refractive error variants were compared with those associated with height, a highly polygenic trait with

little or no evidence of a contribution from GxE and GxG interactions (Wood et al., 2014; Abadi et al., 2017).

As an alternative to Levene's test and the heteroskedastic linear model tested previously, I used conditional quantile regression (CQR) (Koenker and Hallock, 2001) for these analyses. CQR has been shown to have greater statistical power in cases were population distribution is asymmetric and when interactions display antagonistic effects (URL: https://macsphere.mcmaster.ca/handle/11375/23291, accessed: 29 October 2019). The critical difference between quantile regression and linear regression is that the former provides multiple estimates of SNP effect sizes depending on how the sample is stratified, while the latter provides a single estimate that summarizes the average SNP effect for the whole sample.

4.2. Methods

4.2.1. Study participants and sample quality control

A sample of 72,985 unrelated, white-British participants with the autorefraction-measured refractive error was studied. The selection of this sample is outlined in Section 1.4.3.

4.2.2. Selection of genetic variants associated with refractive error and height

For refractive error, the selection of genetic variants was restricted to those attaining genome-wide significance threshold ($p < 5 \times 10^{-8}$) in the *CREAM* Consortium and *23andMe* meta-analysis and that replicated ($p < 0.05$) in a UK Biobank sample (Tedja et al., 2018). Of the initial 149 variants, rs74764079, rs73730144 and rs17837871 were excluded from the analysis because their minor allele frequencies (MAF) were less than 5% (3%,1% and 1% respectively). Before the analysis, the risk allele for each SNP was coded as the allele associated with a more negative refractive error.

Similarly, a list of genetic variants associated with height at genome-wide significance was obtained from the *GIANT* Consortium (Wood et al., 2014). Analyses were restricted to the 148 independent genetic variants with the strongest association (i.e. those with the lowest p-values) and MAF > 0.05. The purpose of using height phenotype was to compare and contrast this trait with refractive error.

4.2.3. Statistical analysis

Linear regression was performed to estimate the effects of the 146 refractive error variants assuming that the effect sizes are constant across the full sample distribution. Refractive error was modelled as the outcome variable, while the genotype, age, age-squared, sex and genotyping array were fitted as predictors/covariates.

To estimate SNP effect sizes at different parts of the sample outcome distribution, CQR was performed. I used the *rq* function from the *quantreg* package in R. The same set of covariates as described above was selected when conducting CQR. Standard errors for the CQR estimates were obtained using 10,000 Markov-chain-marginal-bootstraps. In addition, sensitivity analyses were performed with the first ten principal components also included as covariates. Fitting the more complicated model did not change the estimates substantially. Therefore, the rest of the chapter will discuss the results obtained without using principal components as covariates.

To quantitatively assess the extent of effect size heterogeneity, SNP effect estimates and their standard errors were meta-regressed using a mixed-effects model with the *metafor* package in R (Viechtbauer, 2010). The CQR estimated SNP effects at 9 different quantiles (0.1, 0.2, 0.3, …, 0.9; see section 4.2.5. for the rationale explaining the selection of this set of quantiles) were modelled as the dependent variable and the quantile at which these estimates were obtained as the independent variable. The non-linear nature of SNP effects was assessed by including a quantile-squared term in the meta-regression (MR), resulting in the final model: $y = \beta_0 + \beta_1 q + \beta_2 q^2 + e$ (where β_0 is an meta regression (MR) intercept term, β_1 and β_2 represent the linear and quadratic change in SNP effect across quantiles, respectively, q are the quantiles, and e is the residual error). An illustration of the model fitting strategy is outlined in Figure 4.1.

4.2.4. Assessment of type-1 error rate and power using a permutation-based approach

The gold-standard approach of permutation was used to assess the type-1 error rate and statistical power of the CQR-MR model. Two near-identical methods were used in these simulations. The first method tested for an association between the autorefraction-measured refractive error and the genotype of a simulated bi-allelic

SNP with no marginal effect. The second method evaluated type-1 error rate after permuting phenotype values amongst individuals in the sample and testing for an association between the simulated phenotype and the raw (actual) SNP genotypes of participants. Details are given below.

Creating simulated SNPs. Biallelic SNPs with MAF ranging from 0.05 to 0.45 were simulated from a binomial distribution and assigned for all 72,985 participants. Association between the simulated genotype and (actual) refractive error was assessed using the CQR-MR framework, as shown in Figure 4.1. The procedure was repeated 10,000 times, and the type-1 error was calculated as the proportion of SNPs with $p < 0.05$ for each of the three MR coefficients (β_0, β_1 and β_2).

Creating simulated phenotypes. In total, the observed *avMSE* refractive error phenotype was permuted 100 times using all 72,985 individuals in the sample. Using the simulated phenotype, the association with each of the 146 genetic variants was assessed by using CQR-MR framework. The type-1 error rate was calculated as the proportion of SNPs with $p < 0.05$ for each of the MR coefficients (β_0, β_1 and β_2).

The relative improvement in statistical power was assessed by varying the sample size. A random sample of 10,000 to 70,000 individuals, in steps of 10,000, was selected from the full sample of 72,985 participants. In each case, an association between the 146 genetic variants and *avMSE* was examined. This procedure was repeated 20 times. Statistical power was calculated as the proportion of replicates in which the null hypothesis of no association was rejected at $\alpha = 0.05$. As a result of this sampling procedure, power evaluations were based on a total of 149 x 7 x 20 = 20,860 tests. When assessing statistical power and type-1 error rate, the same covariates were included as in the real data analysis.

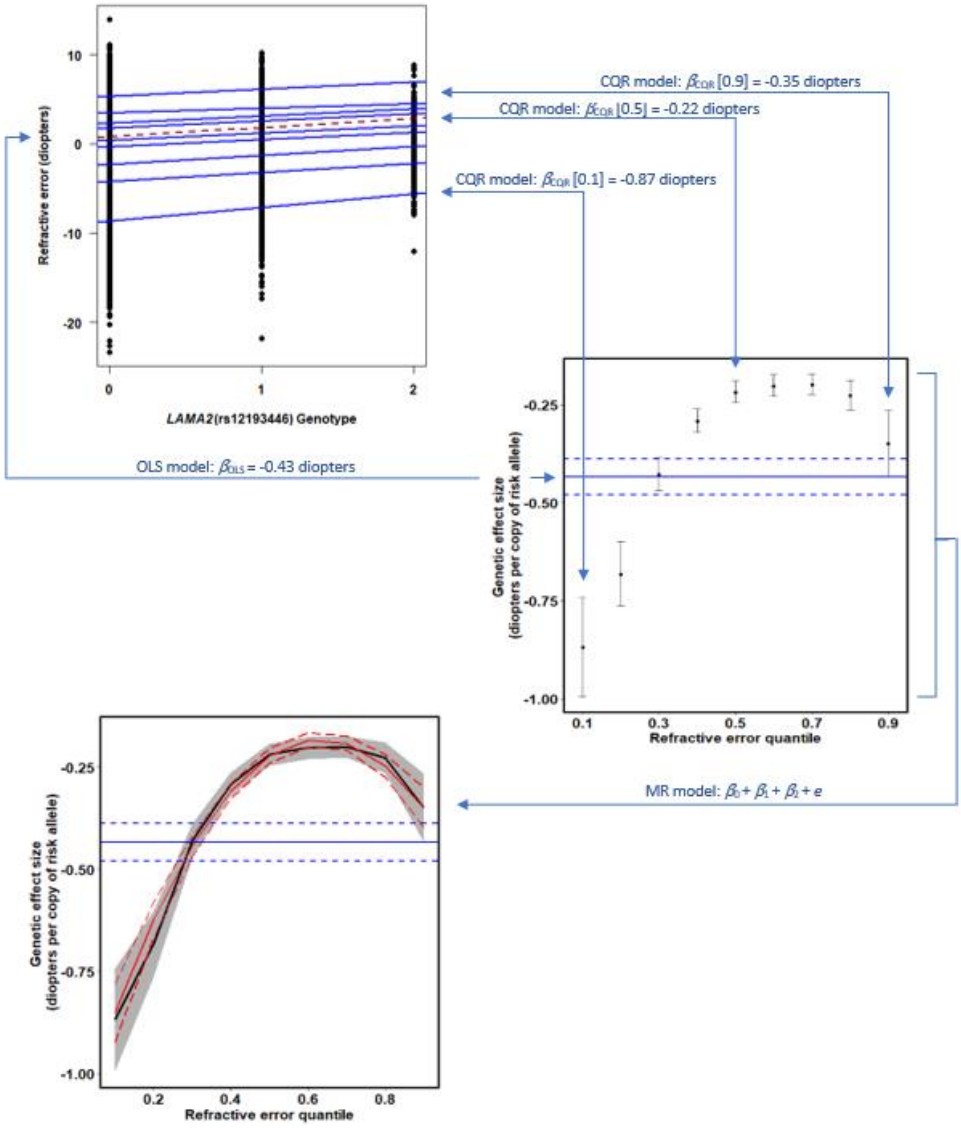


Figure 4.1. Conditional quantile regression (CQR) and meta-regression (MR) analysis framework used to assess the extent of effect size heterogeneity across quantiles of the refractive error distribution. The standard approach of ordinary least-squares (OLS) linear regression, assumes that the SNP effect size does not vary from person to person in the sample under consideration. Therefore, the effect size is estimated by calculating the change in refractive error with respect to the number of minor alleles (dashed red line in the top-left graph). A CQR relaxes the assumption of OLS and allows for SNP effects to vary across the outcome distribution. Similar to OLS, the effect size of a SNP is represented by the slope of the *quantile* regression line (demonstrated in the top-left graph, the nine blue lines correspond to quantile regression fits for a range of quantiles). A hypothetical example for the rs12193446 variant is shown, where the effect size for individuals varies depending on the quantile of the trait distribution that they are found in. The middle-right graph provides a better representation of the differences in the estimated SNP effects (black circles with error bars showing 95% CI). The blue horizontal line in the middle-right graph shows a constant SNP effect estimated by OLS and the dotted lines indicate 95% CI. To quantitatively assess the extent of effect size heterogeneity across different quantiles, an MR can be used (solid red line in the bottom-left graph, with dashed red lines showing 95% CI).

4.2.5. Assessment of the optimal number of quantiles using simulations

The selection of quantiles to be used in CQR-MR analysis is arbitrary. Therefore, additional simulations were performed to find the optimal set of quantiles that could be used to model effect size heterogeneity for refractive error-associated variants. I explored four different sets of quantiles: a) 19 quantiles, q = 0.05 to 0.95 in steps of 0.05; b) 10 quantiles, q = 0.05 to 0.95 in steps of 0.1; c) 9 quantiles, q = 0.1 to 0.9 in steps of 0.1; d) 5 quantiles, q = 0.1 to 0.9 in steps of 0.2. The best set was selected based on type-1 error rate and statistical power, which were calculated in the same manner as described in section 4.2.4. More specifically, the model that showed the lowest rate of inflation of false-positive findings, while maintaining relatively high statistical power was preferred. For simplicity, the rest of the chapter describes different CQR-MR models in terms of the number of quantiles included in the CQR, i.e. 5, 9, 10 or 19.

4.2.6. Correction for the inflation of the false-positive findings

The simulations described in section 4.2.4. revealed inflation of the test statistics for all three MR coefficients. To adjust for the inflation of type-1 error rate, correction factors for each component ($\lambda_{\beta 0}$, $\lambda_{\beta 1}$ and $\lambda_{\beta 2}$) were calculated in a manner similar to genomic control used in GWAS (Devlin and Roeder, 1999). Correction factors for p-values and confidence intervals for each term (β_0, β_1 and β_2) were calculated using the results from the phenotype permutation analyses. The adjustment was achieved with the equation: $\chi^2_{\text{adjusted}} = \chi^2_{\text{observed}} / \lambda$, where λ was calculated as the observed median chi-squared statistic from the simulations divided by the expected median chi-squared statistic with one degree of freedom (Devlin and Roeder, 1999). Given that Z-statistic = β/s.e.$_{unadjusted}$ and X^2-statistic = Z^2, corrected MR confidence intervals were calculated by adjusting standard errors: $s.e._{adjusted} = |\beta / \sqrt{X^2_{adjusted}}|$.

4.2.7. Polygenic risk score effect in different educational attainment strata

The involvement of gene-environment interaction was evaluated by using a polygenic risk score (PRS). The PRS was created by counting the number of risk alleles (0, 1 or 2 for a biallelic SNP) carried by each participant. The PRS did not take into account SNP effect sizes to avoid bias due to using the weights obtained from and applied in UK Biobank. An environmental exposure factor representing 'age completed full-time

education' (*EduYears*) was selected. Individuals with a university degree were not asked the age they completed full-time education. Hence, I assumed that these individuals had completed their education at the age of 21 years. To reduce heterogeneity that arises because some 'age completed full-time education' categories had low counts, adjacent categories were merged, resulting in four final *EduYears* categories: 13-15, 16, 17-20 and 21-26 years.

4.3. Results

4.3.1. Analysis of variants associated with refractive error using ordinary least squares

First, a standard approach of testing SNP effects under the assumption of constant effect sizes was carried out using ordinary least square regression (Appendix B). Among the variants tested, rs12193446 near *LAMA2* gene showed the strongest effect, which was associated with a -0.43 D more negative refractive error (95% CI -0.39 to -0.48, $p = 1.1 \times 10^{-77}$).

4.3.2. Assessment of type-1 error rate and statistical power using simulations

The performance of CQR-MR framework was examined by assessing the type-1 error and power for all three terms estimated in the MR while varying the number of quantiles. The main findings were: first, the CQR-MR model showed systematic inflation of the type-1 error rate for the β_0, β_1 and β_2 components (Appendix C top panel). Second, the MAF did change the type-1 error rate of CQR-MR (Appendix C middle panel). Third, the model that considered 5 quantiles was overly conservative (Appendix C bottom panel). Adjustment for the observed systematic bias is described in section 4.3.3.

Simulations showed that the number of quantiles considered in the CQR-MR model had a substantial contribution towards determining the type-1 error rate. For example, the inflation of false-positive findings for the CQR-MR intercept term (β_0) became progressively worse when the outcome distribution was split into a greater number of quantiles. The type-1 error rate for the model with 19 quantiles was approximately 0.30, while the model with 10 quantiles had a type-1 error rate of 0.16. Inflation of false-positive findings, slightly above the expected ($\alpha = 0.05$; the correct type-1 error

rate), was observed for the model with only 5 quantiles. A similar pattern of inflation was observed for β_1 and β_2 coefficients. The statistical power was very similar across the CQR-MR models that considered 9, 10 or 19 quantiles, while the model with 5 quantiles showed reduced power.

Thus, in summary, simulations provided evidence that MR leads to increased type-1 error rate for the intercept (β_0), linear (β_1) and quadratic components (β_2). The error also depends on the number of quantiles included and is elevated when the number of quantiles considered increases. The statistical power for models with 9, 10 or 19 quantiles was similar. The model that included 9 quantiles (0.1 to 0.9 in steps of 0.1) was considered to be the optimal model and was selected for refractive error and height CQR analyses.

4.3.3. Correction for the inflation of the false-positive findings

As discussed in section 4.3.2., the CQR-MR analysis led to inflated type-1 error rate for the intercept, linear and quadratic components. To correct for this source of bias, p-values and confidence intervals were adjusted using λ coefficients of 1.66, 1.23 and 1.10, respectively (i.e. observed Chi-squared statistics were divided by the relevant λ coefficient, followed by recalculation of confidence intervals and p-values using adjusted Chi-squared statistic). Correction factors were obtained during phenotype simulation.

4.3.4. Widespread evidence of non-uniform refractive error-associated variant effects sizes

The majority of the 146 genetic variants examined had an inverse-U shaped effect size profiles (Figure 4.2. and Appendix D). Individuals near the centre of the distribution (i.e. emmetropic participants) displayed relatively small effect sizes for refractive error-associated variants, while the individuals at the extremes of the distribution displayed greater effects. For example, the strongest effect size heterogeneity was observed for rs12193446 (*LAMA2*). The effect size varied from -0.20 D (95% CI -0.18 to -0.23) in emmetropes to -0.89 D (95% CI -0.71 to -1.07) for highly-myopic participants (Figure 4.1). However, there were exceptions to the inverse-U shape profile. For example, rs1649068 (*BICC1*) and rs9388766 (*L3MBTL3*) showed non-constant, almost linear changes in effect sizes (Appendix D). In addition, SNPs such as rs9680365 (*GRIK1*) and

rs7449443 (*FLJ16171-DRD1*) did not show evidence of effect size heterogeneity (i.e. the estimates across the quantiles were in the range of those expected under the assumption of uniform effect size across the sample distribution) (Figure 4.2. and Appendix D).

4.3.5. Quantitative analysis of non-uniform effects using MR

Meta-regression was used to quantitatively assess the extent of heterogeneity across the refractive error-associated variants. A *Bonferroni* adjusted *p*-value threshold of $0.05/(3 \times 146) = 1.1 \times 10^{-4}$ (where 3 represents the number of components estimated and 146 is the number of SNPs tested) was used to account for multiple comparisons. In total, 66 (45%) of genetic variants displayed significant non-uniform distribution of effects, i.e. $p < 1.1 \times 10^{-4}$ for the β_1 (linear) or β_2 (quadratic) model coefficients (Table 4.1., Appendix D and Appendix E). This suggests that 45% of tested genetic variants could potentially be involved in either gene-gene or gene-environment interaction. On the other hand, 18 (12%) of genetic variants did not show evidence of at least nominally significant effect size heterogeneity (i.e. β_1 component and β_2 component, $p > 0.05$).

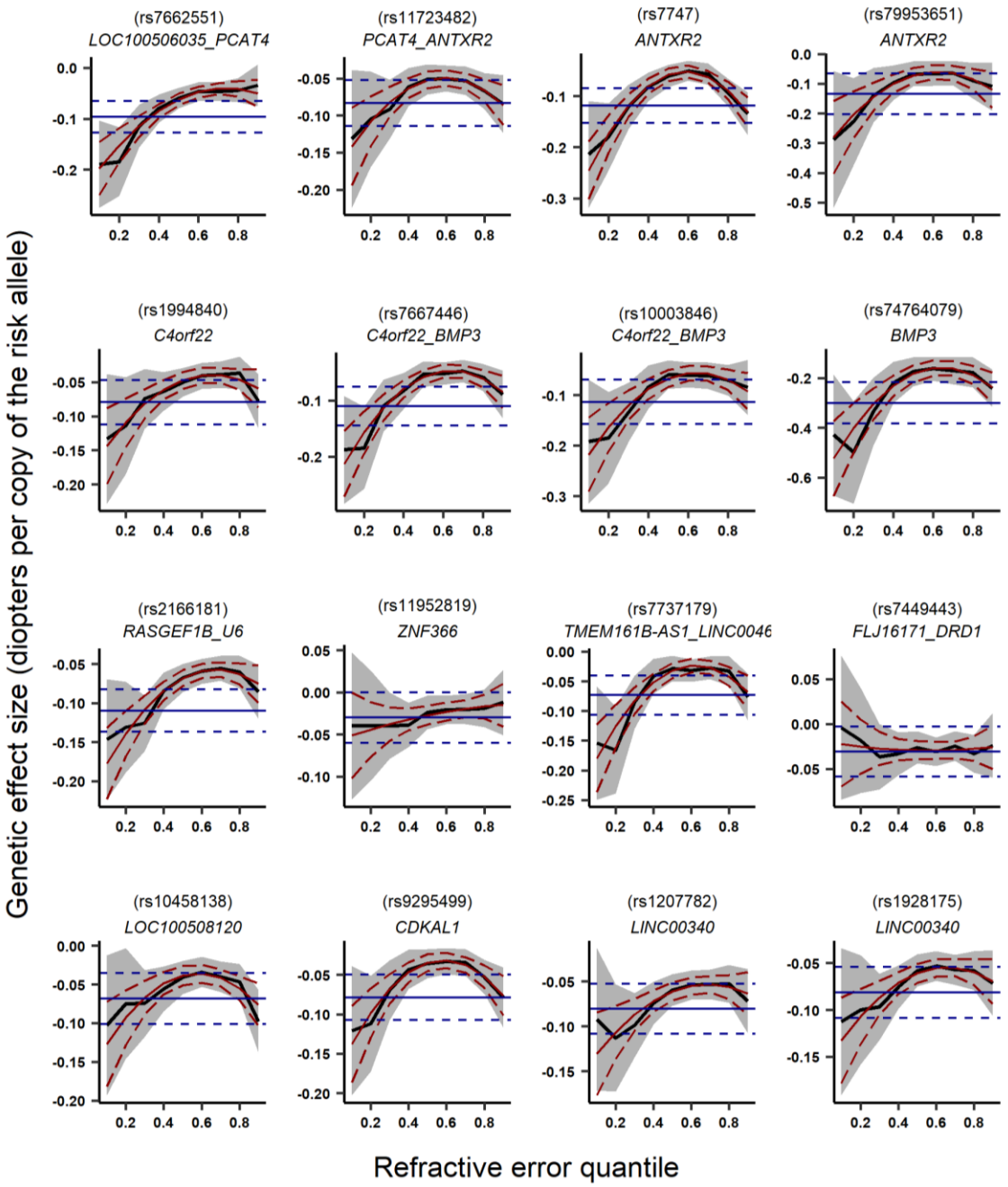


Figure 4.2. Example of genetic effect size heterogeneity for variants associated with refractive error. The solid black line represents genetic effect size estimates obtained using CQR, and the shaded grey region represents their 95% confidence intervals. The solid blue line shows the estimate obtained from standard OLS, and it's corresponding 95% confidence intervals are depicted by the dashed blue lines. The solid red line represents MR estimates, and the dashed red lines correspond to its 95% confidence interval.

Table 4.1. Summary for top 10 genetic variants showing the strongest evidence of association with refractive error according to conditional quantile regression – meta-regression (CQR-MR). Results for confidence intervals and *p*-values are provided after correcting for the inflated type I error rate of MR.

SNP	Gene(s)	β_0 component		β_1 component		β_2 component	
		Beta [95% CI]	*P*	Beta [95% CI]	*P*	Beta [95% CI]	*P*
rs12193446	*BC035400 LAMA2*	-1.130 [-1.272; -0.988]	8.07×10^{-55}	2.995 [2.529; 3.461]	2.12×10^{-36}	-2.363 [-2.765; -1.961]	1.19×10^{-30}
rs524952	*GOLGA8B GJD2*	-0.673 [-0.758; -0.588]	4.83×10^{-54}	1.797 [1.534; 2.06]	7.47×10^{-41}	-1.417 [-1.634; -1.200]	1.68×10^{-37}
rs7744813	*KCNQ5*	-0.543 [-0.631; -0.455]	7.24×10^{-34}	1.402 [1.132; 1.672]	2.15×10^{-24}	-1.092 [-1.314; -0.870]	5.75×10^{-22}
rs11602008	*LRRC4C*	-0.669 [-0.79; -0.548]	2.60×10^{-27}	1.612 [1.250; 1.974]	2.71×10^{-18}	-1.131 [-1.421; -0.841]	2.25×10^{-14}
rs1550094	*PRSS56*	-0.521 [-0.624; -0.418]	4.77×10^{-23}	1.441 [1.118; 1.764]	2.08×10^{-18}	-1.142 [-1.409; -0.875]	4.90×10^{-17}
rs72621438	*SNORA51 CA8*	-0.441 [-0.530; -0.352]	2.06×10^{-22}	1.089 [0.817; 1.361]	4.46×10^{-15}	-0.821 [-1.044; -0.598]	5.85×10^{-13}
rs2326823	*BC035400*	-0.680 [-0.830; -0.530]	6.17×10^{-19}	1.815 [1.341; 2.289]	6.45×10^{-14}	-1.429 [-1.831; -1.027]	3.09×10^{-12}
rs10500355	*RBFOX1*	-0.400 [-0.490; -0.310]	3.63×10^{-18}	1.011 [0.734; 1.288]	8.39×10^{-13}	-0.775 [-1.003; -0.547]	2.76×10^{-11}
rs6495367	*RASGRF1*	-0.374 [-0.459; -0.289]	7.17×10^{-18}	1.009 [0.747; 1.271]	4.38×10^{-14}	-0.833 [-1.049; -0.617]	3.89×10^{-14}
rs2573210	*PRSS56*	-0.501 [-0.621; -0.381]	2.91×10^{-16}	1.414 [1.037; 1.791]	1.94×10^{-13}	-1.121 [-1.434; -0.808]	2.26×10^{-12}

Abbreviations: *SNP*: single nucleotide polymorphism, *CHR*: chromosome, *BP*: base pair, *EA*: effect allele, *CI*: confidence interval, β_0: meta-regression intercept effect size in dioptres per copy of the risk allele, β_1: meta-regression coefficients for the linear term and β_2: meta-regression coefficients for the quadratic term.

4.3.6. Interaction between the polygenic risk score and educational attainment

I observed that the polygenic risk score showed a similar effect size distribution profile across quantiles as that observed for the majority of SNPs individually (Figure 4.3.). Moreover, the effect of the PRS varied across different educational attainment strata. A greater PRS effect size was observed for individuals from the myopic tail of refractive error distribution (quantiles < 0.4), and the effect was exacerbated with additional years spent in education. For example, a one standard deviation increase in PRS was associated with a -0.82 D (95% CI -0.73 to -0.90, $p = 8.9 \times 10^{-83}$) more negative refractive error in the lowest educational stratum (13-15 years) for individuals in refractive error quantile 0.1, while the effect in the highest education stratum was -1.11 D (95% CI -1.02 to -1.18, $p = 1.17 \times 10^{-155}$). The difference in PRS effect across education strata could be as large as 0.57 D (at quantile 0.2).

Similar to the analysis of individual SNPs, the PRS effect was smallest in emmetropes, and the difference associated with educational attainment was within a narrow range of -0.25 to -0.37 D for participants in quantile 0.6, irrespective of their level of education. Individuals in the hyperopic tail of the refractive error distribution (quantiles > 0.8) displayed a reverse relationship compared to myopes. Namely, the

effect of PRS was smaller in participants with greater educational attainment. For example, a one standard deviation reduction in PRS was associated with a +0.62 D (95% CI +0.55 to +0.69) effect in the lowest education stratum, yet only a +0.41 D (95% CI +0.38 to +0.44) effect in the highest education stratum (p = 6.55 x 10^{-68} and p = 9.53 x 10^{-193}, respectively) for hyperopic participants in quantile 0.9.

4.3.7. Quantitative analysis of GIANT Consortium variants associated with height

In addition to performing CQR for genetic variants associated with refractive error, I performed the same set of analyses for height associated SNPs. Given the limited evidence regarding the involvement of GxG and GxE for height, this trait was used to illustrate the difference in effect size distribution. The standard analysis of marginal SNP effect revealed rs143384 located near the *GDF5* gene to have the largest effect size and the strongest association with height (effect size = +0.64 cm per copy of the risk allele, 95% CI 0.57 to 0.70, p = 2.14 x 10^{-80}). In the CQR analysis, the majority of height-associated variants did not show evidence of effect size heterogeneity across quantiles, i.e. in marked contrast to the results observed for refractive error (Appendix F). In cases where there was a deviation from the estimates obtained using OLS, the strongest effect was found at either quantile 0.05 (e.g. *CENPO* variant rs2278483 and *STAU1* variant rs17450430) or quantile 0.95 (e.g. *HHIP* variant rs1812175 and *FAM46A* variant rs310421). Appendix G provides the full results for height variants estimated using OLS, while Appendix H provides MR estimates.

Permutation-based correction to control the type-1 error rate was performed for height. In total, after correction for the inflated type-1 error rate, 53% of the GIANT Consortium genetic variants displayed a marginal association with height (i.e. p < 0.05/(3 x 148) = 3.34 x 10^{-4} for MR intercept β_0). The largest effect was observed for rs143384 located near the *GDF5* gene (β_0 = +0.53cm per copy of the risk allele, 95% CI 0.45 to 0.62, p = 1.14 x 10^{-17}). Not a single height associated variant displayed evidence of a non-uniform effect size across the sample distribution (p > 3.34 x 10^{-4} for β_1 and β_2) (Appendix H).

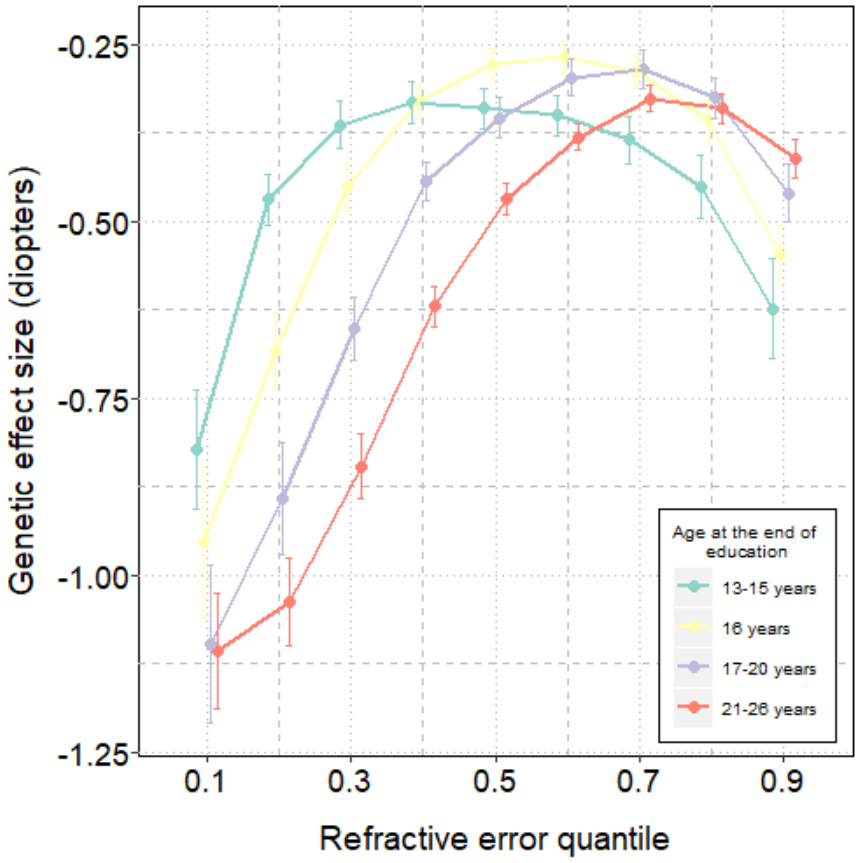


Figure 4.3. The PRS effect size distribution across refractive error quantiles in different educational attainment strata. The difference in PRS effect size depending on the years spent in education is depicted with a coloured line. Error bars represent 95% confidence intervals.

4.4. Discussion

Approximately 88% of the 146 refractive error-associated variants tested in this chapter, showed evidence of effect size heterogeneity. The highly non-linear distribution of genetic effect sizes suggested that in certain individuals SNPs exerted up to 4-fold stronger effects (i.e. in myopes compared to emmetropes). This inter-individual variation is hidden when a conventional approach of assessing SNP effects is used.

Heterogeneity of SNP effects for refractive error-associated variants showed a high degree of non-linearity such that the strongest effects were found at the extremes of the sample outcome distribution, compared to emmetropes (i.e. those around 0.5 quantile). Only a minor fraction of variants did not vary dramatically across quantiles, resulting in estimates that were similar to those expected under the assumption of constant effect across the distribution. I suggest that the process of emmetropization

could be responsible for such findings. Emmetropization maintains a sharp retinal image by controlling the rate of axial eye elongation during infancy via a fine-tuned visual feedback loop (Atkinson et al., 2000). I speculate that it might have a protective role against myopia- or hyperopia-predisposing effects of genetic risk variants. For example, it has been suggested that emmetropization has a limit to the amount of axial elongation that it can compensate for (Mutti et al., 2012). Therefore, for those individuals whose limit is surpassed, genetic risk factors could lead to greater effects. To test the hypothesis that emmetropization is directly responsible for the observed heterogeneity of effect sizes would require studies in animal models. Discovery of genetic determinants leading to susceptibility of visually-induced myopia could be the first step in this direction (Huang et al., 2019).

Previous studies that examined the role of GxE interactions in refractive error found only a few compelling examples (Chen et al., 2011b; Fan et al., 2014; Tkatchenko et al., 2015; Fan et al., 2016). No evidence of GxG interactions influencing changes in refractive error exists. Given the high degree of variance heterogenity observed in this study, a signature that could arise due to the involvement of GxE and GxG, it is likely that interaction effects influencing the change in refractive error are widespread. To show that this feature is not an inherent property of genetic variants, a comparison with the phenotype height was made. Far fewer variants displayed effect size heterogeneity suggestive of GxE or GxE (6% for height vs 88% for refractive error had at least nominal evidence of non-uniform effect sizes). Future studies could focus on the distribution of genetic effect sizes across quantiles for axial eye length, given that both height and axial length display some genetic overlap (genetic correlation r_g = 0.1 to 0.2) (Zhang et al., 2011a; Guggenheim et al., 2013b).

Analysis of PRS effects reinforced the results obtained for the individual SNPs. Highly myopic and highly hyperopic participants were not protected by emmetropization against the genetic risk burden, as was evident from the substantially larger effect sizes in these individuals compared to emmetropes. There was evidence that educational attainment further modified the effect of genetic risk factors on the phenotype. Individuals who spent more time in education displayed larger PRS effects.

The CQR-MR framework used to quantify the difference in genetic effects across individuals builds upon a statistical test used to assess variance heterogeneity across genotypes (Paré et al., 2010; Struchalin et al., 2010; Sun et al., 2013; Zhang et al., 2016). As discussed in Chapter 3, the explanation for this phenomenon is not limited to GxE and GxG interactions and can extend to include other biological mechanisms. For example, parent-of-origin effects, where the effect size of a locus depends on which parent transmitted the effect allele, are expected to contribute towards the increased variance heterogeneity in heterozygous individuals (Struchalin et al., 2010). In addition, variance heterogeneity might arise due to indirect genetic effects, which influence the phenotype through untransmitted alleles in parents (Kong et al., 2018). For example, the risk alleles inherited from the parent could show an interaction with untransmitted parental alleles if the environment to which a child is exposed to is influenced by the parent's genotype. Another situation in which variance heterogeneity can arise is when multiple closely located genotypes are correlated as a consequence of being in linkage disequilibrium with the causal variant (Wood et al., 2014; Forsberg et al., 2015; Ek et al., 2018). This is a well-known problem, which makes the causal inference in GWAS analyses particularly difficult (Wray, 2005; Bush and Moore, 2012). It is also worth bearing in mind that complex traits are influenced by a multitude of environmental and genetic risk factors. Therefore, this could explain why the non-uniform effect size heterogeneity persisted after PRS was stratified according to different educational attainment groups.

In summary, the findings in this chapter suggest that the majority of currently-known refractive error-associated variants exert varying effects in different individuals. A simple explanation for observed inverted-U profile among the examined genetic variants is that they are involved in GxE or GxG interactions. For some variants, effects could vary by as much as 4-fold. There was a stark difference in the magnitude of effect size between high-myopes/high-hyperopes and emmetropes. This inter-individual variation is not captured by using standard methods, thus limiting our understanding of risks associated with developing myopia. Future studies should focus on identifying the driving mechanisms behind such non-uniform distribution of effects across quantiles.

Chapter 5

Prediction of refractive error using gene-environment interactions

5.1. Introduction

One of the aims of quantitative genetics is to make an accurate prediction of phenotypes. To accomplish this, various methods have been developed in the context of animal and plant breeding using raw genotype information (Henderson, 1975; Meuwissen et al., 2001; Gianola et al., 2006; Gianola and van Kaam, 2008; Habier et al., 2011; Gianola, 2013; Hickey et al., 2013; Meuwissen et al., 2014; Morota et al., 2014; MacLeod et al., 2016). Most of these methods have been based on a best linear unbiased prediction (*BLUP*) framework (Henderson, 1975). For example, Gianola et al. described various implementations of Bayesian linear regression that differ in the priors adopted, while otherwise sharing the same sampling model (Gianola, 2013). Typically, the priors reflect one's beliefs about the distribution of SNP effects. However, these methods are computationally intensive. Hence, several improvements that utilize summary statistics from GWAS have been developed (Vilhjálmsson et al., 2015; Robinson et al., 2017; Turley et al., 2018).

Recent studies have demonstrated that improved prediction accuracy could be achieved by a more advanced modelling of the underlying genetic architecture of a trait (Vilhjálmsson et al., 2015; Turley et al., 2018; Chung et al., 2019; Wainschtein et al., 2019). There is evidence that accounting for linkage disequilibrium (*LD*) by incorporating information from a reference sample, considering the impact of rare variants, and analysing the strength of genetic correlation across multiple traits leads to better prediction of the phenotype. However, despite sophisticated statistical modelling strategies, in many situations', prediction accuracy has remained far from being clinically relevant. For example, Vilhjálmsson et al. compared *LD* adjusted

polygenic risk scores to those obtained by using the pruning and thresholding method (*P* + *T*) (Vilhjálmsson et al., 2015). Although accounting for *LD* resulted in 20% more accurate prediction compared to (*P* + *T*) for schizophrenia, the final Nagelkerke prediction R^2 was only 0.25. Similarly, limited prediction accuracy was obtained for multiple sclerosis, breast cancer, and type 2 diabetes phenotypes. Another shortcoming of current genomic prediction methods is that they perform poorly in non-European samples (Martin et al., 2017). Differences in *LD* structure or heterogeneity of causal genetic variants across diverse populations are thought to be responsible for such underperformance of currently available genomic prediction methods.

Dudbridge provided a formula to calculate predictive accuracy of polygenic risk scores based on the assumption that narrow-sense heritability of a trait provides an upper bound estimate of how good a prediction based on genetic information can get (Dudbridge, 2013). However, a recent study suggested that accuracy of prediction could be higher than the estimate of narrow-sense heritability by considering other sources of variation such as indirect genetic effects (e.g. the environment that parents create for their children) (Kong et al., 2018). Some have hypothesized that taking into account gene-environment interaction effects could provide better targeted screening or intervention (Klengel and Binder, 2013; Chatterjee et al., 2016; McAllister et al., 2017). Motivated by this idea, I sought to find out whether a better prediction of refractive error could be achieved by considering gene-environment interaction effects.

5.2. Methods

5.2.1. Sample and SNP quality control

A sample of 72,985 unrelated, white-British participants with autorefraction-measured refractive error were studied. The selection of this sample is outlined in Section 1.4.3.

As the purpose of this study was to determine whether including gene-environment interaction effects would improve genomic prediction, only high-confidence variants were considered. Analyses were restricted to directly genotyped variants with minor allele frequency (MAF) > 0.05, missing genotype rate ≤ 0.01, Hardy-Weinberg

equilibrium < 1 x 10^{-6} and 'rs' variant ID prefix. In total, 308,735 SNPs were retained and selected to test for association with refractive error.

5.2.2. Cross-validation for refractive error prediction

To avoid bias due to overfitting and to obtain reliable prediction estimates, I performed 20-fold cross-validation. In each fold, SNP associations with refractive error measurement were estimated in 95% of individuals (*training sample, N* = 69,336), while refractive error prediction was made in a 5% left-out sample (*testing sample, N* = 3,649).

5.2.3. Genome-wide association analysis for marginal and interaction SNP effects

PLINK (Purcell et al., 2007) was used to estimate SNP effects under an additive genetic model. The following linear regression was fitted to obtain marginal SNP effects:

$$Refractive\ error = SNP + Age + Sex + Array + PC1 + PC2 \dots PC10 \qquad (Eq.1)$$

Where, SNP corresponds to the numeric count of minor alleles carried by a participant (0, 1 or 2) , $Array$ is a binary variable indicating if a participant was genotyped on the UK BiLEVE Axiom array or the UK Biobank Axiom Array (Bycroft et al., 2018) and $PC1 - PC10$ are the first ten principal components.

University degree status was considered as an environmental risk factor for refractive error and was used to estimate gene-environment interaction effects with the following linear regression model:

$$Refractive\ error = SNP + UniEdu + SNP \times UniEdu + \\ + Age + Sex + Array + PC1 + PC2 + \cdots PC10 \qquad (Eq.2)$$

Where, $UniEdu$ corresponds to education coded as *0* if an individual did not have a univ*e*rsity degree and *1* otherwise, and $SNP \times UniEdu$ corresponds to the gene-environment interaction parameter.

5.2.4. Selection of SNPs for construction of polygenic risk scores

Based on the GWAS summary statistics obtained after fitting *Eq. 1* and *Eq. 2*, genetic variants were stratified according to their marginal SNP effect *p*-value threshold: 1 x 10^{-3}, 5 x 10^{-3}, 1 x 10^{-4}, 5 x 10^{-4}, 1 x 10^{-5}, 5 x 10^{-5}, 1 x 10^{-6}, 5 x 10^{-6}, 1 x 10^{-7}, 5 x 10^{-7}, 5 x

10^{-8}. The rationale behind this selection strategy was made because this study aimed to investigate whether the inclusion of interactions in the prediction would improve the existing polygenic risk score approach. The set of independent variants for each of these 11 thresholds was obtained using an LD clumping threshold (r^2) of 0.01 and physical distance threshold of 1Mb in *PLINK* (Purcell et al., 2007). Note that the number of SNPs at each threshold varied across cross-validation folds.

5.2.5. Evaluation of the performance of polygenic interaction scores

The predictive performance of a standard polygenic risk score (*PRS-G*) and a polygenic interaction risk score (*PRS-I*) for refractive error were evaluated. The standard polygenic risk score was calculated with the following formula:

$$Polygenic\ risk\ score\ (PRS-G) = \sum_{i=1}^{M} SNP_i \times Weight_i \qquad (Eq.3)$$

Where, SNP_i corresponds to the numeric count of minor alleles carried by a participant (0, 1 or 2) for the $i = 1, 2, 3, \dots M$ variants and $Weight_i$ corresponds to the effect size of variant i in the training sample. Similarly, I obtained a second polygenic risk score based on the gene-environment interaction effects:

$$Polygenic\ risk\ score\ (PRS-I) = \sum_{i=1}^{M} SNP_iE_i \times GEWeight_i \qquad (Eq.4)$$

Where, $GEWeight_i$ corresponds to the gene-environment interaction effect size of variant i and SNP_iE_i corresponds to the Hadamard product of a SNP matrix and a vector of an environmental factor for each participant as shown below:

$$\begin{bmatrix} SNP_1 & \cdots & SNP_M \\ \vdots & \ddots & \vdots \\ SNP_N & \cdots & SNP_{NM} \end{bmatrix} \circ \begin{bmatrix} Environment_1 \\ \vdots \\ Environment_N \end{bmatrix} = \begin{bmatrix} SNP \times Environment_1 & \cdots & SNP \times Environment_M \\ \vdots & \ddots & \vdots \\ SNP \times Environment_N & \cdots & SNP \times Environment_{NM} \end{bmatrix}$$

Where, N is the number of participants and M is the number of SNPs. I refer to the polygenic risk score described in $(Eq.3)$ as *PRS-G* and polygenic risk score described in $(Eq.4)$ as *PRS-I*. An environmental risk score was calculated in the similar manner:

$$Environmental\ risk\ score = \sum UniEdu \times Weight \qquad (Eq.5)$$

Where, $UniEdu$ corresponds to education coded as *0* if an individual did not have a univ*ers*ity degree and *1* otherwise and $Weight$ corresponds to the estimated effect of education in the training sample. I refer to the environmental risk score described in $(Eq. 5)$ as *ERS*. Therefore, the prediction of refractive error using gene-environment interaction effects is a sum of these three components:

$$Refractive\ error = PRS\text{–}G + ERS + PRS\text{–}I \qquad (Eq. 6)$$

Note that *PRS-G*, *ERS* and *PRS-I* were modelled as independent covariates in *Eq. 6*. As a sensitivity analysis, the performance of the following model was also evaluated:

$$Refractive\ error = PRS\text{–}G + ERS + (PRS\text{–}G \times ERS) \qquad (Eq. 7)$$

I found negligible difference between the two models described by $Eq. 6$ and $Eq. 7$. In practice, it might be less time consuming to use the $Eq. 7$ model because gene-environment interaction effects do not need to be estimated.

To assess the variance explained by a PRS, the improvement in R^2 was assessed for a model including the PRS term(s) relative to a baseline model that included only age and sex as covariates.

5.2.6. Using risk scores to differentiate between myopic and non-myopic individuals

In addition to estimating the amount of phenotypic variance explained by *PRS-I*, I sought to find out whether *PRS-I* is useful in identifying a high-penetrance risk subgroup of individuals. To explore the predictive ability of *PRS-I* in successfully classifying high-myopic vs non-high-myopic individuals, I dichotomized autorefraction-measured refractive error using a classification threshold of ≤ -6.00 D. Only SNPs attaining $p < 5 \times 10^{-4}$ were used to construct a polygenic score because in section 5.2.5 this threshold achieved the best prediction of the phenotype. Predictive performance was assessed using area under the curve (AUC) measurement using the *p*ROC (Robin et al., 2011) package in R. The comparison between the models was assessed using *roc.test* function, specifying 1,000 bootstrap replicates.

5.3. Results

5.3.1. Predictive performance of polygenic scores

PRS-G explained between 4.6% and 6.4% of the phenotypic variance in refractive error over and above that explained by age and sex, depending on the *p*-value threshold chosen for selecting variants. Optimal predictive performance was R^2 = 6.4% (95% CI, 6.1% - 6.7%) at a *p*-value threshold of $p = 5 \times 10^{-4}$ (Figure 5.1.). Thus, in accordance with prior studies (Carey et al., 2016; Mullins et al., 2016), considering more variants at less stringent association thresholds improved phenotype prediction until the noise to signal ratio became too large (e.g. $p = 1 \times 10^{-3}$). In other words, these results demonstrated that a more accurate prediction could be achieved by including not only genome-wide significant variants but less strongly-associated variants, too. In contrast, the amount of phenotypic variance explained by *PRS-I* was low (R^2 = 0.3% to 0.6%) and remained relatively constant regardless of the *p*-value threshold. The highest R^2 observed was 0.63% (95% CI, 0.33% to 0.94%) at $p = 5 \times 10^{-7}$, while at the threshold $p = 5 \times 10^{-4}$ that showed the largest R^2 for *PRS-G, PRS-I* explained only 0.38% (95% CI, 0.17% to 0.61%) of the variance. The environmental predictor, *ERS*, explained approximately 2.6% (95% CI, 2.36% to 2.82%) of the variance. In a model where *PRS-G* and *ERS* were added as two separate covariates, prediction accuracy was increased to 8.42% (95% CI, 8.00% to 8.84%), i.e. close to the level expected assuming their effects were independent (6.4% + 2.6% = 9.0%). Fitting the $Eq. 6$ model provided negligible improvement (R^2 = 8.44%, 95% CI, 8.02% to 8.86%) and the model in $Eq. 7$ showed similarly negligible improvement (R^2 = 8.48%, 95% CI, 8.05% to 8.89%).

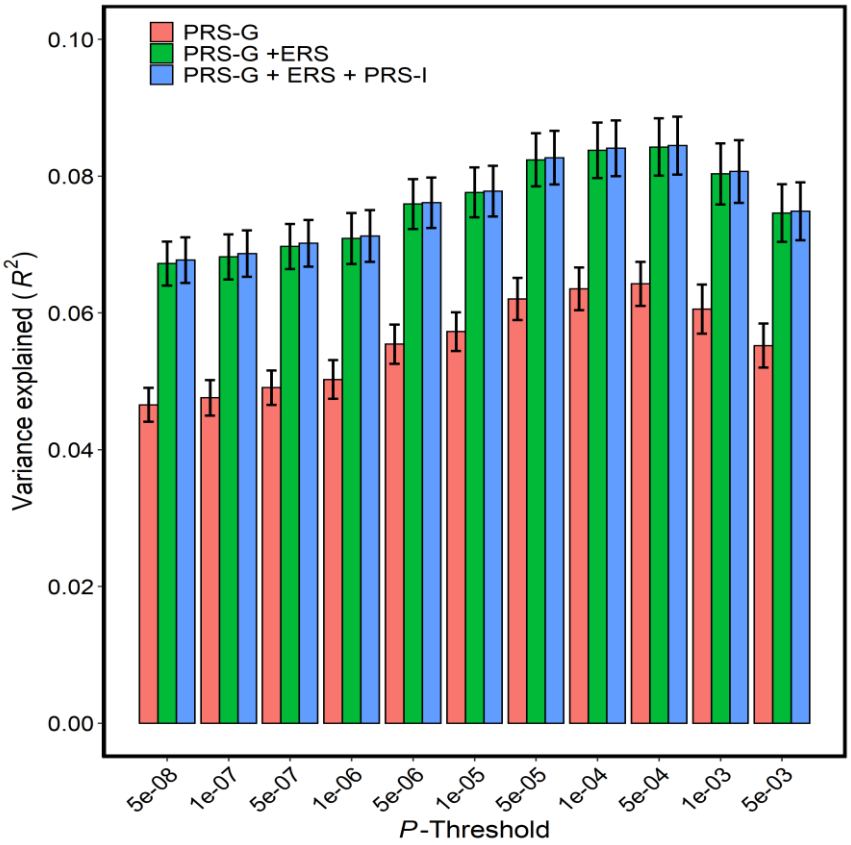


Figure 5.1. Variance explained (R^2) by different polygenic and environmental scores at different refractive error association thresholds. The estimates are averages across 20-folds of cross validation. The error bars represent 95% confidence intervals. Legend labels: *PRS-G* - polygenic marginal score*, PRS-I* - polygenic interaction score, *ERS* - environmental risk score.

5.3.2. Assessment of polygenic risk scores to differentiate between a case-control high-myopia phenotype

The average number of high-myopic (≤ -6.00 D) individuals across cross-validations was 137 ± 13.6. Figure 5.2. shows a representative example AUC from one random cross-validation for different classification models tested. Full results for all cross-validations are presented in Appendix I. The average predictive performance across 20-fold cross-validations of *PRS-G* was 0.68 (95% CI 0.63 to 0.72), which was above the level expected by chance. Inclusion of *ERS* resulted in more accurate classification of high-myopes vs non-high-myopes (AUC 0.71, 95% CI 0.66 to 0.75). A comparison between the two models (*PRS-G* and *PRS-G* + *ERS*) using *roc.test* function showed that in 17 out of 20 cross-validations, addition of *ERS* significantly improved predictive performance. Adding *PRS-I* in the model made, on average, prediction slightly worse (AUC 0.69, 95%

CI 0.64-0.73). Only in 10 out of 20 cross-validation sets inclusion of *PRS-I* showed improved prediction.

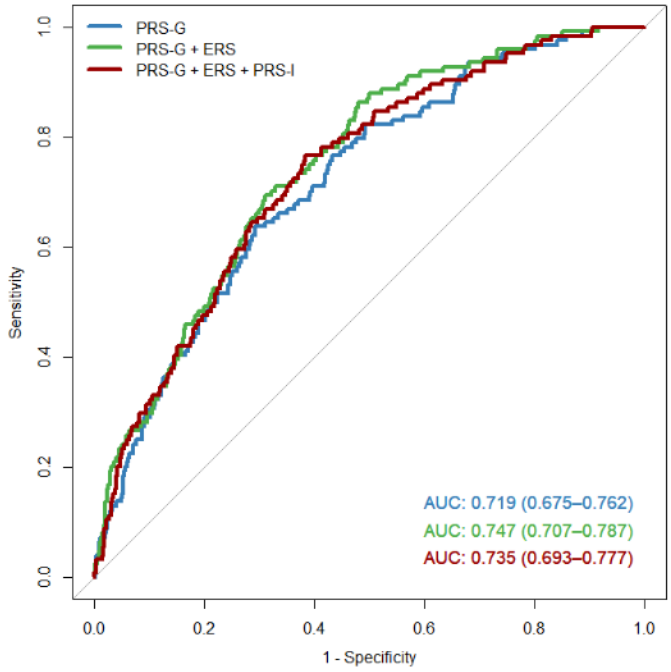


Figure 5.2. Predictive performance of AUC curves for distinguishing between high-myopic and non-high-myopic individuals. An example is given for one of the cross-validations. Numbers in brackets represent 95% confidence intervals. Legend labels: *PRS-G* - polygenic marginal score, *PRS-I* - polygenic interaction score, *ERS* - environmental risk score.

5.4. Discussion

There is immense interest in using genetic and environmental risk factors to improve targeted screening or intervention. Prior research in refractive error prediction focused on utilizing marginal SNP effects (Kiefer et al., 2013; Verhoeven et al., 2013; Guggenheim et al., 2017; Ghorbani Mojarrad et al., 2018; Tedja et al., 2018; Chen et al., 2019). Despite increasing sample sizes leading to more accurate estimation of SNP effects and an increasing number of genetic variants implicated in refractive error development, genetic prediction currently remains too poor to be clinically useful. Given that a causal role of education on refractive error development has been described (Mountjoy et al., 2018), this study aimed to investigate the contribution of gene-environment interaction effects towards the prediction of refractive error. The existing literature lacks a consensus regarding the optimal approach to model the joint

effects of genetic markers and environmental exposures. Here, I focused on a natural extension of the standard polygenic risk score model.

It is common in human genetics to consider several association thresholds when constructing polygenic risk scores (Genetics of personality Consortium, 2015; Carey et al., 2016; Mullins et al., 2016). An explanation for this is that some genetic variants that do not reach genome-wide significance could be informative in making predictions. Accordingly, I found that less stringent thresholds such as $P = 5 \times 10^{-4}$ provided a better prediction of refractive error compared to genome-wide significant SNPs only. The reason for this phenomenon is that current GWAS sample sizes are insufficient to decisively detect all variants that influence a trait (Genetics of personality Consortium, 2015; Carey et al., 2016). Hence, some genetic variants that do not pass the genome-wide significance threshold are nevertheless informative in making predictions.

When the standard polygenic score (*PRS-G*) was used to predict refractive error, it explained up to 6.4% of the trait variance (Figure 5.1.). Other studies have reported similar estimates. For example, 2.6% of refractive error variance was explained in 15 year olds from the ALSPAC cohort (Ghorbani Mojarrad et al., 2018), 4.3% in children from Generation R cohort (Enthoven et al., 2019) and 7.8% was obtained by Consortium for refractive error and myopia (Tedja et al., 2018). In contrast, I observed that a polygenic risk score created using SNP x education effects (*PRS-I*) showed a consistently poor performance irrespective of *p*-value threshold. No more than 0.6% of the variance in refractive error was explained by the *PRS-I*, and when added to *PRS-G* and *ERS* in the model, *PRS-I* did not improve prediction. This suggests that essentially all of the variance of *PRS-I* is already captured by *PRS-G* and *ERS*. This could indicate that future attempts to use a similar approach of incorporating gene-environment interaction effects for prediction will have limited returns as a polygenic risk score constructed using marginal SNP effects will suffice.

Alternative methods have been used previously to study gene-environment interaction effects in prediction. For example, Mullins et al. and other research groups have used the model (*Eq. 7*): *PRS-G* + *ERS* + (*PRS-G* x *ERS*) (Mullins et al., 2016; Abadi et al., 2017; Enthoven et al., 2019), where the direct interaction between an environmental risk exposure and a polygenic risk score is tested without the need to

choose weights for the interaction effects. A comparison of this model (*Eq*. 7) with the empirically determined model (*Eq*. 6) suggested they were similar. Alternatively, Acosta-Pech et al. used an extension of genomic *BLUP* (*GBLUP*) to incorporate interactions in a linear mixed model framework (Acosta-Pech et al., 2017).

One of the goals of studying gene-environment interactions is to identify individuals who are at high risk of developing a severe level of the disease (in case of refractive error - at risk of becoming highly myopic). One step in that direction included a study that investigated the increase in the risk of developing five common diseases, including coronary artery disease and type 2 diabetes (Khera et al., 2018). Their method involved stratification of *PRS-G* into percentiles and examining the change in odds ratios across the stratified *PRS-G* groups. The authors concluded that individuals at the top 0.5% of the distribution were at greater odds of developing a disease compared to other individuals within the top 20% of the distribution. A recent study explored the predictive capability of *PRS-G* x *ERS* for myopia (Enthoven et al., 2019). Weak evidence of discerning between myopic and non-myopic individuals with the help of polygenic and environmental risk score interaction effect was observed. Motivated by this example, I aimed to examine whether the inclusion of *PRS-I* could improve the differentiation between the non-high-myopes and high-myopes in UK Biobank. I found no evidence to suggest that by considering GxE effects improved prediction beyond that using both *PRS-G* and *ERS.* The difference between findings discussed in this chapter and those reported in Enthoven et al. could be due to the sample sizes used. The current study had greater power to detect the predictive ability of GxE. Another reason for differing results could be due to a different classification of myopia status. In the Enthoven et al. study, myopia was defined as ≤ -0.5 D, whereas in my study, high myopia was defined as ≤ -6.00 D because the goal was to see if GxE can help to identify a subgroup of high-risk individuals. However, when I examined the predictive capability of *PRS-I* using ≤ -0.5 D as the phenotype classification threshold, I observed much worse prediction of case-control myopia phenotype: The AUC for PRS-G alone was 0.62 (95% CI 0.6 to 0.64), while the AUC for *PRS-G* + *ERS* + *PRS-I* was 0.64 (95% CI 0.62 to 0.66).

An explanation for why inclusion of *PRS-I* does not lead to an improved prediction could be due to the opposite direction of effect for *PRS-I* compared to *PRS-G*. This is an

example of a qualitative interaction (discussed in Chapter 1 section X.Y.Z.), where two risk factors exert their effects in the opposite direction resulting in a joint effect that is closer to zero (i.e. in the current context, no change in refractive error). However, the current study was underpowered to answer this question definitively.

There are several limitations to using environmental exposures and therefore, gene-environment interaction effects as a way to improve phenotypic prediction. First, for some environmental exposures, it might not be possible to make predictions in individuals whose risk factor exposure profile differs from the reference (model training) sample. In the case of education, for example, my analyses were restricted to a subset of individuals who self-identified as white-British individuals because these individuals were more likely to have experienced similar academic environments. Including individuals from other countries would not have been appropriate because education systems in those countries may differ from that in the UK. Second, lack of information about past or - especially - future environmental risk factor exposure would limit the applicability of such an approach. Furthermore, misclassification or imprecise measurement of exposures could impede on the ability to make predictions. In such situations, more noise could be introduced, leaving *PRS*-G as the best approach.

In summary, these results suggest that including gene-environment interactions is unlikely to lead to a major improvement in prediction of refractive error. This conclusion is in line with previous theoretical work (Aschard et al., 2012a), although new evidence suggests that rare variants could have a substantial contribution towards phenotypic variance influenced by GxE (Kerin and Marchini, 2019). Although *PRS-I* was capable of explaining some of the variation in refractive error, this information was not independent of *PRS-G*, and therefore had limited potential to improve prediction over that obtained using the standard *PRS-G* approach.

Chapter 6

An exploratory analysis of gene-gene interaction involvement in refractive error development using multifactor dimensionality reduction

6.1. Introduction

Multifactor dimensionality reduction (MDR) is a non-parametric approach developed to identify a set of SNPs that best describe the phenotype (Ritchie et al., 2001; Motsinger and Ritchie, 2006). To this day it remains a popular approach for discovering gene-gene interactions (Kaur and Kumari, 2018; Kim et al., 2019; Yang et al., 2019) and has provided valuable biological insight for a wide range of traits including breast cancer (Ritchie et al., 2001), atrial fibrillation (Tsai et al., 2004), myocardial infarction (Coffey et al., 2004), asthma (Su et al., 2012), nicotine dependence (Xu et al., 2014) and coronary artery disease (Hou et al., 2019). The key idea of MDR is to reduce a high-dimensional multi-locus model into two groups – a high-risk group and a low risk group as shown in Figure 6.1. The classification into different groups was initially based on the ratio of affected individuals to unaffected individuals, for example if the ratio exceeded one, the multi-locus combination was assigned a high-risk status and vice versa (Ritchie et al., 2001; Motsinger and Ritchie, 2006). The performance of this binary classification can be evaluated using cross-validation; several metrics such as misclassification and prediction error are calculated and used to select the best multi-locus model.

Since the development of MDR to address classification questions involving high-dimensional datasets such as genotype data, many additional features have been made available. These include extension of the model to accommodate quantitative traits (Lou et al., 2007; Gui et al., 2013; Yu et al., 2016; Jung et al., 2018), survival data (Gui et al., 2011b; Lee et al., 2012; Lee et al., 2015; Lee et al., 2018), family data (Martin

et al., 2006; Lou et al., 2008; Cattaert et al., 2010; Chen et al., 2014) and multiple phenotypes (Choi and Park, 2013; Xu et al., 2014; Kim et al., 2019). Gola et al. provide a comprehensive overview of MDR and MDR-based approaches, along with their strengths and limitations (Gola et al., 2016). Additional features that have made MDR a popular choice included development of methods that eliminate the necessity of selecting arbitrary thresholds to classify SNP combinations into high and low risk groups (Gui et al., 2011a; Jung et al., 2016; Jung et al., 2018) and the development of alternative modes of evaluating hypotheses (Mei et al., 2005; Hua et al., 2010; Winham et al., 2010; Niu et al., 2011; Park and Kim, 2017; Yang et al., 2017).

In this chapter, I focus on the application of the unified model-based multifactor dimensionality reduction (UM-MDR) implementation of MDR (Yu et al., 2016). From a vast range of available MDR-based methods, this implementation of the original algorithm was selected due to several advantages it has over related approaches. First, UM-MDR can accommodate not only non-genetic covariates that describe individuals' ethnicity or age but also adjust the model for marginal SNP effects. Hence, it could be used to identify a combination of SNPs that show evidence of a pure statistical interaction that is not driven by each variants' independent (marginal) effect on refractive error. Second, UM-MDR can provide an estimate of an interaction effect size. Such information could potentially lead to a more detailed prediction of the phenotype in the future or provide a more comprehensive understanding of the importance of gene-gene interactions in refractive error. Third, many existing MDR- based approaches are computationally intensive due to their reliance on permutation to evaluate the significance of the multi-locus model (Pattin et al., 2009; Gola et al., 2016). UM-MDR circumvents this challenge by providing a more efficient semi-parametric correction procedure using a penalized regression framework for every multi-locus combination. This chapter investigated the role of gene-gene interactions in refractive error development, using UM-MDR.

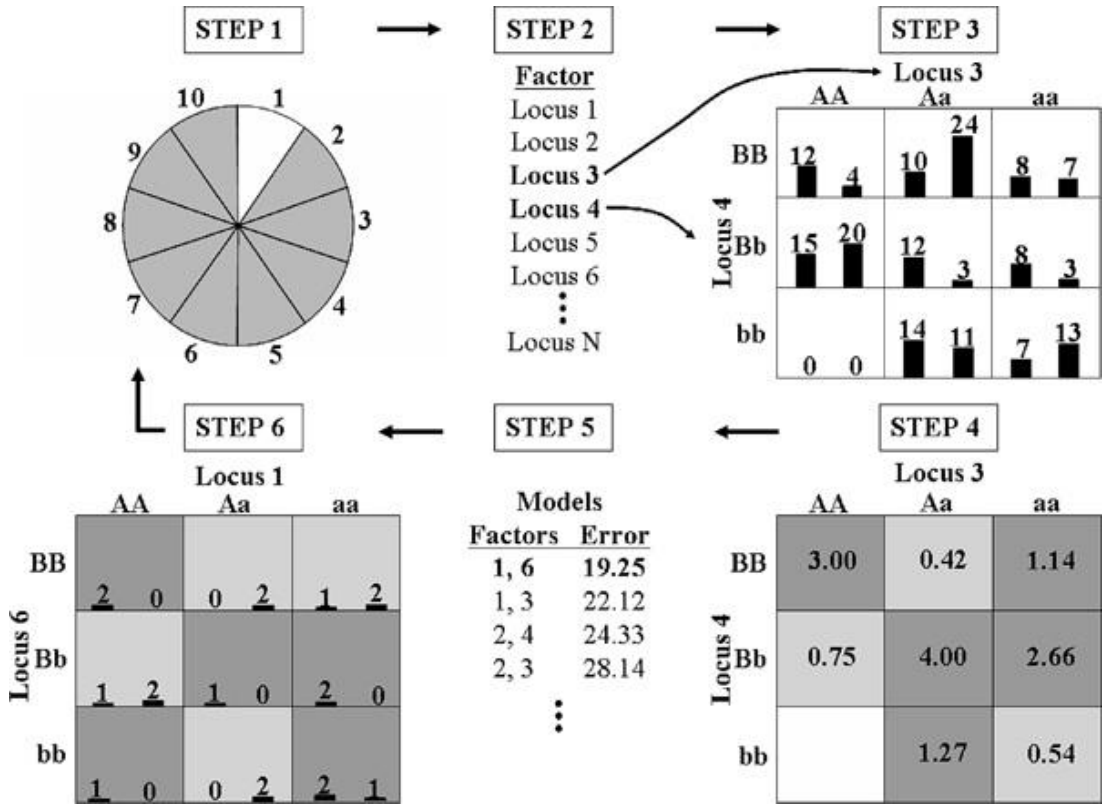


Figure 6.1. Summary of the original MDR algorithm (copied from Ritchie et al., 2003). Step 1 begins by splitting the data into a training and a testing set. One SNP combination from all possible combinations of SNPs is selected in step two and assigned to a low risk or a high-risk category for every possible genotype in step three. In step four, the ratio of affected individuals to unaffected individuals is calculated based on a specific classification rule. In step five, misclassification error of the model is estimated and in step six, the prediction error of the model is estimated using the left-out sample. Steps one through to six are repeated for each possible cross-validation fold. Bars represent hypothetical distributions of cases (left) and controls (right) with each multifactor combination. Dark-shaded cells represent high-risk genotype combinations, whereas light-shaded cells represent low-risk genotype combinations. White cells represent genotype combinations for which no data were observed. Note that this method is not limited for SNP-SNP interactions but can include *n*-dimensional classification, where *n* represents the number of interacting factors.

6.2. Methods

6.2.1. Analysis samples and phenotypes

A discovery sample (N = 197,966) and a non-overlapping replication sample (N = 73,174) of UK Biobank participants were studied. The selection of these two samples is outlined in Section 1.4.2. The optimized model described in Section 1.4.1. was used to derive the "AOSW-inferred refractive error" phenotype for participants in the discovery sample. Autorefraction-measured refractive error was available for participants in the replication sample.

6.2.2. Description of UM-MDR algorithm

A detailed description of UM-MDR algorithm is described in Yu et al. (Yu et al., 2016). In short, the classification step begins by assigning either a low or high-risk category to every multi-locus combination. For a quantitative trait such as AOSW-inferred refractive error, classification depends on the mean phenotype among individuals with a given multi-locus genotype. If this mean value is greater than the mean phenotype of all individuals in the sample, the corresponding multi-locus combination is defined as high-risk or vice versa. Instead of using cross-validation to select the best combination of SNPs, UM-MDR is built under a generalized linear model framework, where the effect of the low or high-risk category is estimated using maximum likelihood. A non-zero effect would suggest an interactive effect of the pair of SNPs on the response variable. Adjustment for marginal SNP effects can be achieved by considering individual SNPs as covariates, and in the presence of strong collinearity between the phenotype and SNPs, interaction effects can be estimated using ridge regression. In the modelling step, the significance of the multi-locus model is evaluated using a semi-parametric correction procedure. A small number of permutations are used to estimate the non-centrality parameter of a chi-square distribution, followed by the recalculation of the *p*-value.

6.2.3. SNP quality control and selection for further analyses

Even though UM-MDR provides one of the most efficient ways to test the significance of multi-locus combinations, doing this on a genome-wide scale remains computationally intractable, currently. Hence, I implemented a stringent SNP filtering strategy before running UM-MDR. First, analyses were restricted to directly genotyped variants with MAF > 0.05, Hardy-Weinberg equilibrium test $p > 1 \times 10^{-6}$ and 100% genotyping rate. After quality control, 345,180 genetic variants were retained and used to run a genome-wide association study in *PLINK* (Purcell et al., 2007). Age, sex, genotyping array and the first ten principal components were used as covariates. A set of independent ("clumped") variants were identified by setting $r^2 = 0.01$ and physical distance of 1Mb using the *PLINK* --clump function (Purcell et al., 2007), resulting in a total of 8,877 independent SNPs with $p < 0.05$. The choice of *p*-value threshold ($p < 5 \times 10^{-8}$) for further analyses was selected as described below.

6.2.4. Selection of optimal p-value threshold

Restricting the UM-MDR analysis to SNPs associated with the trait at genome-wide significance might lead to variants that show an interaction effect but that do not display marginal effects being missed. On the other hand, selecting variants that are nominally significant with the trait ($p < 0.05$) in a GWAS will lead to a higher statistical burden and longer computation times. No universal approach to select the optimal *p*-value threshold for gene-gene interaction testing exists. Hence, to find a suitable balance between true positives and false negatives, I used cross-validation. From the discovery sample, a random subset of 180,000 individuals was selected. Model performance was evaluated using 5-fold cross-validation. The data were split into a training sample (N = 144,000, which was used to discover significant pairs of interacting SNPs) and a testing sample (N = 36,000, which was used to calculate the false discovery rate (FDR)). The FDR was defined as the proportion of variants identified in the training sample that replicated with at least nominal significance level ($p < 0.05$) in the testing sample. The threshold that had the lowest FDR was considered to be the most appropriate for further testing. For each *p*-threshold, I randomly selected 46 SNPs from amongst those showing evidence of association with AOSW-inferred refractive error ($p < 0.05$) in the training sample, so that the FDR was calculated over approximately 1,000 SNP-SNP comparisons (46 x 46 / 2 = 1,058). The rationale behind such random SNP selection was as follows: if less stringent thresholds identify variants that interact in the absence of marginal SNP effects, then the FDR should remain similar to that observed for SNPs attaining the genome-wide significance threshold. Otherwise, selecting a random subset of SNPs would more likely pick up variants that do not interact, hence increasing the FDR.

6.2.5. Statistical analysis

For the optimal *p*-value threshold (see Results), a total of 69 independent genetic variants were associated with the AOSW-inferred refractive error phenotype. These 69 SNPs were analysed by UM-MDR, resulting in 2,346 SNP-SNP pairwise comparisons. As a sensitivity analysis, all comparisons were evaluated using standard linear regression. Associations were considered statistically robust if they passed a Bonferroni corrected threshold ($p < 0.05/2{,}346 = 2.13 \times 10^{-5}$) and were taken forward for testing in the replication sample (for the phenotype, autorefraction-measured refractive error).

Additionally, discovery and replication samples were analysed jointly by using *p*-value based fixed effects meta-analysis in *METAL* software (Willer et al., 2010).

6.2.6. Functional annotation and network analysis

For each variant that showed evidence of involvement in an interaction, the nearest mapped gene was selected and used for gene-based downstream analyses. A list of 14 significant genes identified by UM-MDR was selected for annotation. In order to obtain more detailed insight into the relationship between these genes and the phenotype, publicly-curated databases were used to explore the functional categories and molecular networks by which these interacting genes could be connected. Specifically, *GeneMANIA* (Montojo et al., 2014) was used to find additional genes that could belong to the same pathway or complex, possibly via genetic or physical interactions. To keep the interaction network tractable, analysis was restricted to a maximum of 50 resultant genes. The composite network was created using a query-dependent weighting scheme, where the weights were chosen using linear regression to maximize the number of interactions between the genes from my list. All genes returned by *GeneMANIA* were subject to further analyses, as follows. *PANTHER* (Thomas et al., 2003) was used to identify enrichment of specific biological processes or molecular functions. Overrepresentation analysis was performed using the *REACTOME* pathway database (Fabregat et al., 2018). Within *REACTOME,* non-human identifiers were converted to human equivalents and *IntAct* (Hermjakob et al., 2004) interacting factors, such as proteins, were used to increase the analysis background. *WebGestalt* (Wang et al., 2013) was used to perform a network topology-based analysis (NTA). A network expansion algorithm, where 10 top ranking neighbours were selected to rank all genes in the network based on their proximity to the list of MDR genes, was selected to perform NTA. The advantage of this method over the standard pathway enrichment analysis lies in its ability to model interactions explicitly (Mitrea et al., 2013). Within *WebGestalt, PPI BIOGRID* functional database category was selected to run NTA because the database provides a comprehensive summary for protein and genetic interactions.

6.3. Results

6.3.1. False discovery rate by threshold

Running UM-MDR on a genome-wide scale would pose a substantial computational burden. Figure 6.2. shows a representative example of the time requirements for analysing 2-way or 3-way multi-locus combinations for as many as 21 SNPs using the full discovery sample of AOSW-inferred refractive error (n=197,966). Therefore, under the assumption that SNPs with interaction effects also have main or marginal effects, I attempted to find an optimal main-effect *p*-value threshold that would minimise the number of false-positive findings. While this strategy performed well for the model that did not adjust for marginal SNP effects (Table 6.1.), an FDR could not be reliably estimated for the model where marginal SNP effects were included as covariates. An explanation for this is that the number of true-positive interacting SNP-SNP pairs was scarce in the training datasets (Table 6.2.). For example, at the $p = 1 \times 10^{-5}$ association threshold, an FDR of 1 was calculated in 3 cross-validation folds, while the remaining validation sets contained too few true positives to calculate an FDR reliably. In the scenario where no adjustment for marginal SNP effects was made, a gradually reducing FDR was observed when the association *p*-value threshold became more stringent (Table 6.1.). The lowest FDR of 0.27 (S.E. = 0.039) was estimated for a genome-wide significant threshold ($p = 5 \times 10^{-8}$), and thus only SNPs that passed this threshold were selected for further analyses.

6.3.2. Analysis of SNP-SNP interactions using UM-MDR

In the discovery sample, a total of 69 independent SNPs had marginal evidence of association with the AOSW-inferred refractive error phenotype at the optimal *p*-value threshold of 5×10^{-8}. When these 69 SNPs were assessed for pairwise interactions using UM-MDR in the discovery sample, only one significant SNP-SNP interaction between rs16890054 (*ZMAT4*) and rs5442 (*GNB3*) was identified (Figure 6.3. right panel and Table 6.3.). Both genomic regions were previously known to be involved in myopia development. The interaction exhibited a relatively large effect of 0.1 D (95% CI 0.09 to 0.11, $p = 2.02 \times 10^{-5}$). However, this pair of SNPs did not show evidence of even nominal replication (95% CI -0.09 to 0.21, $p = 0.38$) for the autorefraction measured refractive error phenotype in the replication sample. Furthermore, a sensitivity analysis using linear regression also did not identify a significant interaction

between rs16890054 and rs5442 in either the discovery sample (95% CI -0.01 to 0.05, p = 0.15) or the replication sample (95% CI -0.07 to 0.12, p = 0.62). Indeed, no pair of SNPs was identified as having a significant interaction effect after Bonferroni correction using linear regression in the discovery sample. In general, the distribution of interaction effect sizes obtained with UM-MDR and linear regression differed greatly (Figure 6.4.), with estimates obtained from regression analysis clustered more closely around zero.

6.3.3. Meta-analysis

An additional 13 variants were identified as being involved in SNP-SNP interactions in the combined discovery and replication sample (Table 6.4.). Interestingly,

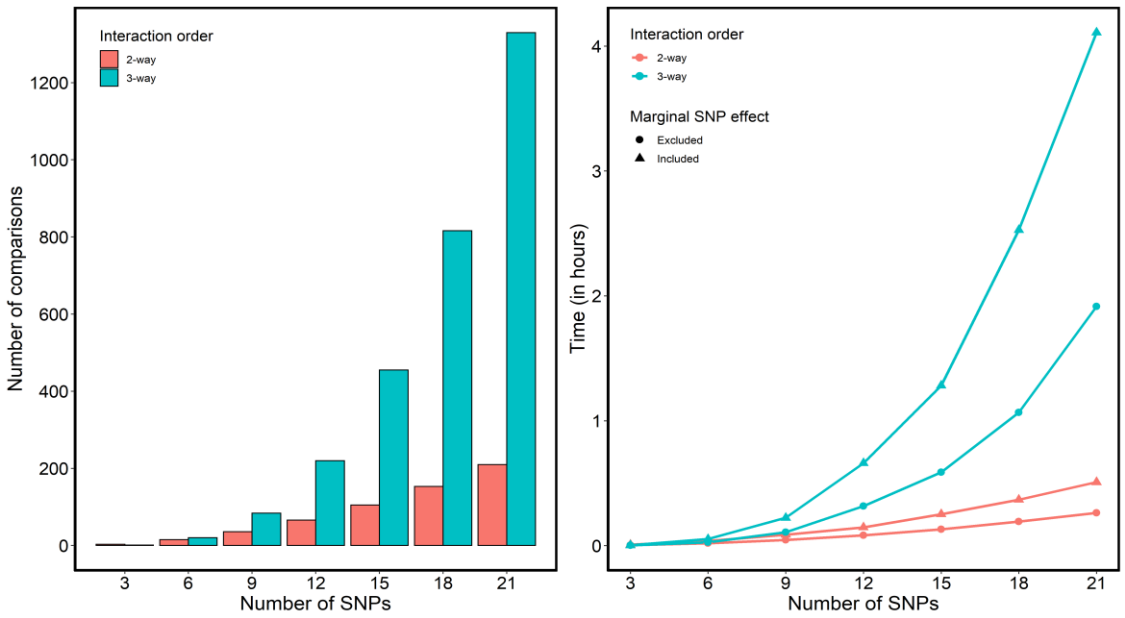


Figure 6.2. Empirical evaluation of time constrains for varying number of gene-gene interactions using UM-MDR. The panel on the left shows the total number of possible SNP-SNP or SNP-SNP-SNP combinations for a small number of genetic variants, while the panel on the right demonstrates the time requirements to perform a UM-MDR analysis for all possible comparisons using the full discovery sample (n=197,966), either adjusting or not-adjusting for marginal SNP effects. Analyses were run on a single processor with 25 GB memory.

Table 6.1. False discovery rate by threshold using the model that did not adjust for marginal SNP effects.

	Association *p*-value threshold											
CV fold	**5 x 10^{-2}**	**1 x 10^{-3}**	**5 x 10^{-3}**	**1 x 10^{-4}**	**5 x 10^{-4}**	**1 x 10^{-5}**	**5 x 10^{-5}**	**1 x 10^{-6}**	**5 x 10^{-6}**	**1 x 10^{-7}**	**5 x 10^{-7}**	**5 x 10^{-8}**
1	NaN	0.866	0.906	0.429	0.726	0.466	0.559	0.303	0.303	0.325	0.247	0.307
2	NaN	0.534	0.428	0.45	0.472	0.396	0.696	0.333	0.327	0.19	0.235	0.16
3	NaN	0.547	0.918	0.62	0.64	0.66	0.509	0.416	0.498	0.376	0.339	0.273
4	NaN	0.583	0.803	0.562	0.721	0.45	0.656	0.52	0.362	0.441	0.41	0.407
5	NaN	0.426	0.649	0.58	0.297	0.528	0.714	0.427	0.403	0.239	0.264	0.264
Mean (se)	NaN NaN	0.575 (0.073)	0.714 (0.091)	0.523 (0.037)	0.542 (0.082)	0.492 (0.045)	0.621 (0.039)	0.392 (0.038)	0.373 (0.034)	0.3 (0.045)	0.292 (0.033)	0.27 (0.039)

*NaN refers to situation where no significant SNP-SNP pair was identified in the training dataset. Abbreviations: CV - cross-validation, se - standard error.

Table 6.2. Total number of significant comparisons at given association threshold.

		Number of SNPs by threshold											
CV fold	**Marginal effect**	**5 x 10^{-2}**	**1 x 10^{-3}**	**5 x 10^{-3}**	**1 x 10^{-4}**	**5 x 10^{-4}**	**1 x 10^{-5}**	**5 x 10^{-5}**	**1 x 10^{-6}**	**5 x 10^{-6}**	**1 x 10^{-7}**	**5 x 10^{-7}**	**5 x 10^{-8}**
1	Excluded	0/0	180/24	75/7	298/170	150/41	574/306	313/138	765/533	623/434	920/621	840/632	956/662
	Included	0/0	0/0	0/0	0/0	0/0	2/0	0/0	0/0	0/0	0/0	0/0	5/1
2	Excluded	0/0	116/54	21/12	280/154	146/77	535/323	316/96	771/514	636/428	881/713	858/656	921/773
	Included	0/0	0/0	0/0	0/0	0/0	5/0	0/0	0/0	2/0	0/0	0/0	1/1
3	Excluded	0/0	168/76	61/5	261/99	175/63	562/191	257/126	797/465	708/355	951/593	916/605	937/681
	Included	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
4	Excluded	0/0	185/77	66/13	318/139	176/49	610/335	349/120	873/419	668/426	970/542	923/544	981/581
	Included	0/0	0/0	0/0	0/0	0/0	4/0	0/0	0/0	0/0	0/0	0/0	1/1
5	Excluded	0/0	115/66	57/20	365/153	111/78	628/296	371/106	831/476	727/434	937/713	903/664	925/680
	Included	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	1/1

Numbers represent significant comparisons in training/testing samples out of a total of 1,035 SNP-SNP comparisons that were evaluated at each *p*-value threshold.

rs16890054 (*ZMAT4*) was present in every single interaction pair, which is more often than would be expected by chance (0.65 out of 13). Following this finding, I confirmed that the distinct peak observed in Figure 6.3. left and right panels was overwhelmingly caused by the interaction between rs16890054 and other genome-wide significant SNPs.

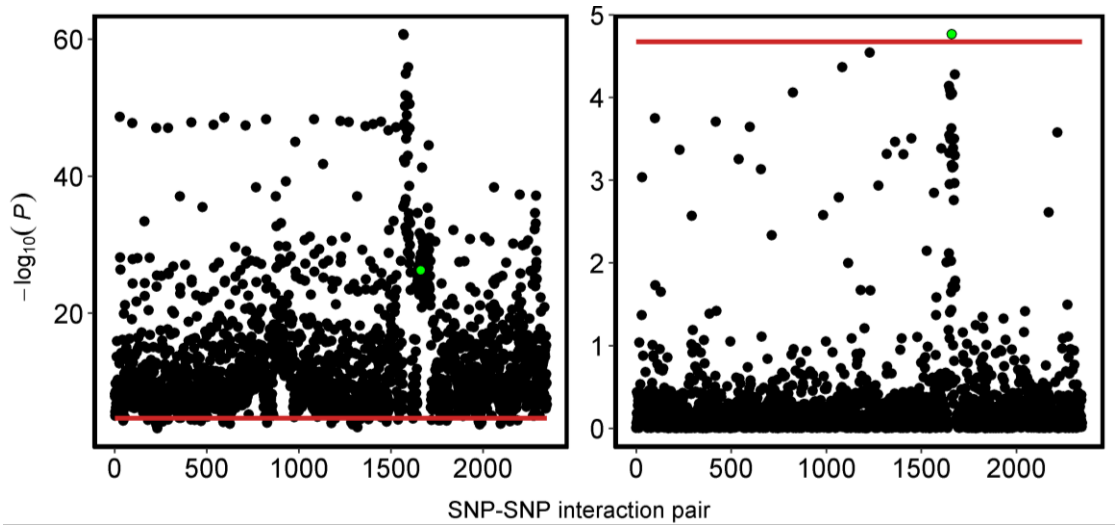


Figure 6.3. Distribution of *p*-values for 69 SNPs with genome-wide significant marginal effects tested for SNP-SNP interaction using UM-MDR. Results are presented for the model where no adjustment for marginal SNP effects was made (left panel) and the model that estimated interaction effects adjusting for marginal SNP effects (right panel). The green dot represents a SNP-SNP pair that showed evidence of an interaction at Bonferroni corrected threshold (p = 0.05/2346, 2.13 x 10^{-5}) in the model accounting for marginal effects. The distribution of *p*-values is arranged on the x-axis arbitrarily depending on the order in which SNP-SNP comparisons were analysed and not the physical position in the genome.

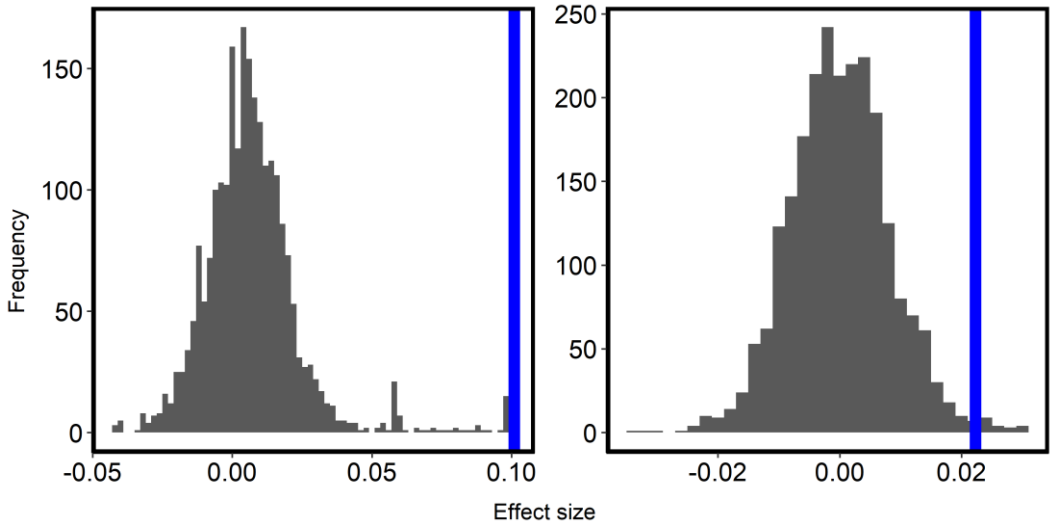


Figure 6.4. Distribution of SNP-SNP interaction effect sizes. Effect sizes estimated using the UM-MDR method (accounting for marginal effects) are shown in the left panel, while those estimated with linear regression are shown in the right panel. The vertical blue line represents the effect size for the rs16890054 vs. rs5442 interaction.

Table 6.3. Summary of interaction between rs16890054 and rs5442. Interactions were assessed with UM-MDR.

						Excluding marginal effects		Including marginal effects	
Sample	**Refractive error (D) Mean (SD)**	**Nearest genes**	**BP**	**CHR**	**Effect allele**	**BETA (SE)**	**P-value**	**BETA (SE)**	**P-value**
Discovery	-0.36 (1.51)	*ZMAT4, GNB3*	40723038, 6954864	8,12	C, A	0.08 (0.0007)	5×10^{-27}	0.1 (0.005)	2.02×10^{-5}
Replication	-0.25 (2.66)	*ZMAT4, GNB3*	40723038, 6954864	8,12	C, A	0.2 (0.003)	9.6×10^{-17}	0.06 (0.08)	0.38
Combined	-	*ZMAT4, GNB3*	40723038, 6954864	8,12	C, A	-	2.3×10^{-42}	-	3.9×10^{-5}

Abbreviations: *BP* - base pair, *CHR* - chromosome, *D* - dioptre, *SD* - standard deviation, *BETA* - effect size, *SE* - standard error, *P* - *p*-value. Combined refers to the pooled discovery and replication samples. Excluding marginal effects refers to the model that did not account for marginal SNP effects.

Table 6.4. Additional SNP-SNP interaction pairs identified as significant in the combined sample of AOSW-inferred and autorefraction-measured refractive error phenotypes. Interactions were assessed with UM-MDR.

					Excluding marginal effects	Including marginal effects
SNP-SNP combinations	**Nearest genes**	**BP**	**CHR**	**Effect allele**	**P-value**	**P-value**
rs16890054, rs7013609	*ZMAT4*, *SNTB1*	40723038, 121593916	8, 8	C, T	1.1×10^{-37}	9.3×10^{-7}
rs35667547, rs16890054	*ADAMTS9*, *ZMAT4*	64547477, 40723038	3, 8	C, C	4.1×10^{-39}	1.5×10^{-6}
rs75698317, rs16890054	*BARHL2*, *ZMAT4*	91181968, 40723038	1, 8	T, C	2.1×10^{-37}	2.7×10^{-6}
rs16890054, rs57324368	*ZMAT4*, *RGR*	40723038, 86014873	8, 10	C, G	2.6×10^{-39}	3.5×10^{-6}
rs1353386, rs16890054	*BMP3*, *ZMAT4*	81947080, 40723038	4, 8	A, C	7.4×10^{-45}	4.1×10^{-6}
rs16890054, rs11145204	*ZMAT4*, *FXN*	40723038, 71714067	8, 9	C, C	6.8×10^{-40}	4.9×10^{-6}
rs16890054, rs10842971	*ZMAT4*, *PZP*	40723038, 9303296	8, 12	C, T	5.9×10^{-36}	6.6×10^{-6}
rs1028308, rs16890054	*HIST1H2BK*, *ZMAT4*	27129757, 40723038	6, 8	A, C	8.3×10^{-37}	6.9×10^{-6}
rs36003362, rs16890054	*ZNF281*, *ZMAT4*	200367088, 40723038	1, 8	G, C	6.2×10^{-37}	6.9×10^{-6}
rs475774, rs16890054	*PLXNA2*, *ZMAT4*	208158787, 40723038	1, 8	T, C	1.7×10^{-37}	7.9×10^{-6}
rs16890054, rs12965607	*ZMAT4*, *MYO5B*	40723038, 47391025	8, 18	C, G	1.6×10^{-45}	1.1×10^{-5}
rs16890054, rs2856250	*ZMAT4*, *GRAMD1B*	40723038, 123418749	8, 11	C, A	1.1×10^{-36}	1.7×10^{-5}
rs16890054, rs17010513	*ZMAT4*, *FRMPD2*	40723038, 49403140	8, 10	C, C	8.1×10^{-37}	1.9×10^{-5}

Abbreviations: *BP* - base pair, *CHR* - chromosome. Excluding marginal effects refers to the model that did not account for marginal SNP effects.

6.3.4. Functional annotation of discovered SNP-SNP interactions

A *GeneMANIA* analysis was undertaken to identify additional potential interactions partners for the 14 genes identified with UM-MDR in the combined discovery and replication samples. In total, 307 connections were established between UM-MDR identified genes and partner genes, including the partner genes *RDH5*, *NRP1*, *FNBP4*, *GNG7*, and *ACTN1* (Figure 6.5.). Only *GRAMD1B* did not have any identifiable interacting members. The vast majority of these links (96.0%) were established based on previous evidence of physical protein-protein interactions, with only a small fraction of these connections (3.9%) attributed to genetic interactions, which could be a signature of functional association between two genes.

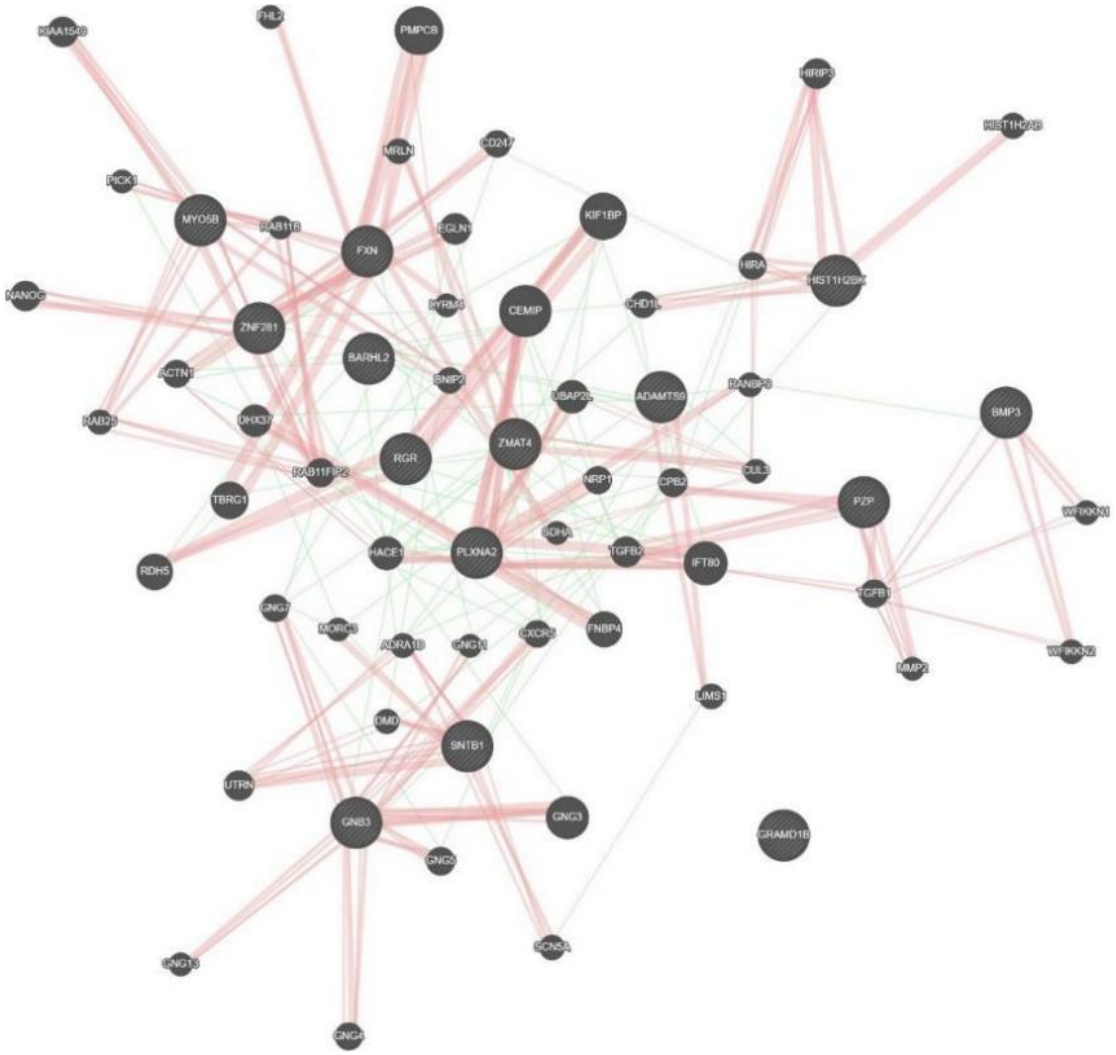


Figure 6.5. Gene-gene interaction network constructed by *GeneMANIA*. The genes indicated with stripes were identified by UM-MDR. Orange connections represent physical interactions. Two genes were linked if they were found to interact in a protein-protein interaction study. Green connections show genetic interactions. Two genes were functionally associated if the effects perturbing one gene were found to be modified by perturbations to a second gene.

The gene list returned by *GeneMANIA* was selected for functional annotation in *Panther*. Regulation of cellular processes (GO:0009987) was identified as a predominant gene ontology category, in which 25 genes were involved in diverse mechanisms such as control of growth, division and maturation of various cell types (e.g. *BMP3* and *TGFB1*), modulation of G-protein (e.g. *PICK1* and *ADRA1D*) and signal transduction (*GNB3*). Approximately half of the genes (26 out of 64) were implicated in binding activity (GO:0005488), including binding F actin filaments (*MYO5B*) and binding DNA as transcription factors (e.g. *ZNF281* and *BARHL2*).

An overrepresentation analysis, which uses a hypergeometric distribution to test whether a list of genes is enriched for pathway X above the level expected by chance, identified presynaptic function kainate receptors as the most enriched pathway (p = 1.3 x 10^{-11}). Table 6.5. shows a summary of the top-ranking pathways returned by *REACTOME*. Overall, the pathways identified were mostly involved in regulation of G-protein related events, signal transduction and transport of metabolites.

Table 6.5. Summary for top ranking *REACTOME* pathways.

Pathway Name	Total number of genes in pathway	Number of genes from *GeneMANIA* list	*p*-value	*FDR*
Presynaptic function of Kainate receptors	23	7	1.3 x 10^{-11}	9.3 x 10^{-9}
Activation of kainite receptors upon glutamate binding	37	7	3.4 x 10^{-10}	9.5 x 10^{-8}
Vasopressin regulates renal water homeostasis via Aquaporins	95	9	3.9 x 10^{-10}	9.5 x 10^{-8}
Glucagon-type ligand receptors	45	7	1.3 x 10^{-9}	2.1 x 10^{-7}
Aquaporin-mediated transport	111	9	1.5 x 10^{-9}	2.1 x 10^{-7}
Cooperation of PDCL and TRiC/CCT in G-protein beta folding	47	7	1.75 x 10^{-9}	2.1 x 10^{-7}
Prostacyclin signalling through prostacyclin receptor	49	7	2.33 x 10^{-9}	2.4 x 10^{-7}
Thromboxane signalling through TP receptor	53	7	3.9 x 10^{-9}	3.6 x 10^{-7}
G-protein activation	56	7	5.8 x 10^{-9}	4.2 x 10^{-7}
ADP signalling through P2Y purinoreceptor 12	56	7	5.8 x 10^{-9}	4.2 x 10^{-7}

Abbreviations: *FDR* - false discovery rate. Results are shown for the top 10 overrepresented pathways.

In addition to overrepresentation analysis, I performed a network topology-based analysis, which takes into account the interaction between the genes in the list. Several enriched gene ontology domains were identified (Table 6.6.). Most notably, multicellular organismal process showed the strongest level of enrichment (p = 1.41 x 10^{-7}), while other domains such as G protein-coupled receptor signalling pathway matched those identified by *REACTOME.*

Table 6.6. Summary of enriched gene ontology categories identified using network topology-based analysis in *WebGestalt.*

Gene Ontology ID	Pathway Name	Total number of genes in pathway	Number of genes from UM-MDR list	*p*-value	*FDR*
GO:0032501	Multicellular organismal process	6115	44	1.42×10^{-7}	0.0022
GO:0023052	Signalling	5257	38	4.81×10^{-6}	0.024
GO:0007154	Cell communication	5304	38	6.14×10^{-6}	0.024
GO:0007186	G protein-coupled receptor signalling pathway	751	13	6.25×10^{-6}	0.024

Abbreviations: *FDR* - false discovery rate. Results are shown for pathways with *FDR* < 0.05.

6.4. Discussion

Gene-gene interactions have been long thought to be one of the causes of the "missing heritability problem" (Manolio et al., 2009; Slatkin, 2009). Experimental studies (Le Rouzic et al., 2007; Le Rouzic and Carlborg, 2008; Forsberg et al., 2017) have demonstrated that modelling genetic interactions can improve prediction of quantitative traits despite theoretical work showing that there is little to no advantage of considering such interactions (Hill et al., 2008). The discrepancy between theoretical and experimental results could lie in the fact that the additive model can capture most of the variance of non-additive genetic effects (Huang and Mackay, 2016). Due to limited examples of gene-gene interactions in refractive error, the current study aimed at providing a comprehensive guide to the extent and importance of such interactions in refractive error development.

In order to discover significant pairs of multi-locus combinations, a recently proposed implementation of the original MDR algorithm was used (Yu et al., 2016). Multifactor dimensionality reduction methods have been demonstrated to have higher statistical power compared to standard linear or logistic regression (Ritchie et al., 2003; Coffey et al., 2004; Motsinger and Ritchie, 2006). This increase in power comes from the fact that MDR relaxes the assumption of a particular inheritance model and does not put strict constraints on the parameters to be estimated. As a result, it allows for the detection of non-linear interactions, a task that is not easily accomplished with regression modelling. A comparison between interaction effects estimated using linear regression and UM-MDR (Figure 6.4.), showed that the latter approach resulted in a more widely spread distribution of effect sizes. In future, such information could

potentially be used to improve phenotypic prediction in a manner similar to polygenic risk scores.

Despite the benefits that come with MDR, only one significant interaction pair between rs16890054 (*ZMAT4*) and rs5442 (*GNB3*) was identified after adjustment for marginal SNP effects. Unfortunately, this interaction did not replicate, which could suggest that my initial finding was a false-positive. Similar to previous findings (Yu et al., 2016), we found that ignoring the contribution from marginal SNP effects in MDR led to severe inflation of false-positive findings (Figure 6.3. left panel). For example, all comparisons reaching the Bonferroni corrected threshold would be considered to have originated due to interactions using the original MDR algorithm (Ritchie et al., 2001; Motsinger and Ritchie, 2006). These false-positive findings were driven by strong individual SNP effects, as evidenced by only one significant SNP-SNP pair remaining after adjustment for marginal effects (Figure 6.3. right panel).

It is generally believed that novel genomic regions associated with a phenotype via gene-gene interactions will be pinpointed as sample sizes become larger. Motivated by this idea, I sought to investigate whether additional interacting pairs of SNPs could be identified in a larger combined sample of AOSW-inferred refractive error and autorefraction measured refractive error. The association between rs16890054 and rs5442 remained strong (Table 6.3.), and an additional 13 interacting pairs were found (Table 6.4.). Notably, *ZMAT4* was involved in every single interaction pair. It has been suggested that interacting loci form highly connected epistatic networks (Carlborg et al., 2006; Forsberg et al., 2017). These are usually arranged such that variants involved in many interactions tie more extensive networks together, while the hub genetic variants are in the centre connecting many such radial networks (Forsberg et al., 2017). Therefore, I speculate that *ZMAT4* could act as a hub gene that is involved in mediating the action of other genes identified by UM-MDR. Since the biological function of ZMAT4 protein is nucleic acid binding, it is reasonable to assume that this gene is responsible for regulating expression of other refractive error associated genes. This gene is highly expressed in brain tissues. More specifically, the highest protein expression was observed in cerebellum and the highest RNA expression was observed in pons and medulla according to the human protein atlas (URL:

https://www.proteinatlas.org/ENSG00000165061-ZMAT4/tissue, accessed: 29 October 2019).

A list of 14 unique genes identified by UM-MDR in the combined sample was selected for further characterisation. Using *GeneMANIA*, an additional 50 genes were identified that were not part of the original list of candidate genes (Figure 6.5.). Of these, only *RDH5* has been previously associated with refractive error (Tedja et al., 2018). A few other genes have been previously associated with other ocular traits or traits that affect facial features, including the eyes. For example, according to GeneCards (URL: https://www.genecards.org/, accessed: 29 October 2019), the *PICK1* gene has been associated with age-related macular degeneration (Lin et al., 2012), *FNBP4* with micropthalmia (Kondo et al., 2013), *TGFB1* associated with Camurati-Engelmann disease (Janssens et al., 2003), which in some cases can lead to loss of vision, *TGFB2* is associated with macular holes (Liu and Qiu, 1998) and *RAB25* is associated with binocular vision deficits. Regarding enrichment of biological process gene ontology categories, 25 of the 64 genes showed evidence of involvement in the control of cellular processes (GO:0009987), followed by 13 genes responsible for the control of the metabolic process (GO:0008152). Approximately half of the genes showed evidence of binding cytoskeleton and DNA as their primary molecular function (GO:0005488), followed by 16 genes acting as catalytic activators (GO:0003824). *REACTOME* overrepresentation analysis revealed presynaptic function of kainate receptors and activation of kainate receptors upon glutamate binding as two most enriched pathways. A noteworthy function of kainate receptors includes modulation of the release of neurotransmitters like glutamate and gamma amino butyric acid (GABA). Both glutamate and GABA affect the transmission of visual information and shape the development of the retina (Guoping et al., 2017). The functions of other enriched pathways could be broadly divided into signalling of various molecular entities, such as thromboxane and prostacyclin, or transport of metabolites such as potassium and calcium. Network topology-based analysis, which took into consideration not only a list of genes but also the interaction between them, identified multicellular biological processes. This gene ontology category includes any biological process that occurs at the level of a multicellular organism and can range from cytokine production to circadian regulation of heart rate. Three other categories with FDR < 0.05

were related to cell signalling. Consistent results from the overrepresentation analysis and network topology-based analysis reinforces the importance of signalling cascades in the development of refractive error.

MDR has been used to study gene-gene interactions in myopia previously. Chen et al. (Chen et al., 2011c) used MDR to look for the relationship between steroidogenesis enzyme genes and high myopia in Taiwanese individuals. They concluded that an interaction between steroidogenesis genes might be a modulating factor in sex hormone metabolism and high-myopia risk. In another study, the same research team looked for the association of the lumican gene with high myopia susceptibility (Chen et al., 2009b). In addition, MDR has been used in the past to study other ocular traits such as open-angle glaucoma (Jia et al., 2009) and eye colour (Pośpiech et al., 2011; Zidkova et al., 2013).

One limitation of this study was restricting analyses to genome-wide significant variants. However, selecting SNPs from candidate genes is a common practice for this type of analysis (Coffey et al., 2004; Tsai et al., 2004; Su et al., 2012; Xu et al., 2014). Moreover, the cross-validation strategy (Table 6.1.) suggested that less stringent association thresholds are not likely to identify variants that are important in shaping the phenotype. Although the FDR could not be estimated reliably when adjusting for marginal SNP effects (Table 6.2.), I made the assumption that only variants with marginal SNP effects would display evidence of interaction (i.e. no interaction in the absence of marginal effects) based on the FDR analysis when not adjusting for marginal effects.

In summary, these findings provide tentative evidence that gene-gene interactions contribute to refractive error development. The *ZMAT4* gene emerged as the most promising candidate interacting gene, potentially involved in a diverse range of gene-gene interactions. Network topology-based analysis identified enriched pathways with plausible roles in myopia susceptibility. Further validation of the identified SNP-SNP interaction combinations and enriched gene pathways is required, in independent datasets.

Chapter 7

Concluding remarks

As discussed in section 1.1.3., despite overwhelming evidence of involvement of genetic and environmental components in determining shifts in refractive error (Morgan and Rose, 2019), only a handful of gene-environment interaction effects have been identified to date (Chen et al., 2011a; Tkatchenko et al., 2015; Fan et al., 2016), while no clear and replicable evidence of gene-gene interactions exist. This can be explained by GxE and GxG interaction studies being underpowered compared to the standard GWAS approach that tests for marginal SNP effects (Duncan and Keller, 2011). Motivated by the release of UK Biobank, where the large amount of data pertaining to the genetic variation of approximately 500,000 individuals has been collected and processed, the primary focus of this thesis was to apply existing statistical approaches in order to identify genetic loci with roles in refractive error development via interaction with environmental exposures or other genetic variants.

Throughout the thesis, I tried to stay consistent with respect to methodological procedures. First, statistical analyses in all chapters were restricted to participants who self-identified as white-British. Including individuals from diverse backgrounds could introduce bias when studying GxE effects if environmental exposures differ. For example, University degree was considered as one of the environmental factors in this dissertation. Education systems vary across countries (URL: http://www.indire.it/en/2017/08/29/school-education-in-europe/, accessed: 29 October 2019) therefore the use of non-harmonised environmental exposures could confound observed findings. Second, motivated by the poor replication rate of initial GxE discoveries in candidate gene studies (Duncan and Keller, 2011; Dick et al., 2015; Border and Keller, 2017), Chapter 2, Chapter 3, and Chapter 6 used discovery and replication samples to avoid reporting false positive findings. In all three cases,

association between GxE/GxG interaction effects and AOSW-inferred refractive error was tested in the discovery sample and replicated using the autorefraction phenotype. I chose to use the sample of participants with the carefully measured autorefraction phenotype as the replication sample, so that any replicating loci would provide high confidence of being involved in GxE or GxG interactions.

In Chapter 2, I explored the role of GxE interaction using two different measurement scales (additive and multiplicative). The fact that the *LAMA2* gene was successfully identified and replicated using either of the two risk scales provided a high degree of confidence of its involvement in an interaction with education. GxE interaction tested on the multiplicative scale provided little information about the functional role of interactions even when the discovery and replication samples were meta-analysed. On the other hand, GxE interaction tested on the additive scale provided a more comprehensive view. In particular, it was apparent that a large proportion of the genes identified as having GxE effects were already known to be associated with myopia via their marginal effects (Kiefer et al., 2013; Verhoeven et al., 2013; Tedja et al., 2018). Collectively, these results suggest that if the goal of a study is to identify novel genetic variants that are associated with refractive error, the effort put into investigating GxE interactions might not yield a substantial benefit, since any true findings are likely to re-identify variants already known to have marginal effects on the phenotype. For standard GWAS analyses of refractive error it is relatively easy to combine studies via meta-analysis; however, this approach is much less straightforward in the case of GxE interaction effects. For instance, different studies might have examined distinct environmental exposures or exposure at different stages of childhood. Nevertheless, if attempts to study interactions are made, I suggest that GxE interaction measured on the additive scale might provide a more insightful view into the underlying biology of refractive error in forthcoming studies.

In Chapter 3, I evaluated a rarely-used approach ('variance heterogeneity' analysis) that relaxes the assumption of modelling GxE and GxG interactions without explicitly having any information available about the interacting factor. Instead of focusing on finding the *mean* difference between the genotype groups, this method tests for a difference in the *variance* between the genotype groups. Consistent with the findings from Chapter 2, the majority of the identified genetic variants clustered in genomic

regions already known to be involved in refractive error development. Once again, *LAMA2* was among the top candidate genes showing evidence of variance heterogeneity, providing further support for its modifiable role in the presence of other environmental or genetic risk factors. I observed a systematic increase in the interaction effect size of variants with additional years spent in education. This finding supports the notion that more years spent in education leads to a more negative refractive error (Mountjoy et al., 2018). The fact that effect sizes vary depending on environmental exposures could provide a new avenue in the prediction of complex phenotypes such as refractive error. The *heteroskedastic linear model* analysis in Chapter 3 showed how potential false positive findings arising by virtue of a statistical artefact (a 'mean-variance' relationship) could be eliminated. I propose that future studies investigating genetic determinants that influence refractive error could benefit from studying the joint effect of the *mean* and the *variance* shift across genotype groups. Such approaches could be integrated into sophisticated modelling strategies that build upon the standard polygenic risk score approach, in order to advance the field of personalised medicine.

Given that Chapters 2 and 3 consistently suggested an enrichment of interacting variants among those showing evidence of marginal association, Chapter 4 focused on examining the extent of SNP effect heterogeneity across the sample distribution for refractive error-associated variants. Using conditional quantile regression, I was able to show that genetic variants exert different effects in different individuals. A visual representation of the highly non-linear profile, observed for the majority of tested variants, showed that SNP effects for individuals in the myopic arm of the refractive error distribution were as much as 4-fold higher compared to emmetropes. Although there are several explanations for such phenomena, this is exactly what we would expect to see in case of involvement in GxE or GxG interactions that modify the risk in certain individuals. I speculate that interacting factors might act through the control of emmetropization. Finally, I provided additional evidence that education is a critical determinant for refractive error, by observing that SNP effects in individuals who spent more time in education were stronger compared to those individuals who spent the least time in education.

In Chapter 5, I shifted my focus to exploring the usefulness of GxE interactions in making phenotypic predictions. I found that the inclusion of GxE effects in prediction models is unlikely to substantially improve the prediction of refractive error; indeed, the phenotypic variance explained by SNP x Education interactions did not exceed 1%. This finding is in line with previous studies, for example no improvement in prediction of breast cancer, type 2 diabetes, or rheumatoid arthritis was observed by including GxE or GxG interaction effects (Aschard et al., 2012a). Although future studies utilising larger sample sizes may achieve an improvement in prediction by using GxE, as was the case for major depressive disorder (Arnau-Soler et al., 2019), I argue that phenotype prediction will be largely dominated by purely additive effects.

Chapter 6 was aimed at exploring the role of GxG interaction in refractive error using multifactor dimensionality reduction. I adopted an FDR based method to prioritize genetic variants for testing GxG interaction effects, observing an increasing rate of false positive findings when less-stringently-associated SNPs were included. This once again suggested that variants with marginal effects are also involved in (GxG) interactions. Therefore, using only genetic variants that showed a genome-wide significant main effect association ($p < 5 \times 10^{-8}$), I discovered one region that showed significant evidence of GxG interaction in the discovery sample. However, no evidence of such an interaction was observed in the independent replication sample. To increase the number of potentially interacting loci, I performed a meta-analysis of the discovery and replication samples. This revealed 13 additional candidate regions that could potentially be involved in GxG interactions. Interestingly, the *ZMAT4* gene was implicated in every single comparison. Therefore, I propose that *ZMAT4* might act as a hub gene to influence the mode of action of other genes. I explored the functional consequences of the newly-identified GxG loci by performing pathway enrichment analysis and network topology-based analysis. Although, there was strong enrichment for several pathways, most of these pathways were already known to be relevant for refractive error.

Several limitations of the UK Biobank dataset hindered a more comprehensive investigation of the role of genetic and environmental risk factors in refractive error development. First, only approximately 25% of UK Biobank participants had their refractive error measurement taken, leaving a large proportion of individuals for whom

refractive error was inferred using age of onset of spectacles wear. As discussed in Chapter 1 and 3, the inferred phenotype was imprecise (correlation with true phenotype = 0.55), leading to reduced statistical power to detect associations. This could have been caused by demographic differences between discovery and replication samples which reflect regional demographic differences (Section 1.4.2.). This source of measurement error could have given rise to the strong mean-variance relationship observed in Chapter 3, which differed markedly compared to the mean-variance relationship observed for autorefraction measured refractive error. The imprecision of the AOSW-inferred refractive error phenotype may also have contributed to the observed high rate of false positive findings in other experimental chapters (that is, associations that were not replicated using the autorefraction measured refractive error dataset). Second, refractive error was assessed without cycloplegia, which can result in measurement error due to certain participants accommodating during the test. However, since UK Biobank participants were aged 40 years old, the degree of measurement error was likely to be minimal compared to errors that can arise when non-cycloplegic autorefraction is performed in children. This source of bias could have affected the estimation of SNP effect sizes in the analyses performed in this thesis. For example, in Chapter 4, the measurement error of refractive error phenotype could have affected the estimation of risk factor effect sizes at certain quantiles. This, in turn, could have led to imprecise assessment of interindividual variation in SNP effect sizes. Specifically, if the higher quantiles - those comprising individuals with hyperopia - were relatively more affected by measurement error this may have led to the attenuation of risk factor effect size estimates for these higher quantiles, which would in turn have attenuated the difference in effect size for hyperopes vs. emmetropes. Another source of bias could come from the fact that UK Biobank cohort is not representative of the general population and is enriched in more affluent individuals who have obtained a university degree. This could have caused distorted associations due to collider bias. Specifically, when education level influences selection into a study, then genetic variants associated with education may yield biased associations with traits correlated with education, such as refractive error. Similarly, associations between certain SNPs and refractive error could be subject to bias when including education as a covariate in the analysis.

In summary, my findings demonstrate that searching for GxE or GxG interactions in refractive error development is far less straightforward than identifying genetic variants with marginal effects on the phenotype. Furthermore, extreme care needs to be taken in confirming the validity of novel findings by seeking to replicate the results in an independent sample. (Most of the interactions identified in the discovery sample that were statistically significant after Bonferroni correction were deemed likely to be false positive findings because of failure to replicate in an independent sample). Those loci with convincing evidence of involvement in an interaction had a high likelihood of being enriched among genes already known to be associated with refractive error through standard GWAS analyses. Consequently, downstream analyses such as gene-based association tests and pathway enrichment tests generally led to the strengthening of previously established findings. I demonstrated that the standard ordinary least squares approach is inadequate when it comes to modelling interactions. Future studies should consider alternative strategies, such as joint methods that take into account the main and interaction effect of a SNP or conditional quantile regression that does not assume uniform SNP effect sizes across individuals. More sophisticated methods could be developed to include GxE and GxG interaction effects when making phenotypic predictions. Given the available evidence that complex traits are mediated by complicated networks involving interactions between genetic markers and external environmental exposures, I anticipate that the interest to study GxE and GxG interactions will increase as the size of genomic datasets grows.

Appendix A: Full results for 82 independent genes identified after gene-based clumping of Levene's median test summary statistics

Gene	*CHR*	*START BP*	*STOP BP*	*P discovery*	*P replication*
LAMA2	6	129154286	129887711	1.02×10^{-36}	9.09×10^{-7}
NPLOC4	17	79473913	79654138	6.58×10^{-29}	1.44×10^{-9}
ZMAT4	8	40338109	40805345	4.42×10^{-26}	8.34×10^{-5}
CHRND	2	233340870	233451375	1.29×10^{-22}	1.29×10^{-9}
KCNQ5	6	73281571	73958574	4.99×10^{-19}	2.25×10^{-3}
PDE11A	2	178437977	179023066	1.29×10^{-18}	3.27×10^{-3}
RBFOX1	16	5239469	7813342	5.67×10^{-18}	1.25×10^{-2}
GPD2	2	157241965	157492915	2.88×10^{-17}	1.73×10^{-1}
RCBTB1	13	50056082	50209742	5.95×10^{-17}	2.76×10^{-2}
CTNNB1	3	41186401	41331939	1.02×10^{-16}	6.29×10^{-3}
LRRC4C	11	40085524	41531186	1.89×10^{-15}	2.06×10^{-6}
METAP1D	2	172814804	172997158	2.30×10^{-15}	1.29×10^{-1}
SARNP	12	56096247	56261540	4.43×10^{-14}	4.53×10^{-3}
KCNMA1	10	78579359	79447577	1.22×10^{-12}	3.93×10^{-1}
FRMPD2	10	49314602	49532941	1.88×10^{-12}	7.27×10^{-2}
LRIT1	10	85941276	86051217	2.45×10^{-12}	1.92×10^{-7}
BICC1	10	60222774	60641194	6.31×10^{-12}	4.59×10^{-2}
AKAP6	14	32748479	33352268	1.64×10^{-11}	6.77×10^{-1}
RASGRF1	15	79202289	79433215	1.95×10^{-11}	2.28×10^{-1}
BARHL2	1	91127579	91232794	2.68×10^{-11}	7.65×10^{-3}
ACAA2	18	47259874	47390306	1.63×10^{-10}	3.18×10^{-2}
PPP6C	9	127858852	128002218	1.90×10^{-10}	2.92×10^{-2}
GRIA4	11	105430800	105902819	1.06×10^{-9}	4.25×10^{-4}
ZIC5	13	100565275	100674178	1.49×10^{-9}	1.62×10^{-5}
CCDC89	11	85344893	85447320	1.92×10^{-9}	4.01×10^{-1}
BMP4	14	54366454	54473554	2.24×10^{-9}	1.80×10^{-1}
CD34	1	208009883	208134742	3.24×10^{-9}	9.45×10^{-2}
NTM	11	131190371	132256716	1.32×10^{-8}	8.59×10^{-2}
SOX7	8	10531278	10638084	1.85×10^{-8}	9.92×10^{-4}
HIST1H3J	6	27808093	27908570	1.85×10^{-8}	2.60×10^{-2}
HIVEP3	1	41922036	42551596	2.60×10^{-8}	4.30×10^{-1}
SIX3	2	45119037	45223216	2.83×10^{-8}	6.84×10^{-1}
PLEKHG4	16	67261413	67373403	2.87×10^{-8}	5.15×10^{-1}
PAX6	11	31756340	31889509	3.60×10^{-8}	1.75×10^{-1}
HPD	12	122227433	122376517	3.71×10^{-8}	8.54×10^{-1}
SHISA6	17	11094740	11517380	5.99×10^{-8}	1.45×10^{-2}
AQR	15	35098552	35311995	7.76×10^{-8}	1.29×10^{-1}
BPTF	17	65771644	66030494	1.12×10^{-7}	1.48×10^{-1}
PKHD1	6	51430145	52002423	1.34×10^{-7}	8.47×10^{-2}
TPRA1	3	127241907	127359602	1.42×10^{-7}	6.05×10^{-1}
ZSWIM6	5	60578100	60891999	1.69×10^{-7}	2.05×10^{-2}
CAMSAP2	1	200658686	200879832	2.80×10^{-7}	1.38×10^{-1}
ANTXR2	4	80772771	81044626	3.61×10^{-7}	2.48×10^{-1}
TNKS	8	9362756	9689856	8.24×10^{-7}	1.94×10^{-2}
MMP2	16	55463081	55590586	1.54×10^{-6}	6.22×10^{-1}
SLC16A10	6	111358728	111594608	1.72×10^{-6}	2.56×10^{-1}
ZEB2	2	145091942	145327958	1.86×10^{-6}	7.35×10^{-2}
SLC45A4	8	142167273	142360241	1.87×10^{-6}	4.80×10^{-1}
PPIP5K2	5	102405958	102589224	3.22×10^{-6}	9.26×10^{-1}
DOCK9	13	99395741	99788660	3.46×10^{-6}	1.79×10^{-1}
SLC14A2	18	42742947	43313072	3.67×10^{-6}	7.03×10^{-1}
PPP4R2	3	72995985	73168349	3.89×10^{-6}	1.64×10^{-1}
TET2	4	106017032	106250960	4.16×10^{-6}	2.88×10^{-1}
ST8SIA1	12	22296325	22537648	5.97×10^{-6}	5.08×10^{-1}
SPAG4	20	34153776	34258967	6.98×10^{-6}	6.46×10^{-2}
LOC101927855	17	59390108	59494966	7.18×10^{-6}	7.10×10^{-1}
DSCAM	21	41334343	42269039	7.22×10^{-6}	4.71×10^{-1}

SNTB1	8	121497985	121875599	7.36×10^{-6}	3.45×10^{-1}
UBE2I	16	1307420	1427019	7.48×10^{-6}	5.75×10^{-1}
PTPN5	11	18699475	18864268	8.46×10^{-6}	1.11×10^{-1}
GRAMD1B	11	123275191	123548478	1.00×10^{-5}	8.09×10^{-1}
TRAF3IP1	2	239179185	239359541	1.28×10^{-5}	9.08×10^{-1}
VAV3	1	108063782	108557545	1.36×10^{-5}	8.04×10^{-2}
EIF2B2	14	75419612	75526294	1.38×10^{-5}	8.09×10^{-1}
CHDH	3	53800324	53930420	1.98×10^{-5}	4.44×10^{-1}
NRIP1	21	16283556	16488224	2.19×10^{-5}	2.13×10^{-2}
TMC1	9	75086717	75501267	2.59×10^{-5}	1.70×10^{-1}
BMP2	20	6698745	6810910	2.66×10^{-5}	9.85×10^{-2}
OAZ1	19	2219520	2323487	2.92×10^{-5}	1.08×10^{-2}
THBS4	5	79281170	79429111	3.29×10^{-5}	3.88×10^{-1}
ADAM17	2	9579392	9745917	3.88×10^{-5}	2.85×10^{-1}
AIP	11	67200505	67308579	4.11×10^{-5}	8.23×10^{-2}
SIDT2	11	116999626	117118161	4.40×10^{-5}	9.69×10^{-1}
MEF2C	5	87964058	88249922	4.49×10^{-5}	7.75×10^{-2}
ATXN1L	16	71829894	71941236	4.57×10^{-5}	2.07×10^{-1}
THRB	3	24108644	24586772	4.89×10^{-5}	3.11×10^{-1}
MYO1D	17	30769628	31253902	5.68×10^{-5}	1.59×10^{-1}
MORN4	10	99324310	99443913	6.31×10^{-5}	1.16×10^{-1}
AMT	3	49404211	49510111	6.49×10^{-5}	2.55×10^{-1}
NAA38	17	7710003	7838608	7.05×10^{-5}	2.07×10^{-2}
ZNF689	16	30563879	30672096	7.75×10^{-5}	5.29×10^{-1}
OTOF	2	26630071	26831566	8.68×10^{-5}	2.46×10^{-1}

START and *STOP* base-pair positions include 50kb region on either side of the transcribed gene. After gene-based clumping of vQTL results, 550 independent genes were retained. Results are presented for genes that showed $p < 0.05/550 = 9.1 \times 10^{-5}$ in the discovery sample using AOSW-inferred refractive error phenotype.

Appendix B: Summary statistics for standard linear regression effect size estimates for association with refractive error.

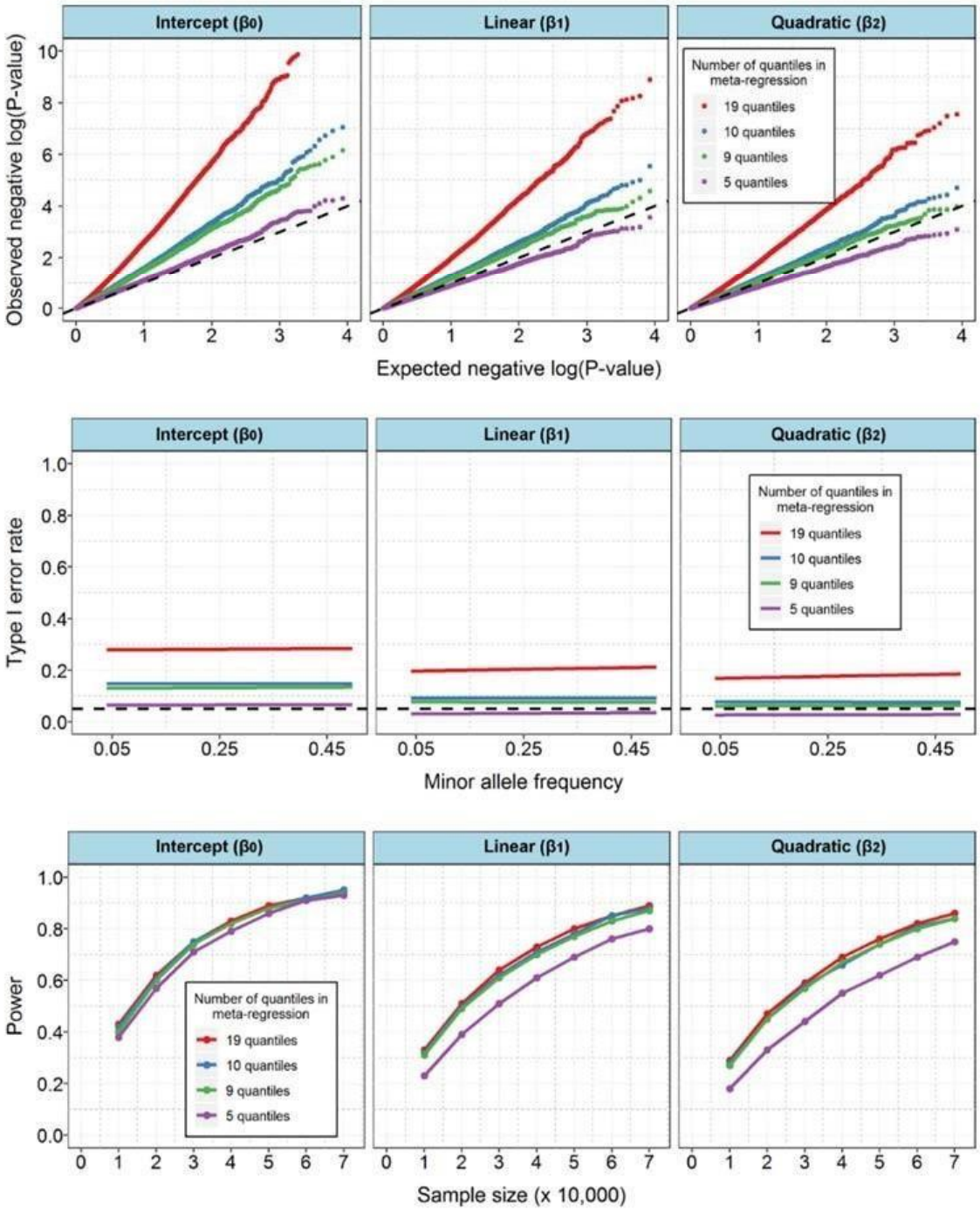
SNP	*Gene*	*CHR*	*BP*	*EA*	*MAF*	*Beta [95% CI]*	*P*
rs524952	*GOLGA8B_GJD2*	15	35005886	A	0.49	-0.258 [-0.285; -0.231]	6.18×10^{-79}
rs12193446	*BC035400_LAMA2*	6	129820038	A	0.10	-0.434 [-0.48; -0.389]	1.06×10^{-77}
rs7744813	*KCNQ5*	6	73643289	A	0.41	-0.218 [-0.246; -0.19]	1.84×10^{-53}
rs1550094	*PRSS56*	2	233385396	G	0.31	-0.204 [-0.234; -0.175]	5.69×10^{-43}
rs11602008	*LRRC4C*	11	40149305	T	0.17	-0.238 [-0.274; -0.202]	5.14×10^{-38}
rs72621438	*SNORA51_CA8*	8	60178580	C	0.35	-0.175 [-0.203; -0.147]	2.99×10^{-34}
rs6495367	*RASGRF1*	15	79375347	A	0.42	-0.158 [-0.185; -0.131]	7.87×10^{-30}
rs10500355	*RBFOX1*	16	7459347	A	0.36	-0.164 [-0.193; -0.135]	2.39×10^{-29}
rs2573210	*PRSS56*	2	233385025	G	0.19	-0.192 [-0.226; -0.157]	4.51×10^{-28}
rs2326823	*BC035400*	6	129842188	C	0.08	-0.252 [-0.301; -0.204]	2.21×10^{-24}
rs2573081	*PDE11A*	2	178828507	G	0.46	-0.137 [-0.164; -0.11]	2.51×10^{-23}
rs5442	*GNB3*	12	6954864	A	0.07	-0.256 [-0.309; -0.204]	1.47×10^{-21}
rs2276560	*EIF4E2_EFHD1*	2	233450919	C	0.24	-0.145 [-0.176; -0.113]	2.56×10^{-19}
rs7895108	*KCNMA1*	10	79061458	T	0.37	-0.126 [-0.154; -0.098]	6.35×10^{-19}
rs6433704	*PDE11A*	2	178847912	G	0.41	-0.127 [-0.155; -0.098]	3.13×10^{-18}
rs2908972	*SHISA6*	17	11407259	A	0.39	-0.121 [-0.149; -0.092]	4.27×10^{-17}
rs3138137	*BLOC1S1-RDH5_RDH5*	12	56116981	C	0.49	-0.116 [-0.143; -0.088]	1.01×10^{-16}
rs17400325	*PDE11A*	2	178565913	C	0.04	-0.28 [-0.348; -0.213]	2.86×10^{-16}
rs2855530	*BMP4*	14	54421917	C	0.49	-0.111 [-0.138; -0.084]	4.76×10^{-16}
rs9517964	*ZIC2_PCCA*	13	100717833	C	0.43	-0.113 [-0.141; -0.086]	7.45×10^{-16}
rs745480	*LRIT2_LRIT1*	10	85986554	G	0.48	-0.109 [-0.136; -0.082]	1.93×10^{-15}
rs2166181	*RASGEF1B_U6*	4	82422327	G	0.47	-0.11 [-0.137; -0.082]	2.77×10^{-15}
rs1858001	*C4BPA_CD55*	1	207488004	G	0.31	-0.117 [-0.146; -0.087]	4.20×10^{-15}
rs62070229	*MYO1D_TMEM98*	17	31227593	G	0.18	-0.137 [-0.172; -0.102]	9.61×10^{-15}
rs7829127	*ZMAT4*	8	40726394	A	0.21	-0.131 [-0.164; -0.098]	1.27×10^{-14}
rs12898755	*APH1B*	15	63574641	G	0.21	-0.130 [-0.163; -0.097]	1.53×10^{-14}
rs511217	*METTL15_KCNA4*	11	30029948	A	0.26	-0.117 [-0.148; -0.087]	8.36×10^{-14}
rs10887262	*RGR*	10	86009171	C	0.29	-0.110 [-0.140; -0.081]	3.03×10^{-13}
rs11118367	*LYPLAL1*	1	219790221	T	0.45	-0.100 [-0.127; -0.073]	3.83×10^{-13}
rs13069734	*ZBTB38*	3	141148419	G	0.35	-0.106 [-0.135; -0.077]	4.59×10^{-13}
rs10511652	*SH3GL2_ADAMTSL1*	9	18362865	G	0.40	-0.103 [-0.131; -0.075]	5.97×10^{-13}
rs12965607	*MYO5B*	18	47391025	G	0.15	-0.135 [-0.173; -0.098]	1.25×10^{-12}
rs1556867	*5S_rRNA_PBX1*	1	164213686	T	0.24	-0.114 [-0.145; -0.082]	1.91×10^{-12}
rs7624084	*ZBTB38*	3	141093285	T	0.44	-0.097 [-0.124; -0.070]	2.88×10^{-12}
rs2573232	*ALPPL2_ALPI*	2	233300046	T	0.09	-0.166 [-0.213; -0.119]	3.07×10^{-12}
rs56075542	*BC040861_PABPC1P2*	2	146882415	T	0.45	-0.096 [-0.124; -0.069]	4.81×10^{-12}
rs7747	*ANTXR2*	4	80827062	C	0.20	-0.119 [-0.153; -0.085]	5.21×10^{-12}
rs12451582	*NOG_C17orf67*	17	54734643	G	0.36	-0.098 [-0.126; -0.070]	8.76×10^{-12}
rs837323	*PCCA*	13	101175664	C	0.47	-0.094 [-0.121; -0.067]	9.78×10^{-12}
rs7042950	*RORB*	9	77149837	G	0.22	-0.112 [-0.144; -0.079]	1.19×10^{-11}
rs2229742	*NRIP1*	21	16339172	C	0.11	-0.151 [-0.195; -0.107]	1.34×10^{-11}
rs2143964	*BMP4_CDKN3*	14	54726800	G	0.26	-0.106 [-0.137; -0.075]	1.57×10^{-11}
rs1954761	*GRIA4*	11	105596885	T	0.37	-0.095 [-0.123; -0.067]	2.98×10^{-11}
rs11210537	*HIVEP3*	1	42345723	G	0.32	-0.098 [-0.127; -0.069]	3.16×10^{-11}
rs2155413	*DLG2*	11	84634790	A	0.47	-0.093 [-0.121; -0.066]	3.81×10^{-11}
rs12883788	*AKAP6_NPAS3*	14	33303540	C	0.46	-0.091[-0.118; -0.064]	5.82×10^{-11}
rs2150458	*PCBP3_COL6A1*	21	47377296	G	0.43	-0.091 [-0.118; -0.063]	7.97×10^{-11}
rs41393947	*PNPT1_EFEMP1*	2	56011517	A	0.14	-0.129 [-0.168; -0.090]	8.67×10^{-11}
rs2225986	*LINC00862*	1	200311910	A	0.39	-0.090 [-0.118; -0.063]	1.37×10^{-10}
rs297593	*GPD2*	2	157363743	T	0.29	-0.096 [-0.126; -0.066]	3.15×10^{-10}
rs11802995	*KIRREL*	1	158053024	C	0.23	-0.104 [-0.136; -0.071]	3.35×10^{-10}
rs7667446	*C4orf22_BMP3*	4	81906024	C	0.18	-0.110 [-0.145; -0.075]	4.69×10^{-10}
rs72826094	*TCF7L2*	10	114801488	T	0.20	-0.107 [-0.140; -0.073]	4.73×10^{-10}
rs34539187	*FBN1*	15	48756536	C	0.13	-0.126 [-0.166; -0.086]	7.19×10^{-10}
rs1237670	*HP08777*	1	113418415	G	0.22	-0.100 [-0.133; -0.068]	1.06×10^{-9}
rs2622646	*NCOA2_TRAM1*	8	71413737	A	0.35	-0.088 [-0.117; -0.060]	1.12×10^{-9}
rs7662551	*LOC100506035_PCAT4*	4	80537638	G	0.25	-0.096 [-0.127; -0.065]	1.20×10^{-9}
rs8073754	*C17orf47*	17	56618030	C	0.17	-0.110 [-0.146; -0.075]	1.20×10^{-9}
rs11101263	*FRMPD2*	10	49414181	T	0.27	-0.092 [-0.123; -0.062]	2.33×10^{-9}
rs6420484	*TSPAN10*	17	79612397	A	0.36	-0.086 [-0.114; -0.058]	2.44×10^{-9}
rs1313240	*JB175233_C14orf39*	14	60848527	T	0.29	-0.090 [-0.119; -0.060]	3.17×10^{-9}
rs115152181	*CDRT15*	17	14136125	T	0.43	-0.082 [-0.109; -0.055]	3.39×10^{-9}
rs1928175	*LINC00340*	6	22079485	A	0.44	-0.081 [-0.109; -0.054]	4.43×10^{-9}
rs9395623	*TFAP2D_TFAP2B*	6	50757699	T	0.33	-0.085 [-0.114; -0.056]	5.54×10^{-9}

rs4793501	*KCNJ2_BC039327*	17	68718734	T	0.42	-0.081 [-0.109; -0.054]	7.14×10^{-9}
rs36024104	*LRFN5*	14	42294993	G	0.19	-0.101 [-0.135; -0.067]	7.33×10^{-9}
rs9547035	*LINC00333_LINC00351*	13	85573496	G	0.26	-0.090 [-0.121; -0.059]	9.02×10^{-9}
rs1207782	*LINC00340*	6	22059967	T	0.38	-0.081 [-0.109; -0.053]	1.21×10^{-8}
rs2745953	*CD34*	1	208062973	T	0.29	-0.086 [-0.116; -0.056]	1.63×10^{-8}
rs9681162	*AK124857_LMCD1-AS1*	3	8194734	C	0.29	-0.085 [-0.114; -0.055]	2.28×10^{-8}
rs28471081	*RBFOX1*	16	7414383	A	0.21	-0.095 [-0.128; -0.062]	2.65×10^{-8}
rs12526735	*KCNQ5*	6	73648822	T	0.48	-0.075 [-0.102; -0.048]	4.83×10^{-8}
rs10853531	*SLC14A2*	18	42824449	G	0.21	-0.093 [-0.126; -0.059]	5.11×10^{-8}
rs4764038	*GRIN2B*	12	14062637	T	0.26	-0.085 [-0.116; -0.054]	6.38×10^{-8}
rs7925340	*FSHB_ARL14EP*	11	30280408	A	0.23	-0.088 [-0.120; -0.056]	6.50×10^{-8}
rs9295499	*CDKAL1*	6	21160689	C	0.32	-0.079 [-0.108; -0.050]	8.53×10^{-8}
rs7122817	*DSCAML1*	11	117657679	G	0.48	-0.074 [-0.101; -0.046]	1.07×10^{-7}
rs11723482	*PCAT4_ANTXR2*	4	80793199	T	0.26	-0.083 [-0.114; -0.052]	1.37×10^{-7}
rs11145465	*TJP2*	9	71766593	A	0.22	-0.087 [-0.119; -0.054]	2.05×10^{-7}
rs1790165	*NTM*	11	131928971	C	0.42	-0.072 [-0.099; -0.045]	2.28×10^{-7}
rs56014528	*NDUFB1*	14	92598635	G	0.15	-0.098 [-0.136; -0.061]	2.42×10^{-7}
rs1969091	*TMC3_MEX3B*	15	82326775	C	0.29	-0.078 [-0.108; -0.049]	2.54×10^{-7}
rs1064583	*COL10A1*	6	116446576	G	0.40	-0.072 [-0.099; -0.044]	3.09×10^{-7}
rs807037	*KAZALD1*	10	102824349	C	0.34	-0.074 [-0.102; -0.045]	3.56×10^{-7}
rs10003846	*C4orf22_BMP3*	4	81923677	T	0.10	-0.114 [-0.158; -0.070]	3.95×10^{-7}
rs7337610	*FLT1*	13	28962666	T	0.37	-0.072 [-0.100; -0.044]	4.15×10^{-7}
rs4237285	*C10orf11*	10	77814981	C	0.46	-0.068 [-0.095; -0.041]	8.12×10^{-7}
rs4795364	*MED1*	17	37576546	G	0.25	-0.077 [-0.108; -0.046]	1.21×10^{-6}
rs1555075	*RALY*	20	32610401	C	0.35	-0.070 [-0.098; -0.042]	1.23×10^{-6}
rs1994840	*C4orf22*	4	81707526	T	0.22	-0.080 [-0.112; -0.047]	1.79×10^{-6}
rs235770	*BMP2*	20	6761765	T	0.39	-0.066 [-0.094; -0.039]	2.59×10^{-6}
rs1649068	*BICC1*	10	60304864	A	0.45	-0.065 [-0.092; -0.038]	2.61×10^{-6}
rs7968679	*PZP*	12	9313304	G	0.31	-0.070 [-0.099; -0.041]	2.72×10^{-6}
rs6753137	*FAM150B_TMEM18*	2	301051	T	0.43	-0.065 [-0.092; -0.038]	3.10×10^{-6}
rs11589487	*AK097193_BC030753*	1	61342229	G	0.44	-0.064 [-0.091; -0.037]	3.34×10^{-6}
rs1983554	*MEI1_bK250D10.C22.8*	22	42194561	A	0.30	-0.067 [-0.097; -0.038]	6.05×10^{-6}
rs284818	*ST18_FAM150A*	8	53363937	C	0.13	-0.091 [-0.131; -0.052]	6.52×10^{-6}
rs10880855	*ARID2*	12	46144855	T	0.50	-0.061 [-0.089; -0.034]	8.89×10^{-6}
rs6903823	*SCAND3*	6	28322296	A	0.25	-0.070 [-0.102; -0.039]	9.56×10^{-6}
rs1150687	*ZNF192P1_TRNA_Ser*	6	28162469	C	0.39	-0.062 [-0.089; -0.034]	1.02×10^{-5}
rs1359543	*RCBTB1*	13	50159165	A	0.35	-0.064 [-0.092; -0.035]	1.03×10^{-5}
rs7737179	*TMEM161B-AS1_LINC00461*	5	87795525	A	0.22	-0.073 [-0.106; -0.041]	1.21×10^{-5}
rs11654644	*B4GALNT2_TRNA_Gln*	17	47263475	T	0.18	-0.078 [-0.113; -0.042]	1.52×10^{-5}
rs9416017	*DNAJB12*	10	74100279	T	0.36	-0.062 [-0.090; -0.033]	1.98×10^{-5}
rs4687586	*CACNA1D*	3	53837971	C	0.32	-0.062 [-0.092; -0.033]	2.83×10^{-5}
rs4808962	*GATAD2A*	19	19579557	G	0.17	-0.076 [-0.112; -0.040]	3.35×10^{-5}
rs10458138	*LOC100508120*	6	2429743	A	0.22	-0.068 [-0.101; -0.036]	4.17×10^{-5}
rs4260345	*THRB*	3	24256698	C	0.37	-0.058 [-0.086; -0.030]	5.03×10^{-5}
rs17428076	*HAT1_METAP1D*	2	172851936	C	0.24	-0.064 [-0.096; -0.033]	5.74×10^{-5}
rs3110134	*SNORA51_CA8*	8	60097984	G	0.31	-0.058 [-0.088; -0.029]	8.72×10^{-5}
rs10122788	*MVB12B*	9	129206832	G	0.43	-0.055 [-0.082; -0.027]	8.92×10^{-5}
rs79266634	*RBFOX1*	16	7309047	C	0.09	-0.093 [-0.140; -0.046]	0.000106
rs28658452	*MYCN_SNORA40*	2	16234068	A	0.08	-0.097 [-0.145; -0.048]	0.000111
rs79953651	*ANTXR2*	4	80982013	C	0.04	-0.135 [-0.204; -0.066]	0.000115
rs1454776	*GALNT15*	3	16009044	G	0.48	-0.053 [-0.080; -0.026]	0.000125
rs17382981	*CYP26A1_MYOF*	10	94953258	T	0.42	-0.053 [-0.080; -0.026]	0.00013
rs72655575	*SNORA51_CA8*	8	60556509	C	0.21	-0.065 [-0.098; -0.032]	0.000133
rs10760673	*TGFBR1*	9	101878622	A	0.20	-0.065 [-0.099; -0.031]	0.000146
rs9516194	*GPC5_GPC6*	13	93859498	G	0.50	-0.052 [-0.078; -0.025]	0.000166
rs17032696	*CAMKMT_SIX3*	2	45137870	A	0.18	-0.067 [-0.102; -0.032]	0.000183
rs8075280	*POLR2A_TNFSF12*	17	7434819	T	0.39	-0.053 [-0.081; -0.025]	0.000188
rs2116093	*BC043573*	8	10613299	G	0.40	-0.052 [-0.080; -0.024]	0.000233
rs17125093	*TTC8_TRNA_Ala*	14	89428948	A	0.20	-0.063 [-0.096; -0.029]	0.000277
rs7107014	*HNRNPKP3_API5*	11	43307811	C	0.48	-0.050 [-0.077; -0.023]	0.000307
rs11088317	*NRIP1_USP25*	21	16574122	T	0.29	-0.054 [-0.084; -0.024]	0.000356
rs7941828	*MPPED2*	11	30430331	C	0.36	-0.051 [-0.079; -0.023]	0.000358
rs11202736	*RNLS*	10	90142203	A	0.30	-0.054 [-0.083; -0.024]	0.000414
rs55885222	*SNTB1*	8	121621473	A	0.36	-0.051 [-0.079; -0.022]	0.000467
rs9388766	*L3MBTL3*	6	130354855	C	0.30	-0.051 [-0.081; -0.022]	0.000568
rs1358684	*SEMA3D_GRM3*	7	86103402	C	0.27	-0.053 [-0.083; -0.023]	0.000613
rs56055503	*MAF_DYNLRB2*	16	80532694	A	0.22	-0.057 [-0.089; -0.024]	0.00064
rs117735470	*ST8SIA1_C2CD5*	12	22592653	A	0.09	-0.079 [-0.125; -0.032]	0.000921

rs35337422	*RD3L*	14	104407243	C	0.15	-0.062 [-0.100; -0.023]	0.00155
rs7933504	*KCNJ5*	11	128783781	G	0.29	-0.047 [-0.077; -0.017]	0.0019
rs11178469	*PTPRR*	12	71275137	T	0.23	-0.050 [-0.082; -0.018]	0.00223
rs7207217	*BC039327_D43770*	17	69528353	A	0.38	-0.043 [-0.071; -0.015]	0.00234
rs931302	*NONE_SETMAR*	3	4210614	T	0.29	-0.045 [-0.074; -0.015]	0.00305
rs10187371	*ZEB2*	2	145226370	T	0.17	-0.050 [-0.086; -0.014]	0.00648
rs2823097	*NRIP1_USP25*	21	16523144	C	0.33	-0.039 [-0.068; -0.011]	0.00752
rs1532278	*CLU*	8	27466315	T	0.40	-0.037 [-0.064; -0.009]	0.00891
rs4894529	*FNDC3B*	3	171959684	G	0.49	-0.032 [-0.059; -0.005]	0.0189
rs7971334	*PDE3A*	12	20720707	T	0.30	-0.034 [-0.063; -0.004]	0.0246
rs7449443	*FLJ16171_DRD1*	5	174720893	T	0.40	-0.031 [-0.059; -0.003]	0.0286
rs11952819	*ZNF366*	5	71780033	T	0.28	-0.030 [-0.060; 0.000]	0.0503
rs9606967	*AK123891_SYN3*	22	32899516	C	0.21	-0.032 [-0.065; 0.001]	0.056
rs9680365	*GRIK1*	21	30928732	A	0.03	-0.062 [-0.136; 0.013]	0.105

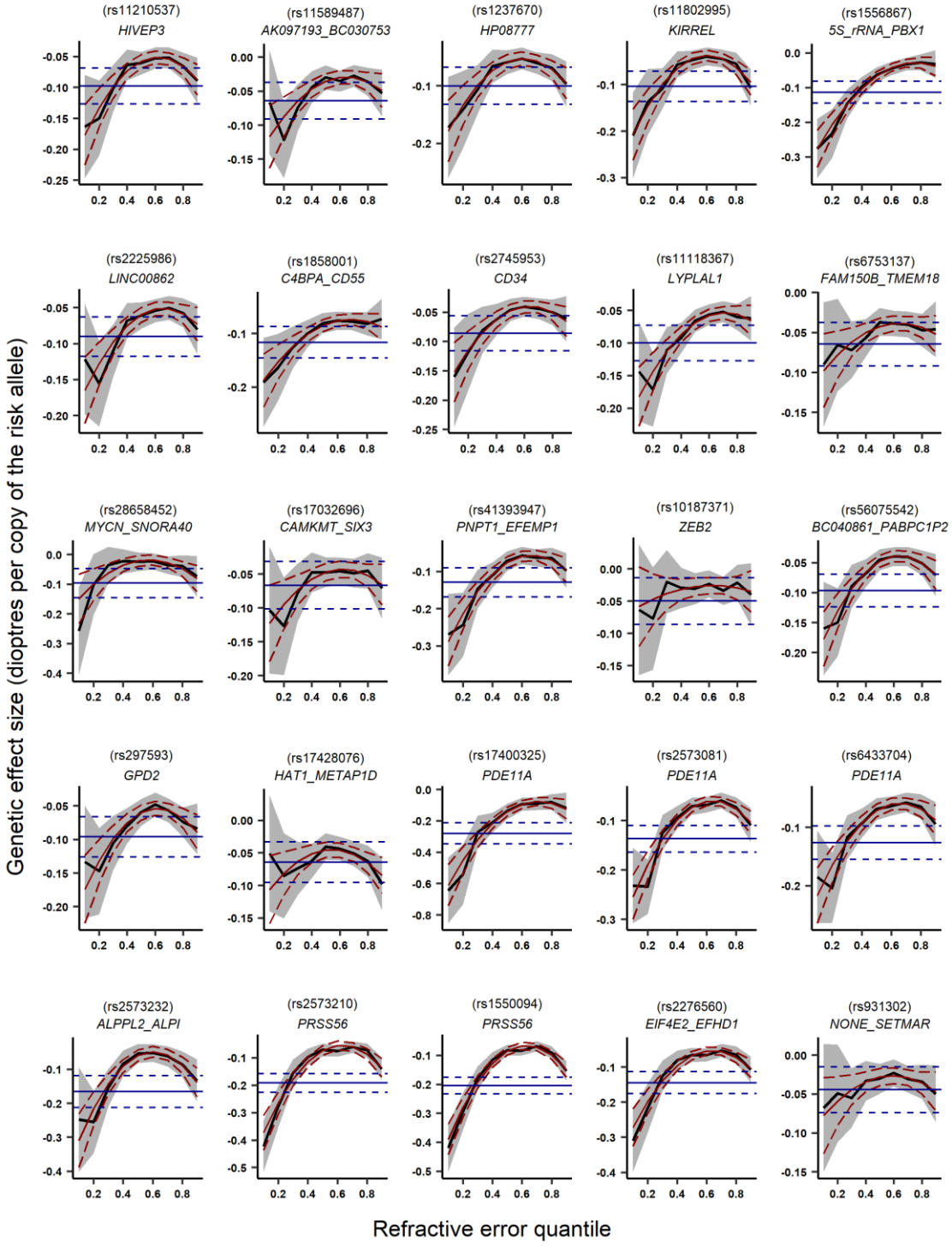
Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *MAF* - minor allele frequency, *Beta* - the effect size in dioptres per copy of the risk allele, *CI* - confidence interval.

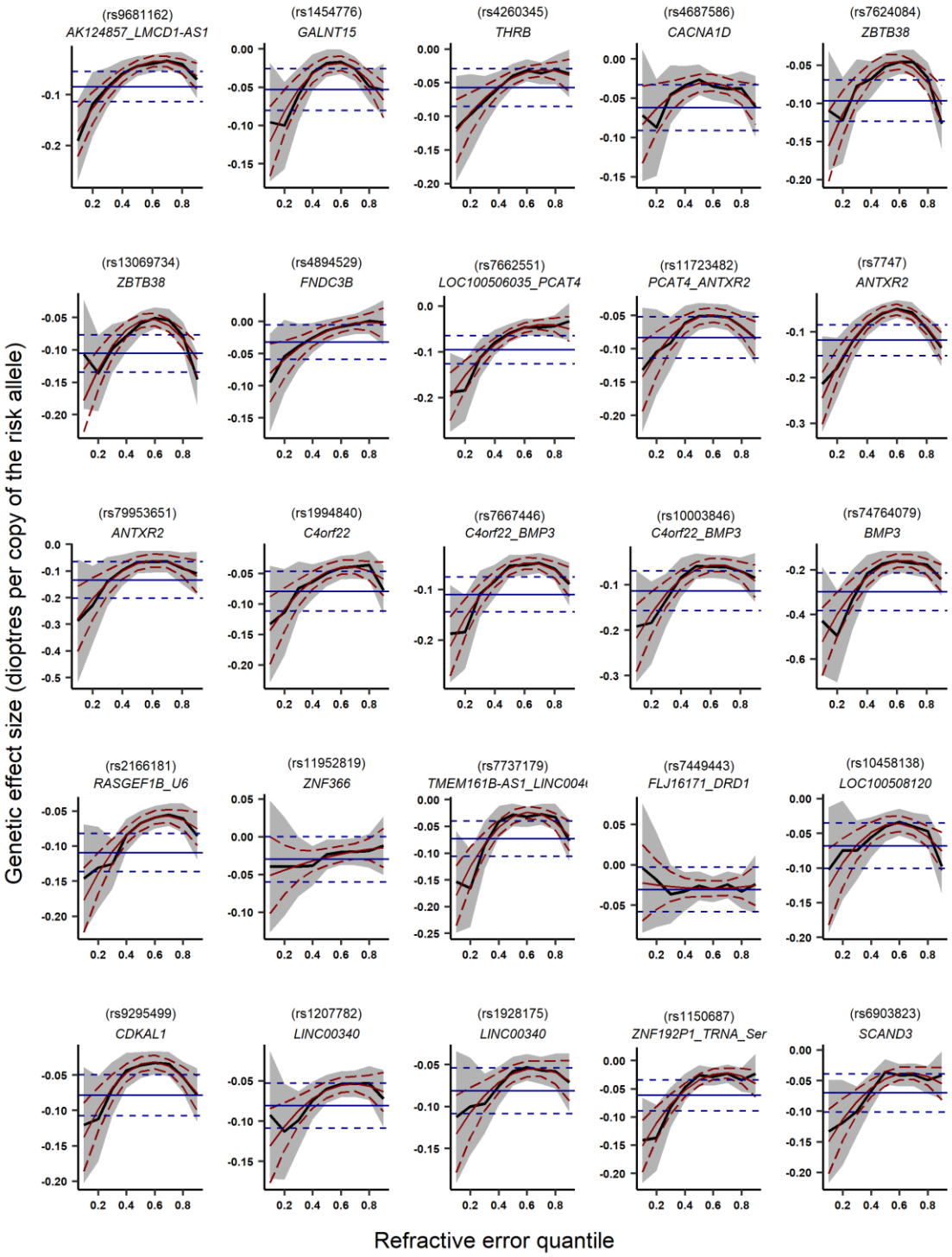
Appendix C: Type-1 error rate and statistical power for CQR-MR models with different number of quantiles.

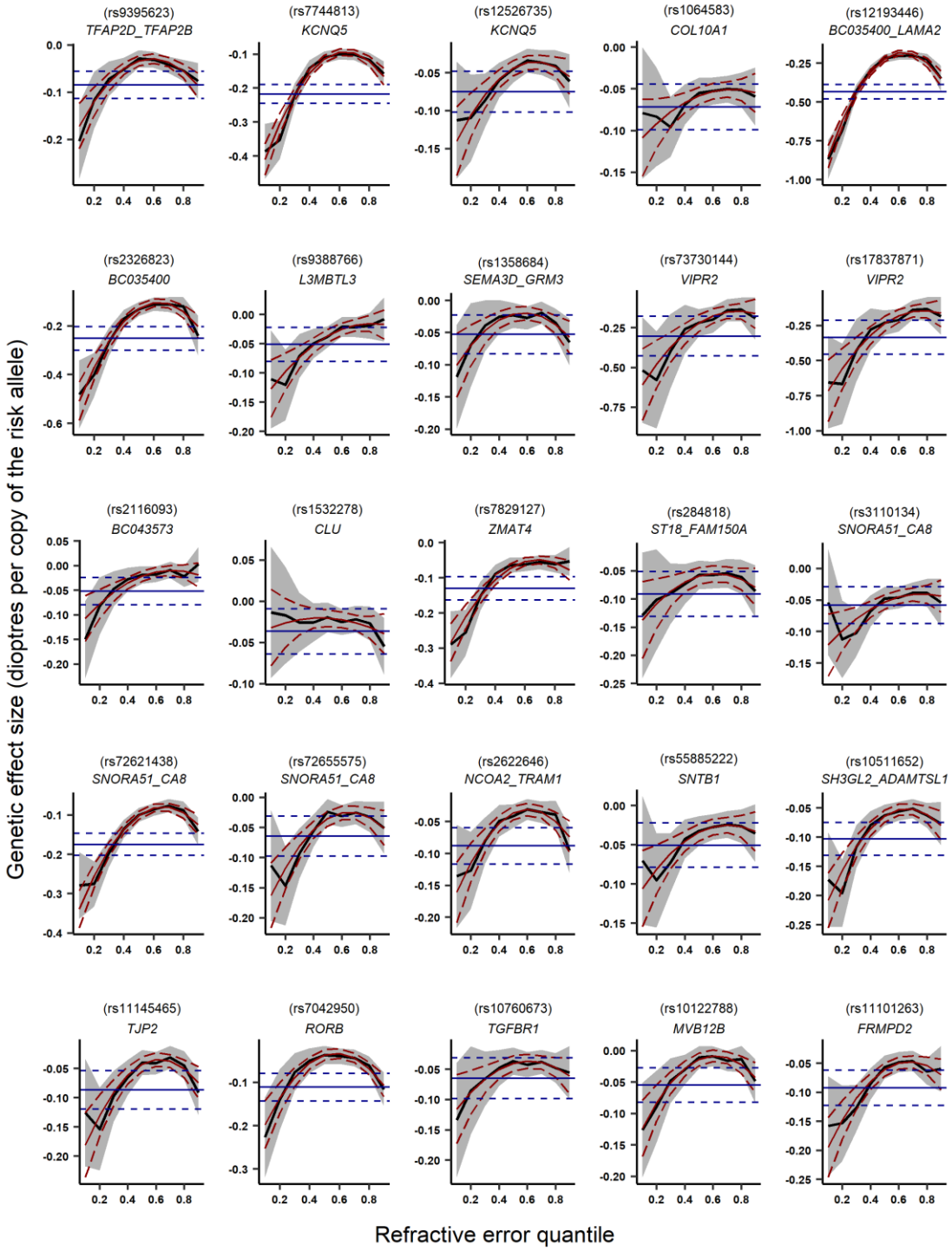


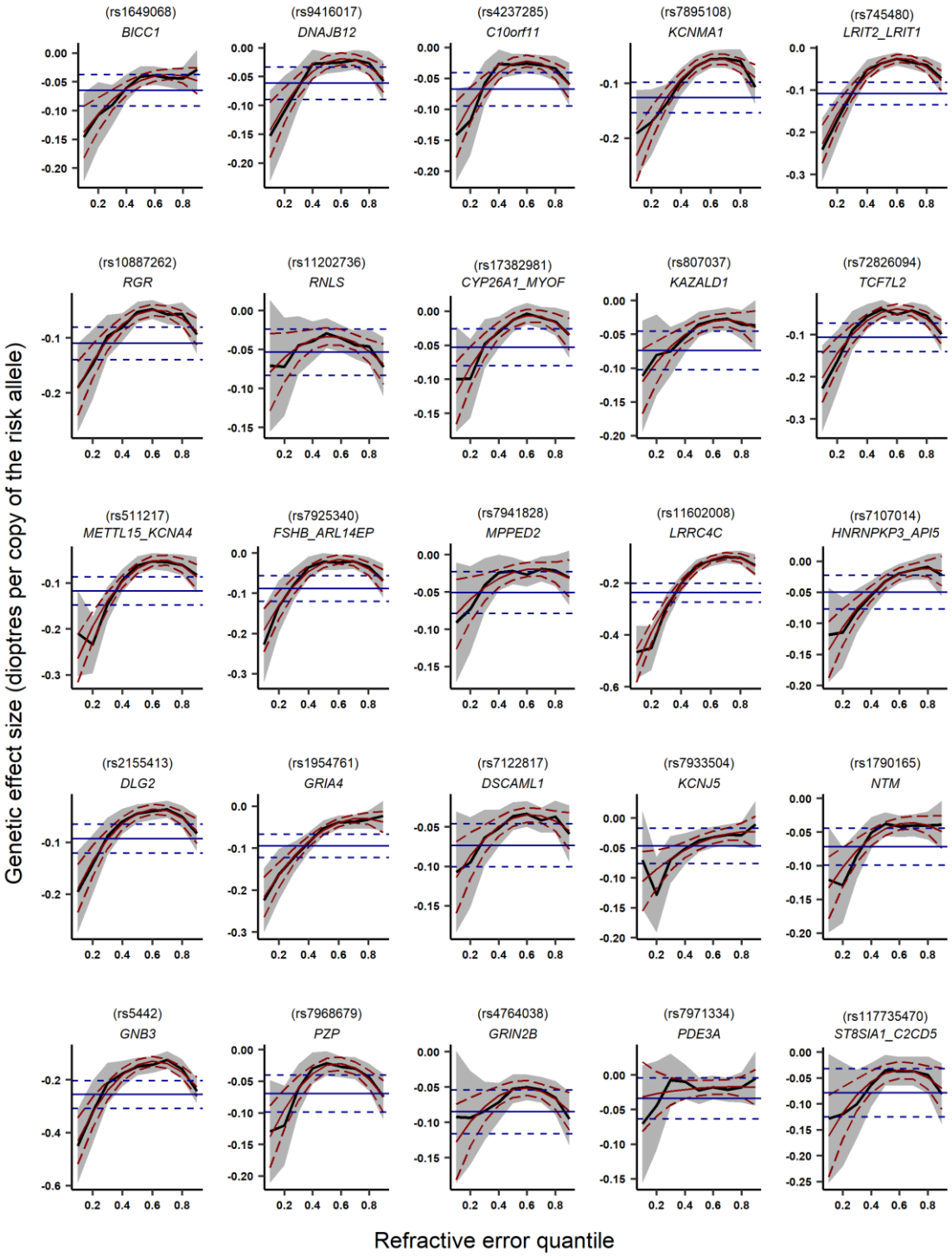
Type-1 error rate and statistical power for CQR-MR models with different number of quantiles. The top panel demonstrates a systematic inflation of type-1 error rate for all three components assessed in meta-regression. The middle panel shows type-1 error for CQR-MR results with respect to different minor allele frequencies. The bottom panel represents statistical power for different CQR-MR models. Note that the observed inflation of the type-1 error rate was corrected in the analysis of statistical power.

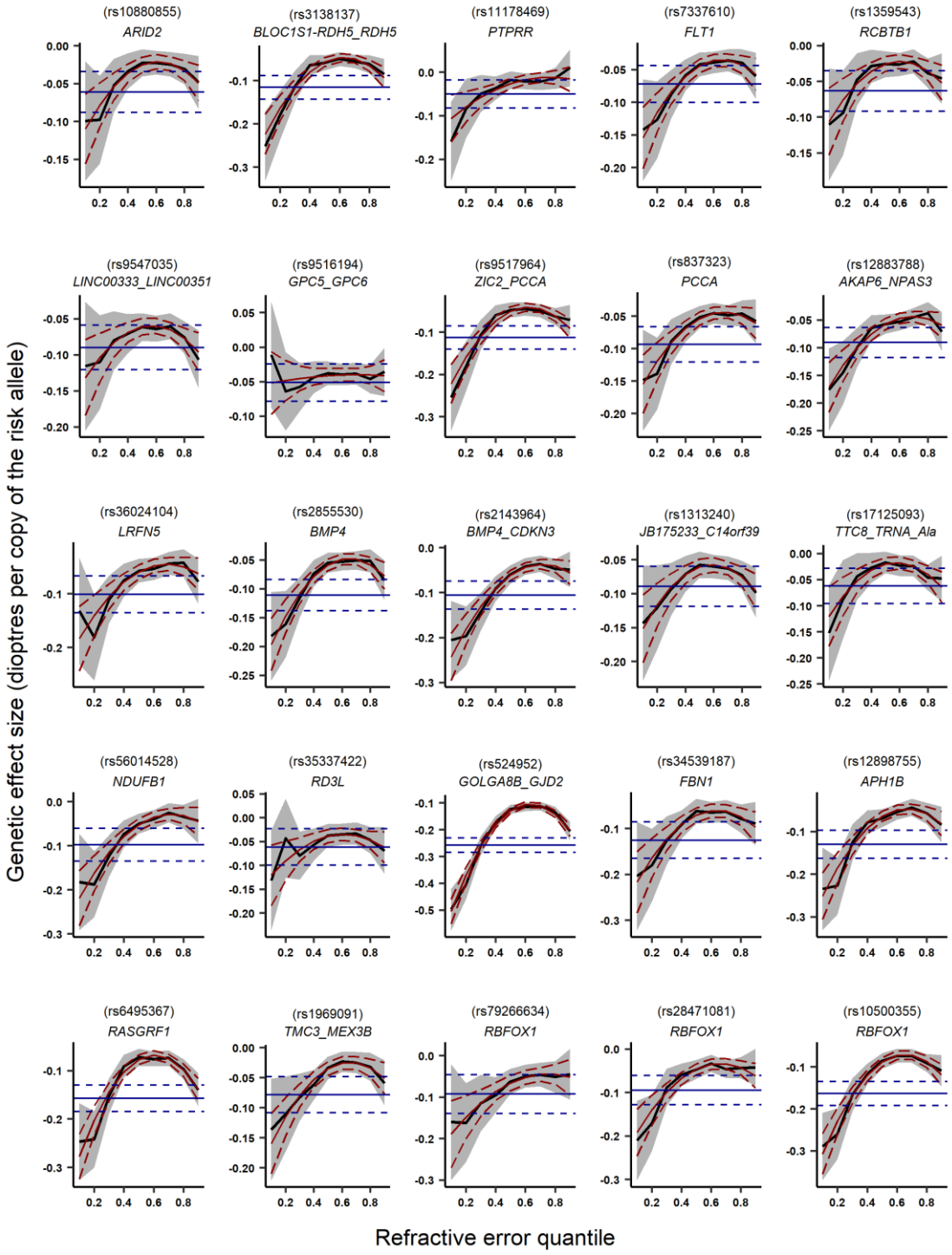
Appendix D: Distribution of effect sizes for all refractive error-associated variants that were obtained from *CREAM* meta-analysis

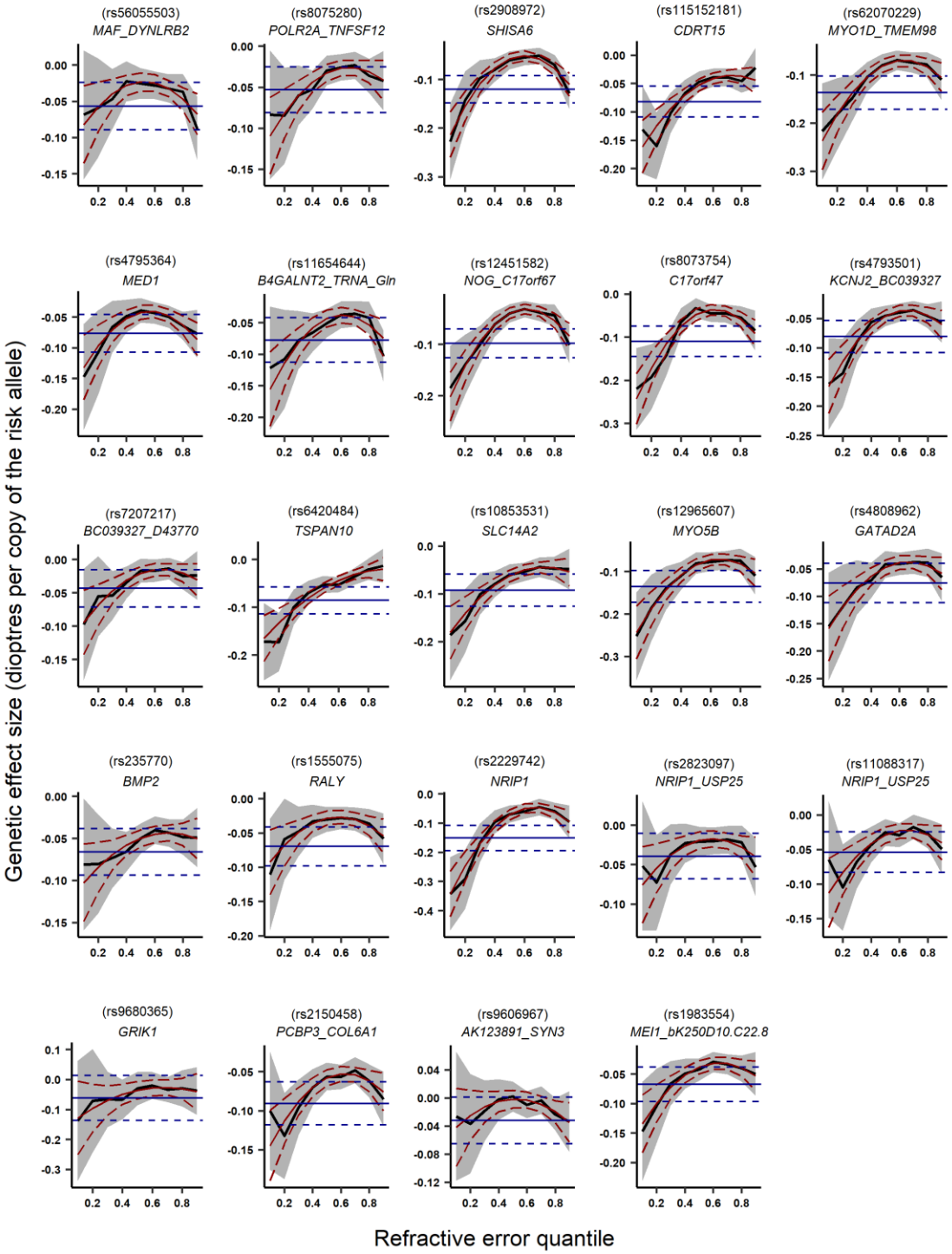



(rs9681162)
AK124857_LMCD1-AS1
(rs1454776)
GALNT15
(rs4260345)
THRB
(rs4687586)
CACNA1D
(rs7624084)
ZBTB38
(rs13069734)
ZBTB38
(rs4894529)
FNDC3B
(rs7662551)
LOC100506035_PCAT4
(rs11723482)
PCAT4_ANTXR2
(rs7747)
ANTXR2
(rs79953651)
ANTXR2
(rs1994840)
C4orf22
(rs7667446)
C4orf22_BMP3
(rs10003846)
C4orf22_BMP3
(rs74764079)
BMP3
(rs2166181)
RASGEF1B_U6
(rs11952819)
ZNF366
(rs7737179)
TMEM161B-AS1_LINC004
(rs7449443)
FLJ16171_DRD1
(rs10458138)
LOC100508120
(rs9295499)
CDKAL1
(rs1207782)
LINC00340
(rs1928175)
LINC00340
(rs1150687)
ZNF192P1_TRNA_Ser
(rs6903823)
SCAND3
Genetic effect size (dioptres per copy of the risk allele)
Refractive error quantile


(rs9395623)
TFAP2D_TFAP2B
(rs7744813)
KCNQ5
(rs12526735)
KCNQ5
(rs1064583)
COL10A1
(rs12193446)
BC035400_LAMA2
(rs2326823)
BC035400
(rs9388766)
L3MBTL3
(rs1358684)
SEMA3D_GRM3
(rs73730144)
VIPR2
(rs17837871)
VIPR2
(rs2116093)
BC043573
(rs1532278)
CLU
(rs7829127)
ZMAT4
(rs284818)
ST18_FAM150A
(rs3110134)
SNORA51_CA8
(rs72621438)
SNORA51_CA8
(rs72655575)
SNORA51_CA8
(rs2622646)
NCOA2_TRAM1
(rs55885222)
SNTB1
(rs10511652)
SH3GL2_ADAMTSL1
(rs11145465)
TJP2
(rs7042950)
RORB
(rs10760673)
TGFBR1
(rs10122788)
MVB12B
(rs11101263)
FRMPD2
Genetic effect size (dioptres per copy of the risk allele)
Refractive error quantile


(rs1649068)
BICC1
(rs9416017)
DNAJB12
(rs4237285)
C10orf11
(rs7895108)
KCNMA1
(rs745480)
LRIT2_LRIT1
(rs10887262)
RGR
(rs11202736)
RNLS
(rs17382981)
CYP26A1_MYOF
(rs807037)
KAZALD1
(rs72826094)
TCF7L2
(rs511217)
METTL15_KCNA4
(rs7925340)
FSHB_ARL14EP
(rs7941828)
MPPED2
(rs11602008)
LRRC4C
(rs7107014)
HNRNPKP3_API5
(rs2155413)
DLG2
(rs1954761)
GRIA4
(rs7122817)
DSCAML1
(rs7933504)
KCNJ5
(rs1790165)
NTM
(rs5442)
GNB3
(rs7968679)
PZP
(rs4764038)
GRIN2B
(rs7971334)
PDE3A
(rs117735470)
ST8SIA1_C2CD5
Genetic effect size (dioptres per copy of the risk allele)
Refractive error quantile


(rs10880855)
ARID2
(rs3138137)
BLOC1S1-RDH5_RDH5
(rs11178469)
PTPRR
(rs7337610)
FLT1
(rs1359543)
RCBTB1
(rs9547035)
LINC00333_LINC00351
(rs9516194)
GPC5_GPC6
(rs9517964)
ZIC2_PCCA
(rs837323)
PCCA
(rs12883788)
AKAP6_NPAS3
(rs36024104)
LRFN5
(rs2855530)
BMP4
(rs2143964)
BMP4_CDKN3
(rs1313240)
JB175233_C14orf39
(rs17125093)
TTC8_TRNA_Ala
(rs56014528)
NDUFB1
(rs35337422)
RD3L
(rs524952)
GOLGA8B_GJD2
(rs34539187)
FBN1
(rs12898755)
APH1B
(rs6495367)
RASGRF1
(rs1969091)
TMC3_MEX3B
(rs79266634)
RBFOX1
(rs28471081)
RBFOX1
(rs10500355)
RBFOX1
Genetic effect size (dioptres per copy of the risk allele)
Refractive error quantile


(rs56055503)
MAF_DYNLRB2
(rs8075280)
POLR2A_TNFSF12
(rs2908972)
SHISA6
(rs115152181)
CDRT15
(rs62070229)
MYO1D_TMEM98
(rs4795364)
MED1
(rs11654644)
B4GALNT2_TRNA_Gln
(rs12451582)
NOG_C17orf67
(rs8073754)
C17orf47
(rs4793501)
KCNJ2_BC039327
(rs7207217)
BC039327_D43770
(rs6420484)
TSPAN10
(rs10853531)
SLC14A2
(rs12965607)
MYO5B
(rs4808962)
GATAD2A
(rs235770)
BMP2
(rs1555075)
RALY
(rs2229742)
NRIP1
(rs2823097)
NRIP1_USP25
(rs11088317)
NRIP1_USP25
(rs9680365)
GRIK1
(rs2150458)
PCBP3_COL6A1
(rs9606967)
AK123891_SYN3
(rs1983554)
MEI1_bK250D10.C22.8
Genetic effect size (dioptres per copy of the risk allele)
Refractive error quantile

Appendix E: Summary statistics for CQR-MR effect estimates for variants associated with refractive error.

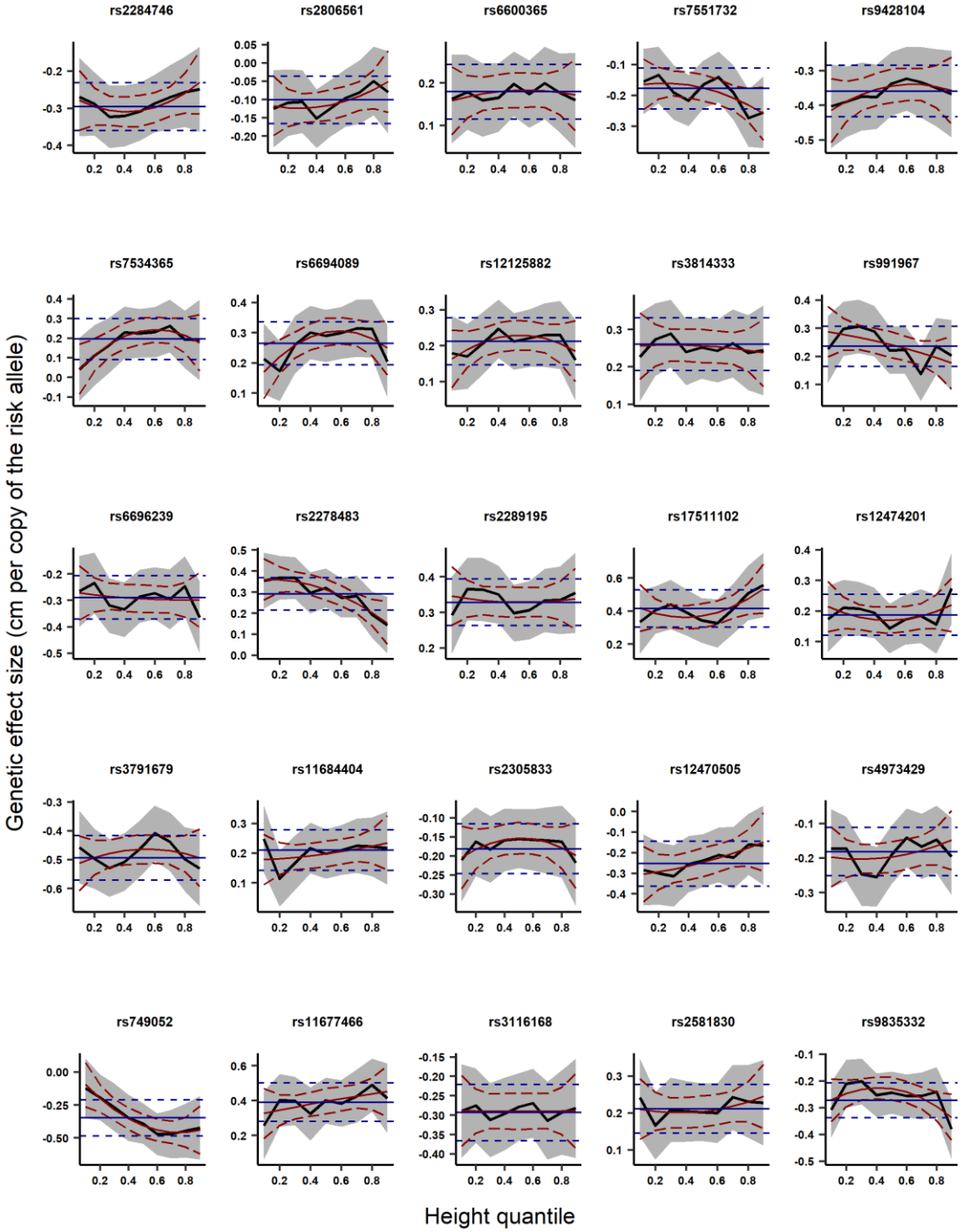
SNP	GENE	β_0		β_1		β_2	
		Beta [95% CI]	**P**	**Beta [95% CI]**	**P**	**Beta [95% CI]**	**P**
rs12193446	*BC035400_LAMA2*	-1.130 [-1.272;-0.988]	8.07E-55	2.995 [2.529;3.461]	2.12E-36	-2.363 [-2.765;-1.961]	1.19E-30
rs524952	*GOLGA8B_GJD2*	-0.673 [-0.758;-0.588]	4.83E-54	1.797 [1.534;2.060]	7.47E-41	-1.417 [-1.634;-1.200]	1.68E-37
rs7744813	*KCNQ5*	-0.543 [-0.631;-0.455]	7.24E-34	1.402 [1.132;1.672]	2.15E-24	-1.092 [-1.314;-0.807]	5.75E-22
rs11602008	*LRRC4C*	-0.669 [-0.790;-0.548]	2.60E-27	1.612 [1.250;1.974]	2.71E-18	-1.131 [-1.421;-0.841]	2.25E-14
rs1550094	*PRSS56*	-0.521 [-0.624;-0.418]	4.77E-23	1.441 [1.118;1.764]	2.08E-18	-1.142 [-1.409;-0.875]	4.90E-17
rs72621438	*SNORA51_CA8*	-0.441 [-0.530;-0.352]	2.06E-22	1.089 [0.817;1.361]	4.46E-15	-0.821 [-1.044;-0.598]	5.85E-13
rs2326823	*BC035400*	-0.680 [-0.830;-0.530]	6.17E-19	1.815 [1.341;2.289]	6.45E-14	-1.429 [-1.831;-1.027]	3.09E-12
rs10500355	*RBFOX1*	-0.004 [-0.490;-0.310]	3.63E-18	1.011 [0.734;1.288]	8.39E-13	-0.775 [-1.003;-0.547]	2.76E-11
rs6495367	*RASGRF1*	-0.374 [-0.459;-0.289]	7.17E-18	1.009 [0.747;1.271]	4.38E-14	-0.833 [-1.049;-0.617]	3.89E-14
rs2573210	*PRSS56*	-0.501 [-0.621;-0.381]	2.91E-16	1.414 [1.037;1.791]	1.94E-13	-1.121 [-1.434;-0.808]	2.26E-12
rs2573081	*PDE11A*	-0.339 [-0.426;-0.252]	1.69E-14	0.871 [0.605;1.137]	1.46E-10	-0.679 [-0.899;-0.459]	1.35E-09
rs745480	*LRIT2_LRIT1*	-0.312 [-0.395;-0.229]	1.68E-13	0.913 [0.658;1.168]	2.28E-12	-0.725 [-0.935;-0.515]	1.34E-11
rs2276560	*EIF4E2_EFHD1*	-0.365 [-0.464;-0.266]	5.12E-13	0.984 [0.683;1.285]	1.43E-10	-0.771 [-1.015;-0.527]	6.10E-10
rs7829127	*ZMAT4*	-0.369 [-0.470;-0.268]	9.91E-13	0.925 [0.611;1.239]	7.74E-09	-0.669 [-0.929;-0.409]	4.48E-07
rs1556867	*5S_rRNA_PBX1*	-0.361 [-0.461;-0.261]	1.77E-12	0.878 [0.569;1.187]	2.65E-08	-0.577 [-0.832;-0.322]	9.35E-06
rs9517964	*ZIC2_PCCA*	-0.296 [-0.381;-0.211]	9.60E-12	0.817 [0.554;1.080]	1.08E-09	-0.654 [-0.870;-0.438]	3.08E-09
rs511217	*METTL15_KCNA4*	-0.343 [-0.443;-0.243]	1.66E-11	0.882 [0.574;1.190]	1.90E-08	-0.664 [-0.917;-0.411]	2.69E-07
rs5442	*GNB3*	-0.561 [-0.728;-0.394]	4.15E-11	1.408 [0.918;1.898]	1.81E-08	-1.152 [-1.538;-0.766]	4.82E-09
rs7895108	*KCNMA1*	-0.299 [-0.388;-0.210]	4.96E-11	0.744 [0.471;1.017]	8.80E-08	-0.567 [-0.789;-0.345]	5.48E-07
rs3138137	*BLOC1S1-RDH5_RDH5*	-0.295 [-0.384;-0.206]	8.84E-11	0.788 [0.515;1.061]	1.46E-08	-0.625 [-0.848;-0.402]	3.96E-08
rs17400325	*PDE11A*	-0.803 [-1.050;-0.556]	1.97E-10	2.036 [1.283;2.789]	1.14E-07	-1.427 [-2.039;-0.815]	4.81E-06
rs2908972	*SHISA6*	-0.287 [-0.376;-0.198]	2.75E-10	0.790 [0.521;1.059]	8.74E-09	-0.661 [-0.879;-0.443]	2.82E-09
rs12898755	*APH1B*	-0.328 [-0.431;-0.225]	4.98E-10	0.846 [0.527;1.165]	2.08E-07	-0.637 [-0.901;-0.373]	2.25E-06
rs2143964	*BMP4_CDKN3*	-0.317 [-0.417;-0.217]	5.15E-10	0.788 [0.477;1.099]	6.79E-07	-0.554 [-0.811;-0.297]	2.38E-05
rs1954761	*GRIA4*	-0.276 [-0.363;-0.189]	5.18E-10	0.646 [0.377;0.915]	2.57E-06	-0.424 [-0.647;-0.201]	0.000188
rs12451582	*NOG_C17orf67*	-0.277 [-0.365;-0.189]	8.51E-10	0.807 [0.537;1.077]	4.57E-09	-0.660 [-0.881;-0.439]	4.45E-09
rs7747	*ANTXR2*	-0.332 [-0.438;-0.226]	9.18E-10	0.953 [0.625;1.281]	1.26E-08	-0.811 [-1.082;-0.540]	4.75E-09
rs6433704	*PDE11A*	-0.281 [-0.371;-0.191]	9.32E-10	0.677 [0.400;0.954]	1.73E-06	-0.514 [-0.743;-0.285]	1.06E-05
rs2229742	*NRIP1*	-0.460 [-0.608;-0.312]	1.06E-09	1.239 [0.794;1.684]	4.91E-08	-0.926 [-1.286;-0.566]	4.59E-07
rs2855530	*BMP4*	-0.256 [-0.341;-0.171]	3.94E-09	0.639 [0.377;0.901]	1.74E-06	-0.490 [-0.706;-0.274]	8.93E-06
rs41393947	*PNPT1_EFEMP1*	-0.379 [-0.506;-0.252]	4.48E-09	0.972 [0.586;1.358]	8.14E-07	-0.732 [-1.048;-0.416]	5.46E-06
rs10511652	*SH3GL2_ADAMTSL1*	-0.267 [-0.356;-0.178]	4.65E-09	0.654 [0.379;0.929]	3.05E-06	-0.496 [-0.722;-0.270]	1.64E-05
rs8073754	*C17orf47*	-0.323 [-0.434;-0.212]	1.31E-08	0.916 [0.577;1.255]	1.15E-07	-0.731 [-1.009;-0.453]	2.54E-07
rs2155413	*DLG2*	-0.253 [-0.340;-0.166]	1.34E-08	0.694 [0.426;0.962]	3.79E-07	-0.556 [-0.777;-0.335]	8.28E-07
rs2573232	*ALPPL2_ALPI*	-0.422 [-0.569;-0.275]	2.01E-08	1.237 [0.779;1.695]	1.22E-07	-1.024 [-1.404;-0.644]	1.24E-07
rs11802995	*KIRREL*	-0.282 [-0.382;-0.182]	3.13E-08	0.810 [0.509;1.111]	1.36E-07	-0.671 [-0.916;-0.426]	8.03E-08
rs56075542	*BC040861_PABPC1P2*	-0.233 [-0.317;-0.149]	5.20E-08	0.608 [0.350;0.866]	3.84E-06	-0.476 [-0.689;-0.263]	1.21E-05
rs10887262	*RGR*	-0.251 [-0.342;-0.160]	6.33E-08	0.650 [0.373;0.927]	4.40E-06	-0.522 [-0.749;-0.295]	6.46E-06
rs11118367	*LYPLAL1*	-0.226 [-0.309;-0.143]	1.03E-07	0.481 [0.224;0.738]	0.000245	-0.335 [-0.547;-0.123]	0.00196
rs62070229	*MYO1D_TMEM98*	-0.302 [-0.413;-0.191]	1.07E-07	0.717 [0.370;1.064]	5.00E-05	-0.552 [-0.839;-0.265]	0.000161
rs7667446	*C4orf22_BMP3*	-0.280 [-0.383;-0.177]	1.08E-07	0.727 [0.410;1.044]	7.09E-06	-0.566 [-0.827;-0.305]	2.15E-05
rs12965607	*MYO5B*	-0.313 [-0.430;-0.196]	1.48E-07	0.738 [0.381;1.095]	5.13E-05	-0.562 [-0.854;-0.270]	0.000166
rs2166181	*RASGEF1B_U6*	-0.222 [-0.305;-0.139]	1.64E-07	0.490 [0.234;0.746]	0.000174	-0.363 [-0.574;-0.152]	0.000727
rs12883788	*AKAP6_NPAS3*	-0.221 [-0.304;-0.138]	1.84E-07	0.512 [0.258;0.766]	7.96E-05	-0.370 [-0.578;-0.162]	0.000498
rs7042950	*RORB*	-0.274 [-0.378;-0.170]	2.28E-07	0.834 [0.520;1.148]	1.94E-07	-0.720 [-0.975;-0.465]	3.08E-08
rs11101263	*FRMPD2*	-0.251 [-0.347;-0.155]	3.32E-07	0.601 [0.305;0.897]	6.94E-05	-0.443 [-0.687;-0.199]	0.000367
rs13069734	*ZBTB38*	-0.237 [-0.328;-0.146]	3.71E-07	0.650 [0.367;0.933]	6.64E-06	-0.572 [-0.805;-0.339]	1.43E-06
rs9395623	*TFAP2D_TFAP2B*	-0.234 [-0.324;-0.144]	3.89E-07	0.695 [0.416;0.974]	1.06E-06	-0.588 [-0.818;-0.358]	5.20E-07
rs7662551	*LOC100506035_PCAT4*	-0.251 [-0.349;-0.153]	4.96E-07	0.570 [0.268;0.872]	0.000218	-0.384 [-0.634;-0.134]	0.00258
rs7925340	*FSHB_ARL14EP*	-0.267 [-0.372;-0.162]	6.68E-07	0.821 [0.496;1.146]	7.35E-07	-0.669 [-0.937;-0.401]	9.56E-07
rs4793501	*KCNJ2_BC039327*	-0.218 [-0.304;-0.132]	7.01E-07	0.555 [0.290;0.820]	4.04E-05	-0.423 [-0.641;-0.205]	0.000143
rs297593	*GPD2*	-0.227 [-0.318;-0.136]	1.11E-06	0.560 [0.279;0.841]	9.43E-05	-0.451 [-0.682;-0.220]	0.000131
rs1858001	*C4BPA_CD55*	-0.230 [-0.323;-0.137]	1.22E-06	0.453 [0.169;0.737]	0.00178	-0.329 [-0.564;-0.094]	0.006
rs56014528	*NDUFB1*	-0.287 [-0.404;-0.170]	1.40E-06	0.722 [0.362;1.082]	8.28E-05	-0.503 [-0.799;-0.207]	0.000869
rs11210537	*HIVEP3*	-0.231 [-0.325;-0.137]	1.53E-06	0.581 [0.293;0.869]	7.86E-05	-0.469 [-0.706;-0.232]	0.000108
rs1237670	*HP08777*	-0.234 [-0.330;-0.138]	1.57E-06	0.599 [0.304;0.894]	6.71E-05	-0.494 [-0.736;-0.252]	6.44E-05
rs11145465	*TJP2*	-0.244 [-0.344;-0.144]	1.71E-06	0.670 [0.363;0.977]	1.83E-05	-0.536 [-0.787;-0.285]	2.85E-05
rs72826094	*TCF7L2*	-0.272 [-0.384;-0.160]	1.79E-06	0.763 [0.424;1.102]	1.01E-05	-0.627 [-0.903;-0.351]	8.31E-06
rs2225986	*LINC00862*	-0.210 [-0.297;-0.123]	2.02E-06	0.487 [0.223;0.751]	3.00E-04	-0.372 [-0.587;-0.157]	0.000706
rs9681162	*AK124857_LMCD1-AS1*	-0.229 [-0.323;-0.135]	2.02E-06	0.604 [0.313;0.895]	4.78E-05	-0.468 [-0.709;-0.227]	0.000137
rs28471081	*RBFOX1*	-0.252 [-0.356;-0.148]	2.10E-06	0.658 [0.336;0.980]	6.21E-05	-0.496 [-0.761;-0.231]	0.000245
rs2622646	*NCOA2_TRAM1*	-0.218 [-0.308;-0.128]	2.11E-06	0.616 [0.338;0.894]	1.45E-05	-0.506 [-0.736;-0.276]	1.60E-05
rs7624084	*ZBTB38*	-0.209 [-0.296;-0.122]	2.32E-06	0.578 [0.310;0.846]	2.37E-05	-0.508 [-0.729;-0.287]	6.52E-06
rs1150687	*ZNF192P1_TRNA_Ser*	-0.203 [-0.288;-0.118]	2.54E-06	0.533 [0.274;0.792]	5.55E-05	-0.394 [-0.608;-0.180]	0.00031
rs837323	*PCCA*	-0.197 [-0.280;-0.114]	3.24E-06	0.453 [0.197;0.709]	0.000519	-0.336 [-0.548;-0.124]	0.0019
rs7337610	*FLT1*	-0.202 [-0.289;-0.115]	4.73E-06	0.505 [0.238;0.772]	0.000214	-0.382 [-0.603;-0.161]	0.000696
rs7737179	*TMEM161B-AS1_LINC00461*	-0.248 [-0.354;-0.142]	4.82E-06	0.724 [0.401;1.047]	1.09E-05	-0.583 [-0.845;-0.321]	1.31E-05
rs6420484	*TSPAN10*	-0.203 [-0.291;-0.115]	6.81E-06	0.388 [0.115;0.661]	0.0054	-0.207 [-0.434;0.020]	0.0743

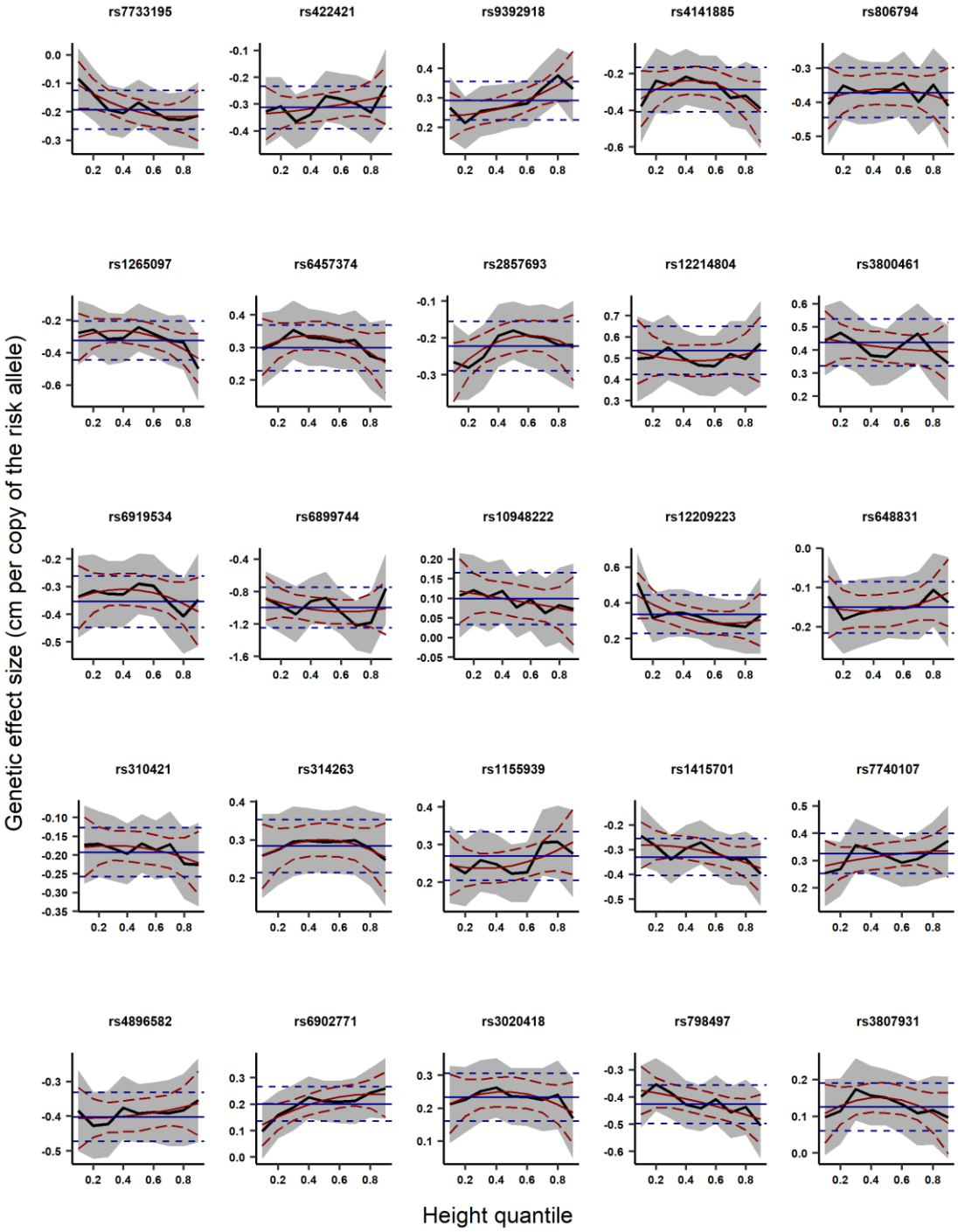
rs1969091	*TMC3_MEX3B*	-0.213 [-0.308;-0.118]	1.09E-05	0.576 [0.283;0.869]	0.000118	-0.440 [-0.683;-0.197]	0.000382
rs115152181	*CDRT15*	-0.199 [-0.288;-0.110]	1.10E-05	0.444 [0.173;0.715]	0.00134	-0.300 [-0.523;-0.077]	0.00847
rs7968679	*PZP*	-0.193 [-0.281;-0.105]	1.69E-05	0.595 [0.327;0.863]	1.37E-05	-0.515 [-0.734;-0.296]	4.18E-06
rs34539187	*FBN1*	-0.281 [-0.410;-0.152]	1.84E-05	0.693 [0.298;1.088]	0.000589	-0.544 [-0.868;-0.220]	0.000984
rs4237285	*C10orf11*	-0.182 [-0.265;-0.099]	1.90E-05	0.525 [0.269;0.781]	5.68E-05	-0.433 [-0.644;-0.222]	5.62E-05
rs9416017	*DNAJB12*	-0.193 [-0.282;-0.104]	1.95E-05	0.563 [0.293;0.833]	4.50E-05	-0.453 [-0.673;-0.233]	5.33E-05
rs10853531	*SLC14A2*	-0.228 [-0.333;-0.123]	2.17E-05	0.503 [0.177;0.829]	0.00247	-0.344 [-0.614;-0.074]	0.0124
rs7107014	*HNRNPKP3_API5*	-0.182 [-0.266;-0.098]	2.34E-05	0.440 [0.179;0.701]	0.000953	-0.283 [-0.499;-0.067]	0.0102
rs2150458	*PCBP3_COL6A1*	-0.182 [-0.267;-0.097]	2.57E-05	0.411 [0.149;0.673]	0.00215	-0.326 [-0.543;-0.109]	0.00326
rs1313240	*JB175233_C14orf39*	-0.195 [-0.286;-0.104]	2.69E-05	0.468 [0.186;0.750]	0.00114	-0.399 [-0.632;-0.166]	0.00079
rs2745953	*CD34*	-0.198 [-0.291;-0.105]	3.26E-05	0.492 [0.206;0.778]	0.000738	-0.384 [-0.619;-0.149]	0.00137
rs72655575	*SNORA51_CA8*	-0.214 [-0.316;-0.112]	3.69E-05	0.578 [0.263;0.893]	0.000323	-0.441 [-0.702;-0.180]	0.000921
rs12526735	*KCNQ5*	-0.181 [-0.267;-0.095]	3.78E-05	0.442 [0.176;0.708]	0.00115	-0.335 [-0.555;-0.115]	0.00287
rs1649068	*BICC1*	-0.175 [-0.260;-0.090]	5.90E-05	0.397 [0.132;0.662]	0.00337	-0.285 [-0.506;-0.064]	0.0115
rs10003846	*C4orf22_BMP3*	-0.280 [-0.417;-0.143]	5.93E-05	0.695 [0.281;1.109]	0.000988	-0.540 [-0.879;-0.201]	0.00177
rs9295499	*CDKAL1*	-0.185 [-0.275;-0.095]	6.06E-05	0.527 [0.248;0.806]	0.000217	-0.451 [-0.683;-0.219]	0.000138
rs36024104	*LRFN5*	-0.233 [-0.347;-0.119]	6.16E-05	0.546 [0.201;0.891]	0.00191	-0.397 [-0.677;-0.117]	0.00545
rs10122788	*MVB12B*	-0.172 [-0.257;-0.087]	7.27E-05	0.533 [0.271;0.795]	6.77E-05	-0.433 [-0.650;-0.216]	9.47E-05
rs1790165	*NTM*	-0.171 [-0.257;-0.085]	8.90E-05	0.398 [0.133;0.663]	0.0032	-0.292 [-0.511;-0.073]	0.00889
rs17382981	*CYP26A1_MYOF*	-0.166 [-0.249;-0.083]	9.26E-05	0.503 [0.248;0.758]	0.000108	-0.397 [-0.608;-0.186]	0.000222
rs6903823	*SCAND3*	-0.194 [-0.292;-0.096]	9.72E-05	0.466 [0.166;0.766]	0.00236	-0.347 [-0.595;-0.099]	0.00604
rs1454776	*GALNT15*	-0.169 [-0.255;-0.083]	0.000126	0.524 [0.257;0.791]	0.000123	-0.457 [-0.679;-0.235]	5.37E-05
rs1983554	*MEI1_bK250D10.C22.8*	-0.176 [-0.269;-0.083]	0.000195	0.448 [0.163;0.733]	0.00206	-0.347 [-0.582;-0.112]	0.00378
rs11654644	*B4GALNT2_TRNA_Gln*	-0.207 [-0.317;-0.097]	0.000239	0.557 [0.219;0.895]	0.00124	-0.458 [-0.733;-0.183]	0.00111
rs1928175	*LINC00340*	-0.163 [-0.252;-0.074]	0.000311	0.329 [0.057;0.601]	0.0178	-0.250 [-0.475;-0.025]	0.0294
rs11723482	*PCAT4_ANTXR2*	-0.184 [-0.284;-0.084]	0.000321	0.457 [0.15;0.764]	0.00352	-0.387 [-0.639;-0.135]	0.00264
rs4808962	*GATAD2A*	-0.206 [-0.318;-0.094]	0.00033	0.512 [0.169;0.855]	0.00342	-0.387 [-0.667;-0.107]	0.00682
rs1994840	*C4orf22*	-0.188 [-0.292;-0.084]	0.000409	0.458 [0.145;0.771]	0.0041	-0.353 [-0.605;-0.101]	0.00597
rs11589487	*AK097193_BC030753*	-0.155 [-0.241;-0.069]	0.000417	0.393 [0.126;0.660]	0.00396	-0.305 [-0.527;-0.083]	0.00704
rs1207782	*LINC00340*	-0.158 [-0.248;-0.068]	0.000614	0.305 [0.028;0.582]	0.0311	-0.222 [-0.450;0.006]	0.0568
rs10880855	*ARID2*	-0.148 [-0.233;-0.063]	0.000636	0.415 [0.151;0.679]	0.00204	-0.339 [-0.557;-0.121]	0.00232
rs7122817	*DSCAML1*	-0.146 [-0.23;-0.062]	0.000662	0.355 [0.095;0.615]	0.0074	-0.282 [-0.497;-0.067]	0.0101
rs4795364	*MED1*	-0.174 [-0.275;-0.073]	0.00076	0.470 [0.160;0.780]	0.00301	-0.414 [-0.670;-0.158]	0.00151
rs4260345	*THRB*	-0.154 [-0.244;-0.064]	0.000777	0.340 [0.065;0.615]	0.0154	-0.236 [-0.462;-0.010]	0.0406
rs807037	*KAZALD1*	-0.152 [-0.241;-0.063]	0.000793	0.359 [0.085;0.633]	0.0103	-0.261 [-0.487;-0.035]	0.0238
rs2116093	*BC043573*	-0.147 [-0.233;-0.061]	0.000816	0.390 [0.124;0.656]	0.00406	-0.277 [-0.497;-0.057]	0.0138
rs9388766	*L3MBTL3*	-0.160 [-0.254;-0.066]	0.000872	0.356 [0.067;0.645]	0.0156	-0.219 [-0.457;0.019]	0.0708
rs9547035	*LINC00333_LINC00351*	-0.166 [-0.264;-0.068]	0.000894	0.375 [0.073;0.677]	0.015	-0.330 [-0.579;-0.081]	0.00946
rs4764038	*GRIN2B*	-0.161 [-0.258;-0.064]	0.00116	0.376 [0.078;0.674]	0.0135	-0.323 [-0.569;-0.077]	0.01
rs8075280	*POLR2A_TNFSF12*	-0.143 [-0.231;-0.055]	0.00136	0.357 [0.089;0.625]	0.00914	-0.272 [-0.492;-0.052]	0.0152
rs79953651	*ANTXR2*	-0.371 [-0.598;-0.144]	0.00137	0.994 [0.306;1.682]	0.0046	-0.794 [-1.355;-0.233]	0.00558
rs11088317	*NRIP1_USP25*	-0.151 [-0.244;-0.058]	0.00153	0.392 [0.106;0.678]	0.00718	-0.299 [-0.533;-0.065]	0.0121
rs3110134	*SNORA51_CA8*	-0.145 [-0.235;-0.055]	0.00163	0.272 [-0.01;0.554]	0.059	-0.176 [-0.412;0.060]	0.144
rs10458138	*LOC100508120*	-0.170 [-0.276;-0.064]	0.00172	0.460 [0.136;0.784]	0.00534	-0.396 [-0.66;-0.132]	0.00331
rs17125093	*TTC8_TRNA_Ala*	-0.169 [-0.275;-0.063]	0.00182	0.528 [0.201;0.855]	0.00157	-0.457 [-0.726;-0.188]	0.000855
rs79266634	*RBFOX1*	-0.233 [-0.38;-0.086]	0.0019	0.5 [0.054;0.946]	0.0282	-0.332 [-0.698;0.034]	0.0751
rs1359543	*RCBTB1*	-0.143 [-0.234;-0.052]	0.002	0.402 [0.125;0.679]	0.00451	-0.334 [-0.563;-0.105]	0.00419
rs1064583	*COL10A1*	-0.126 [-0.208;-0.044]	0.00258	0.201 [-0.056;0.458]	0.126	-0.134 [-0.348;0.08]	0.22
rs1358684	*SEMA3D_GRM3*	-0.14 [-0.232;-0.048]	0.0029	0.418 [0.135;0.701]	0.00378	-0.363 [-0.596;-0.13]	0.00226
rs55885222	*SNTB1*	-0.133 [-0.221;-0.045]	0.0031	0.305 [0.035;0.575]	0.0267	-0.216 [-0.438;0.006]	0.0562
rs117735470	*ST8SIA1_C2CD5*	-0.216 [-0.36;-0.072]	0.00335	0.585 [0.147;1.023]	0.00893	-0.472 [-0.833;-0.111]	0.0105
rs17428076	*HAT1_METAP1D*	-0.139 [-0.235;-0.043]	0.00458	0.348 [0.053;0.643]	0.0209	-0.321 [-0.565;-0.077]	0.00984
rs10760673	*TGFBR1*	-0.152 [-0.258;-0.046]	0.00483	0.373 [0.046;0.7]	0.0255	-0.303 [-0.575;-0.031]	0.0292
rs235770	*BMP2*	-0.123 [-0.209;-0.037]	0.00498	0.228 [-0.038;0.494]	0.0924	-0.163 [-0.383;0.057]	0.146
rs1555075	*RALY*	-0.124 [-0.211;-0.037]	0.00527	0.333 [0.066;0.6]	0.0145	-0.285 [-0.505;-0.065]	0.0113
rs7207217	*BC039327_D43770*	-0.126 [-0.215;-0.037]	0.00566	0.342 [0.068;0.616]	0.0145	-0.262 [-0.487;-0.037]	0.0224
rs6753137	*FAM150B_TMEM18*	-0.121 [-0.207;-0.035]	0.00598	0.258 [-0.006;0.522]	0.0557	-0.201 [-0.418;0.016]	0.0691
rs17032696	*CAMKMT_SIX3*	-0.156 [-0.267;-0.045]	0.00601	0.366 [0.025;0.707]	0.0356	-0.296 [-0.576;-0.016]	0.0382
rs7933504	*KCNJ5*	-0.13 [-0.223;-0.037]	0.00632	0.252 [-0.037;0.541]	0.0874	-0.149 [-0.388;0.09]	0.222
rs28658452	*MYCN_SNORA40*	-0.214 [-0.373;-0.055]	0.00848	0.679 [0.185;1.173]	0.00703	-0.59 [-1.001;-0.179]	0.00488
rs284818	*ST18_FAM150A*	-0.172 [-0.3;-0.044]	0.00856	0.379 [-0.011;0.769]	0.0567	-0.306 [-0.623;0.011]	0.0585
rs35337422	*RD3L*	-0.156 [-0.275;-0.037]	0.0103	0.398 [0.034;0.762]	0.0322	-0.328 [-0.628;-0.028]	0.0322
rs11178469	*PTPRR*	-0.132 [-0.235;-0.029]	0.0121	0.313 [-0.004;0.63]	0.053	-0.203 [-0.463;0.057]	0.126
rs4894529	*FNDC3B*	-0.104 [-0.191;-0.017]	0.0185	0.259 [-0.008;0.526]	0.0568	-0.163 [-0.384;0.058]	0.148
rs4687586	*CACNA1D*	-0.109 [-0.201;-0.017]	0.0201	0.28 [-0.006;0.566]	0.0549	-0.248 [-0.485;-0.011]	0.0404
rs7941828	*MPPED2*	-0.105 [-0.194;-0.016]	0.0208	0.265 [-0.012;0.542]	0.061	-0.204 [-0.434;0.026]	0.0823
rs2823097	*NRIP1_USP25*	-0.104 [-0.194;-0.014]	0.0237	0.294 [0.019;0.569]	0.0364	-0.249 [-0.475;-0.023]	0.0306
rs56055503	*MAF_DYNLRB2*	-0.114 [-0.215;-0.013]	0.0263	0.347 [0.034;0.66]	0.0297	-0.33 [-0.59;-0.07]	0.0128
rs931302	*NONE_SETMAR*	-0.101 [-0.193;-0.009]	0.0306	0.252 [-0.029;0.533]	0.0787	-0.213 [-0.443;0.017]	0.0695
rs11202736	*RNLS*	-0.105 [-0.2;-0.01]	0.0307	0.273 [-0.021;0.567]	0.0689	-0.259 [-0.503;-0.015]	0.0373
rs9680365	*GRIK1*	-0.167 [-0.394;0.06]	0.15	0.414 [-0.28;1.108]	0.242	-0.303 [-0.869;0.263]	0.294
rs9516194	*GPC5_GPC6*	-0.056 [-0.139;0.027]	0.189	0.047 [-0.214;0.308]	0.724	-0.033 [-0.246;0.18]	0.762
rs11952819	*ZNF366*	-0.058 [-0.152;0.036]	0.227	0.078 [-0.213;0.369]	0.599	-0.033 [-0.272;0.206]	0.787
rs10187371	*ZEB2*	-0.072 [-0.189;0.045]	0.228	0.148 [-0.213;0.509]	0.421	-0.119 [-0.417;0.179]	0.433
rs9606967	*AK123891_SYN3*	-0.063 [-0.169;0.043]	0.244	0.238 [-0.087;0.563]	0.152	-0.231 [-0.498;0.036]	0.0896
rs1532278	*CLU*	-0.040 [-0.128;0.048]	0.372	0.087 [-0.182;0.356]	0.527	-0.096 [-0.314;0.122]	0.389
rs7971334	*PDE3A*	-0.036 [-0.132;0.060]	0.462	0.051 [-0.244;0.346]	0.734	-0.035 [-0.282;0.212]	0.781

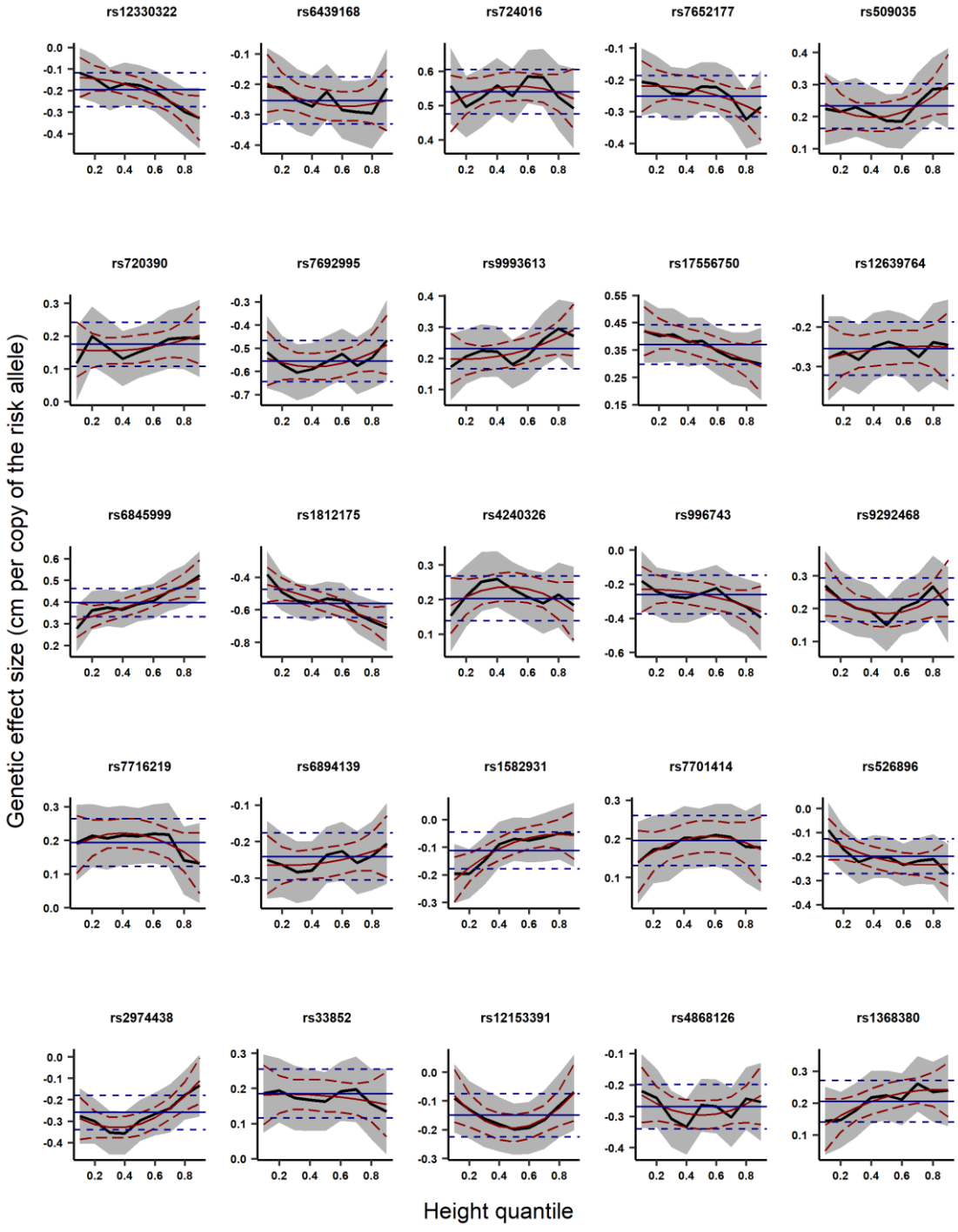
rs7449443	*FLJ16171_DRD1*	-0.019 [-0.106;0.068]	0.668	-0.037 [-0.308;0.234]	0.789	0.0310 [-0.188;0.250]	0.782

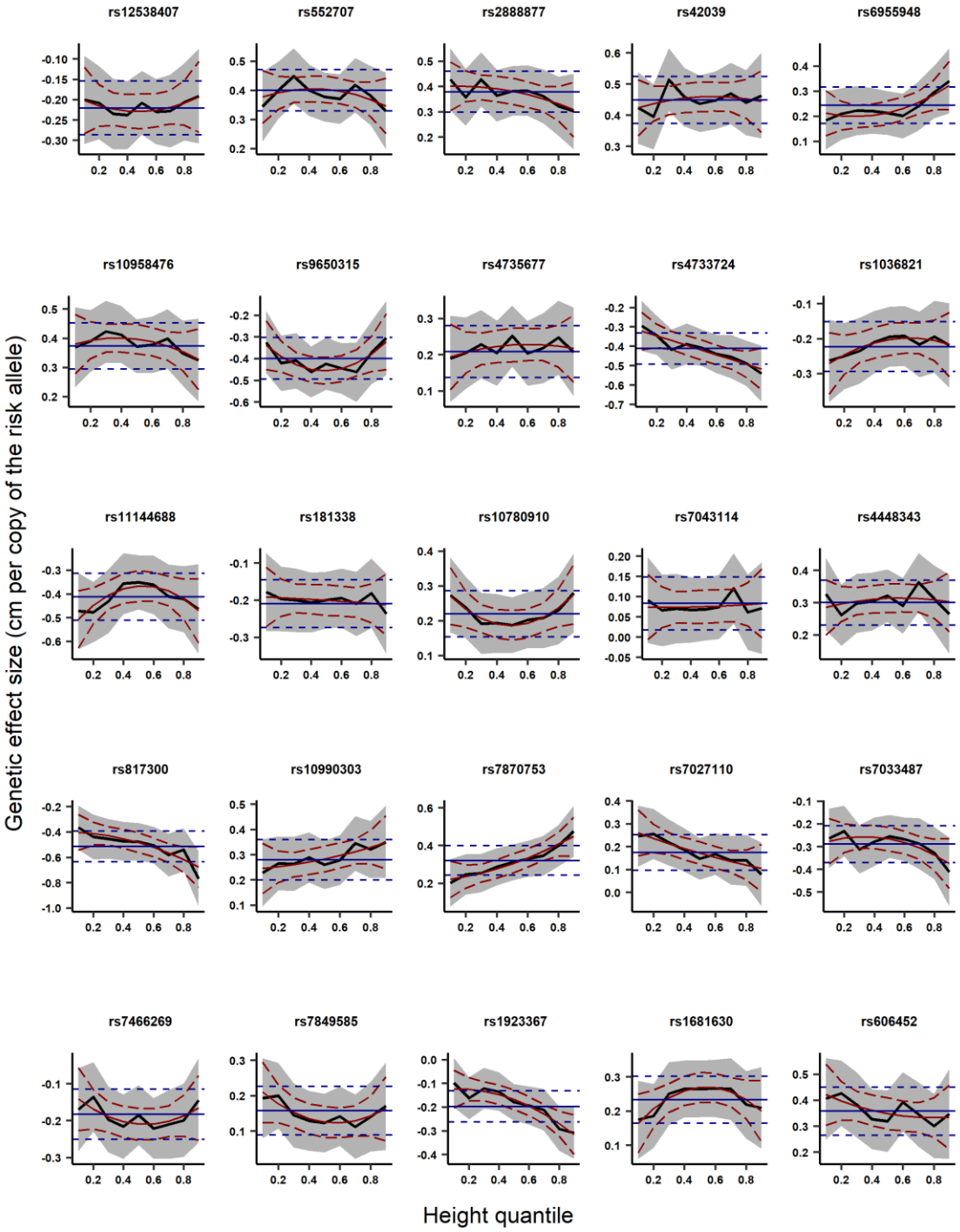
Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *CI* - confidence interval, β_0 - meta-regression intercept effect size in dioptres per copy of the risk allele, β_1 - meta-regression coefficients for the linear term and β_2 - meta-regression coefficients for the quadratic term.

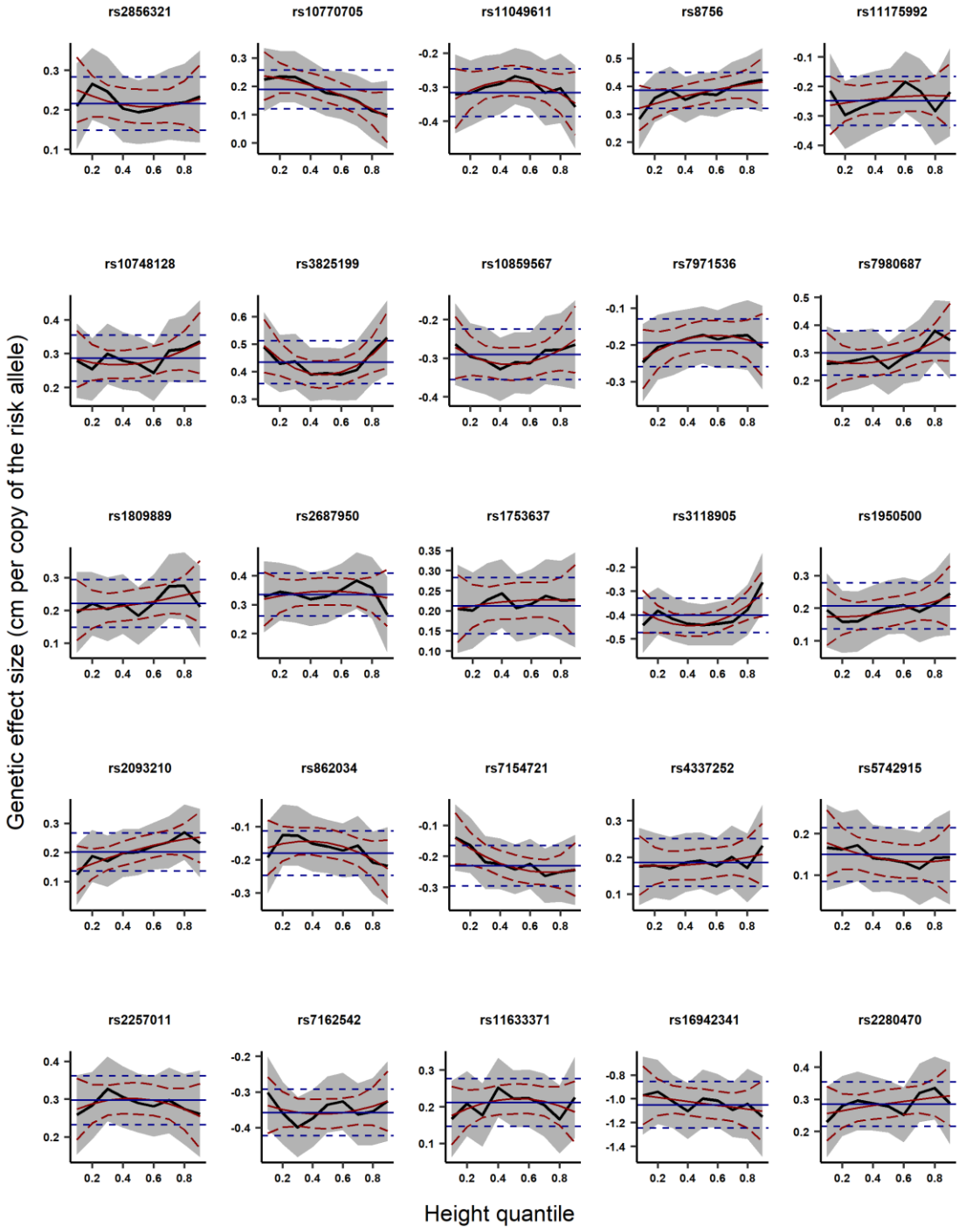
Appendix F: Distribution of effect sizes for all height-associated variants that were obtained from *GIANT* Consortium

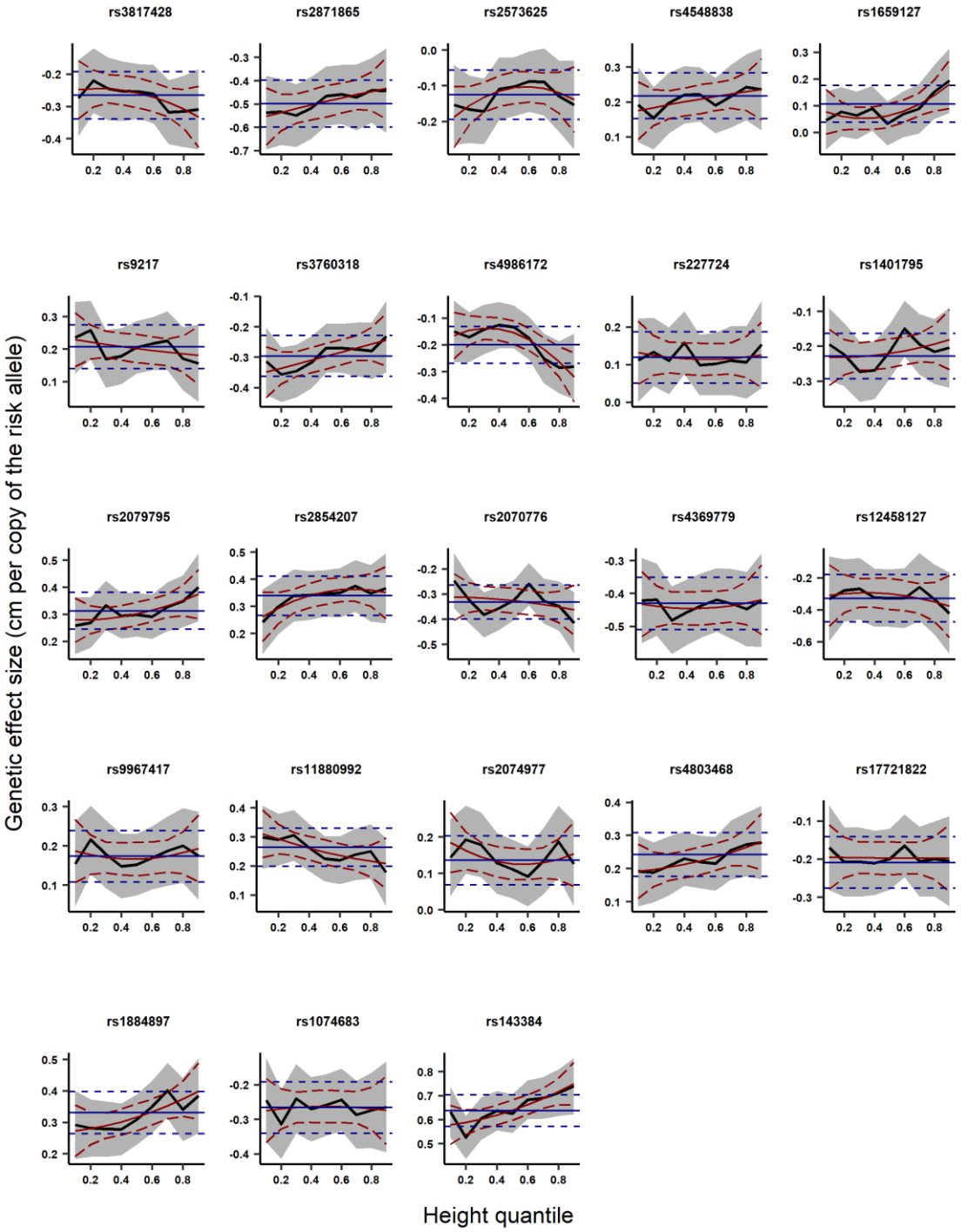



rs7733195
rs422421
rs9392918
rs4141885
rs806794
rs1265097
rs6457374
rs2857693
rs12214804
rs3800461
rs6919534
rs6899744
rs10948222
rs12209223
rs648831
rs310421
rs314263
rs1155939
rs1415701
rs7740107
rs4896582
rs6902771
rs3020418
rs798497
rs3807931
Genetic effect size (cm per copy of the risk allele)
Height quantile


rs12330322
rs6439168
rs724016
rs7652177
rs509035
rs720390
rs7692995
rs9993613
rs17556750
rs12639764
rs6845999
rs1812175
rs4240326
rs996743
rs9292468
rs7716219
rs6894139
rs1582931
rs7701414
rs526896
rs2974438
rs33852
rs12153391
rs4868126
rs1368380
Genetic effect size (cm per copy of the risk allele)
Height quantile


rs12538407
rs552707
rs2888877
rs42039
rs6955948
rs10958476
rs9650315
rs4735677
rs4733724
rs1036821
rs11144688
rs181338
rs10780910
rs7043114
rs4448343
rs817300
rs10990303
rs7870753
rs7027110
rs7033487
rs7466269
rs7849585
rs1923367
rs1681630
rs606452
Genetic effect size (cm per copy of the risk allele)
Height quantile


rs2856321
rs10770705
rs11049611
rs8756
rs11175992
rs10748128
rs3825199
rs10859567
rs7971536
rs7980687
rs1809889
rs2687950
rs1753637
rs3118905
rs1950500
rs2093210
rs862034
rs7154721
rs4337252
rs5742915
rs2257011
rs7162542
rs11633371
rs16942341
rs2280470
Genetic effect size (cm per copy of the risk allele)
Height quantile


rs3817428
rs2871865
rs2573625
rs4548838
rs1659127
rs9217
rs3760318
rs4986172
rs227724
rs1401795
rs2079795
rs2854207
rs2070776
rs4369779
rs12458127
rs9967417
rs11880992
rs2074977
rs4803468
rs17721822
rs1884897
rs1074683
rs143384
Genetic effect size (cm per copy of the risk allele)
Height quantile

Appendix G: Summary statistics for standard linear regression effect size estimates for association with height.

SNP	Gene	CHR	BP	EA	MAF	Beta [95% CI]	P
rs143384	*GDF5*	20	34025756	G	0.40	0.637 [0.571; 0.702]	2.14×10^{-80}
rs724016	*ZBTB38*	3	141105570	G	0.45	0.540 [0.475; 0.605]	1.17×10^{-59}
rs1812175	*HHIP*	4	145574844	A	0.17	-0.562 [-0.649; -0.476]	2.05×10^{-37}
rs3791679	*EFEMP1*	2	56096892	G	0.22	-0.495 [-0.572; -0.417]	6.99×10^{-36}
rs7692995	*LCORL*	4	17936634	C	0.16	-0.556 [-0.645; -0.467]	2.63×10^{-34}
rs6845999	*HHIP*	4	145565826	T	0.43	0.397 [0.331; 0.462]	9.2×10^{-33}
rs798497	*GNA12*	7	2795957	G	0.30	-0.427 [-0.498; -0.357]	3.21×10^{-32}
rs42039	*CDK6*	7	92244422	T	0.25	0.448 [0.373; 0.523]	1.62×10^{-31}
rs8756	*HMGA2*	12	66359752	C	0.48	0.385 [0.32; 0.449]	3.41×10^{-31}
rs4896582	*GPR126*	6	142703877	A	0.30	-0.403 [-0.474; -0.332]	6.46×10^{-29}
rs552707	*JAZF1*	7	28205303	T	0.30	0.401 [0.33; 0.471]	1.65×10^{-28}
rs3825199	*SOCS2*	12	93976954	G	0.22	0.434 [0.356; 0.511]	7.31×10^{-28}
rs3118905	*DLEU7*	13	51105334	A	0.28	-0.402 [-0.474; -0.329]	1.32×10^{-27}
rs7162542	*ADAMTSL3*	15	84514290	C	0.45	-0.358 [-0.423; -0.293]	3.23×10^{-27}
rs4369779	*CABLES1*	18	20735408	T	0.21	-0.431 [-0.51; -0.352]	1.58×10^{-26}
rs16942341	*ACAN*	15	89388905	T	0.03	-1.052 [-1.247; -0.857]	3.52×10^{-26}
rs17556750	*PRKG2*	4	82155568	A	0.29	0.370 [0.298; 0.442]	6.16×10^{-24}
rs4733724	*MLZE*	8	130723728	G	0.20	-0.413 [-0.493; -0.332]	7.92×10^{-24}
rs806794	*HIST1H2BF*	6	26200677	G	0.27	-0.373 [-0.446; -0.300]	1.52×10^{-23}
rs2289195	*DNMT3A*	2	25463483	A	0.41	0.328 [0.262; 0.393]	1.39×10^{-22}
rs2070776	*CD79B*	17	62007498	A	0.35	-0.333 [-0.401; -0.265]	4.54×10^{-22}
rs1884897	*BMP2*	20	6612832	A	0.37	0.330 [0.263; 0.398]	5.77×10^{-22}
rs2871865	*IGF1R*	15	99194896	G	0.12	-0.499 [-0.600; -0.397]	5.84×10^{-22}
rs9428104	*SPAG17*	1	118855587	A	0.25	-0.36 [-0.434; -0.285]	3.54×10^{-21}
rs2425163	*PHF20*	20	34432670	G	0.17	0.412 [0.327; 0.498]	4.27×10^{-21}
rs10958476	*PLAG1*	8	57095808	C	0.21	0.373 [0.294; 0.452]	1.93×10^{-20}
rs12214804	*HMGA1*	6	34188866	C	0.09	0.535 [0.422; 0.649]	2.27×10^{-20}
rs2854207	*CSH2*	17	61947107	G	0.28	0.338 [0.266; 0.411]	3.85×10^{-20}
rs2888877	*CDK6*	7	92228400	T	0.20	0.379 [0.298; 0.460]	5.79×10^{-20}
rs2079795	*C17orf82*	17	59496649	T	0.33	0.313 [0.245; 0.381]	2.34×10^{-19}
rs2284746	*MFAP2*	1	17306675	C	0.48	-0.296 [-0.361; -0.232]	2.64×10^{-19}
rs2257011	*SH3GL3*	15	84266145	T	0.49	0.296 [0.231; 0.361]	4.77×10^{-19}
rs11049611	*CCDC91*	12	28600244	T	0.31	-0.317 [-0.387; -0.247]	8.4×10^{-19}
rs2687950	*KCNRG*	13	50718468	T	0.25	0.334 [0.260; 0.408]	9.79×10^{-19}
rs9392918	*BMP6*	6	7708631	C	0.48	0.290 [0.225; 0.355]	3.05×10^{-18}
rs3760318	*CENTA2*	17	29247715	A	0.38	-0.296 [-0.363; -0.23]	3.12×10^{-18}
rs7740107	*L3MBTL3*	6	130374461	T	0.26	0.325 [0.252; 0.398]	3.12×10^{-18}
rs10859567	*CRADD*	12	94126925	G	0.45	-0.29 [-0.356; -0.225]	3.16×10^{-18}
rs1415701	*L3MBTL3*	6	130345835	A	0.26	-0.33 [-0.405; -0.256]	4.52×10^{-18}
rs4448343	*PTCH1*	9	98266370	G	0.33	0.300 [0.230; 0.369]	2.23×10^{-17}
rs3800461	*C6orf106*	6	34616322	C	0.11	0.431 [0.330; 0.533]	7.64×10^{-17}
rs817300	*PTCH1*	9	98380222	A	0.08	-0.516 [-0.637; -0.394]	1.09×10^{-16}
rs6457374	*HLA-C*	6	31272261	C	0.30	0.298 [0.227; 0.368]	1.1×10^{-16}
rs10748128	*FRS2*	12	69827658	T	0.35	0.286 [0.218; 0.354]	1.32×10^{-16}
rs9835332	*C3orf63*	3	56667682	C	0.46	-0.273 [-0.338; -0.208]	1.53×10^{-16}
rs1155939	*C6orf173*	6	126866133	A	0.49	0.270 [0.205; 0.334]	2.67×10^{-16}
rs11144688	*PCSK5*	9	78542286	A	0.12	-0.412 [-0.512; -0.313]	3.67×10^{-16}
rs9650315	*CHCHD7*	8	57155598	T	0.13	-0.400 [-0.496; -0.304]	3.94×10^{-16}
rs2280470	*ACAN*	15	89395626	A	0.32	0.284 [0.215; 0.353]	7.08×10^{-16}
rs7870753	*HABP4*	9	99201585	G	0.22	0.320 [0.242; 0.398]	8.87×10^{-16}
rs314263	*LIN28B*	6	105392745	C	0.32	0.283 [0.214; 0.353]	1.05×10^{-15}
rs3116168	*DIS3L2*	2	232989831	T	0.27	-0.294 [-0.366; -0.222]	1.68×10^{-15}
rs11880992	*DOT1L*	19	2176403	A	0.41	0.264 [0.198; 0.330]	3.98×10^{-15}
rs422421	*FGFR4*	5	176517326	T	0.22	-0.314 [-0.392; -0.235]	5.53×10^{-15}
rs7652177	*FNDC3B*	3	171969077	C	0.50	-0.252 [-0.317; -0.188]	2.11×10^{-14}
rs4868126	*FBXW11*	5	171283469	T	0.40	-0.270 [-0.34; -0.200]	4.57×10^{-14}
rs606452	*SERPINH1*	11	75276178	A	0.14	0.357 [0.264; 0.450]	5.12×10^{-14}
rs12639764	*TET2*	4	106216205	C	0.39	-0.255 [-0.321; -0.188]	7.76×10^{-14}
rs6919534	*ZNF76*	6	35246903	G	0.14	-0.355 [-0.448; -0.262]	8.08×10^{-14}
rs2278483	*CENPO*	2	25040082	T	0.22	0.291 [0.213; 0.369]	2.45×10^{-13}
rs3814333	*GLT25D2*	1	184007119	T	0.32	0.260 [0.190; 0.330]	3.16×10^{-13}
rs6894139	*MEF2C*	5	88327782	G	0.47	-0.241 [-0.306; -0.176]	3.35×10^{-13}
rs6694089	*DNM3*	1	172083881	A	0.29	0.264 [0.192; 0.335]	4.35×10^{-13}
rs17511102	*CDC42EP3*	2	37960613	T	0.09	0.413 [0.301; 0.526]	5.16×10^{-13}
rs4803468	*BCKDHA*	19	41922352	A	0.39	0.242 [0.176; 0.308]	6.69×10^{-13}
rs3817428	*ACAN*	15	89415247	G	0.27	-0.266 [-0.339; -0.193]	1.11×10^{-12}
rs1074683	*PXMP4*	20	32304653	G	0.26	-0.266 [-0.340; -0.192]	2.07×10^{-12}
rs7033487	*PAPPA*	9	119129257	C	0.20	-0.29 [-0.371; -0.209]	2.35×10^{-12}
rs9993613	*ADAMTS3*	4	73476014	T	0.47	0.231 [0.166; 0.295]	2.93×10^{-12}
rs6696239	*ZNF678*	1	227750068	A	0.19	-0.290 [-0.372; -0.208]	4.16×10^{-12}
rs7154721	*TRIP11*	14	92427348	C	0.42	-0.230 [-0.296; -0.165]	5.04×10^{-12}
rs11677466	*DIS3L2*	2	232982257	T	0.09	0.390 [0.279; 0.501]	6.21×10^{-12}
rs1401795	*C17orf67*	17	54839652	G	0.50	-0.228 [-0.294; -0.163]	6.64×10^{-12}
rs10990303	*PTCH1*	9	98410405	T	0.22	0.279 [0.199; 0.359]	9.21×10^{-12}
rs17450430	*STAU1*	20	47772264	T	0.24	0.261 [0.185; 0.337]	1.66×10^{-11}
rs9292468	*C5orf23*	5	32819073	T	0.40	0.226 [0.160; 0.292]	2.27×10^{-11}
rs1681630	*PTPRJ*	11	47969152	T	0.34	0.232 [0.164; 0.301]	3.22×10^{-11}
rs12538407	*IGF2BP3*	7	23521316	G	0.42	-0.221 [-0.287; -0.155]	5.16×10^{-11}

rs509035	*GHSR*	3	172163449	A	0.32	0.232 [0.163; 0.302]	5.62×10^{-11}
rs6955948	*TMEM176A*	7	150508720	T	0.27	0.242 [0.170; 0.315]	6.66×10^{-11}
rs4548838	*ADAMTS17*	15	100761190	T	0.46	0.217 [0.152; 0.282]	6.76×10^{-11}
rs2857693	*BAT2*	6	31588384	T	0.36	-0.223 [-0.29; -0.156]	7.6×10^{-11}
rs10780910	*SPIN1*	9	90849255	T	0.41	0.220 [0.153; 0.286]	9.12×10^{-11}
rs991967	*TGFB2*	1	218615451	C	0.29	0.235 [0.164; 0.307]	1.13×10^{-10}
rs11633371	*ACAN*	15	89356832	T	0.48	0.211 [0.146; 0.276]	1.69×10^{-10}
rs6439168	*H1FX*	3	129050943	A	0.22	-0.254 [-0.332; -0.176]	1.72×10^{-10}
rs2974438	*SLIT3*	5	168250903	A	0.21	-0.260 [-0.340; -0.180]	1.92×10^{-10}
rs181338	*ZCCHC6*	9	89108161	C	0.50	-0.210 [-0.275; -0.145]	2.2×10^{-10}
rs12125882	*DNM3*	1	172141403	T	0.42	0.212 [0.146; 0.277]	2.31×10^{-10}
rs2856321	*ETV6*	12	11855773	G	0.36	0.216 [0.149; 0.283]	3.12×10^{-10}
rs3020418	*ESR1*	6	152345162	A	0.27	0.232 [0.160; 0.305]	3.5×10^{-10}
rs2581830	*RFT1*	3	53134098	T	0.40	0.211 [0.145; 0.277]	3.71×10^{-10}
rs1368380	*FBXW11*	5	171285632	T	0.43	0.205 [0.140; 0.270]	6.61×10^{-10}
rs1036821	*ZFAT*	8	135650483	A	0.30	-0.223 [-0.294; -0.152]	7.2×10^{-10}
rs4240326	*ANAPC10*	4	145839264	A	0.45	0.203 [0.138; 0.268]	8.64×10^{-10}
rs12209223	*FILIP1*	6	76164589	A	0.10	0.335 [0.227; 0.443]	1.19×10^{-9}
rs17721822	*BMP2*	20	6469596	A	0.36	-0.209 [-0.277; -0.142]	1.22×10^{-9}
rs6902771	*ESR1*	6	152157881	T	0.46	0.200 [0.136; 0.265]	1.3×10^{-9}
rs9217	*ZBTB4*	17	7363088	C	0.36	0.207 [0.139; 0.274]	2×10^{-9}
rs11684404	*EIF2AK3*	2	88924622	C	0.34	0.209 [0.140; 0.277]	2.03×10^{-9}
rs2093210	*C14orf39*	14	60957279	C	0.39	0.201 [0.135; 0.267]	2.35×10^{-9}
rs1923367	*ZCCHC24*	10	81132829	C	0.46	-0.198 [-0.263; -0.133]	2.7×10^{-9}
rs1753637	*DLEU7*	13	51084173	T	0.31	0.212 [0.142; 0.282]	2.94×10^{-9}
rs1809889	*FAM101A*	12	124801226	T	0.27	0.220 [0.148; 0.293]	3.15×10^{-9}
rs7701414	*PDLIM4*	5	131585958	G	0.46	0.195 [0.130; 0.260]	3.91×10^{-9}
rs7971536	*CCDC53*	12	102373788	A	0.50	-0.195 [-0.26; -0.130]	4.17×10^{-9}
rs11175992	*HMGA2*	12	66391396	A	0.24	-0.250 [-0.333; -0.166]	4.71×10^{-9}
rs310421	*FAM46A*	6	81792063	G	0.45	-0.193 [-0.258; -0.128]	6.8×10^{-9}
rs4986172	*ACBD4*	17	43216281	T	0.34	-0.201 [-0.269; -0.132]	8.84×10^{-9}
rs4735677	*PXMP3*	8	78148191	T	0.29	0.208 [0.136; 0.280]	1.25×10^{-8}
rs1950500	*NFATC4*	14	24830850	T	0.29	0.206 [0.135; 0.277]	1.26×10^{-8}
rs4337252	*LOXL1*	15	74226765	C	0.49	0.186 [0.121; 0.250]	1.89×10^{-8}
rs7733195	*FAM44B*	5	172994624	A	0.35	-0.194 [-0.262; -0.126]	2.07×10^{-8}
rs12474201	*SOCS5*	2	46921285	A	0.37	0.187 [0.120; 0.254]	4.23×10^{-8}
rs2305833	*VIL1*	2	219305404	G	0.43	-0.182 [-0.248; -0.117]	5.37×10^{-8}
rs526896	*PITX1*	5	134356705	G	0.28	-0.200 [-0.272; -0.127]	5.96×10^{-8}
rs10770705	*SLCO1C1*	12	20857467	A	0.34	0.189 [0.121; 0.257]	6.03×10^{-8}
rs6600365	*SCMH1*	1	41556253	C	0.44	0.179 [0.114; 0.244]	7.2×10^{-8}
rs7716219	*SLC38A9*	5	54955071	T	0.30	0.193 [0.122; 0.263]	8.85×10^{-8}
rs1265097	*PSORS1C1/PSORS1C2*	6	31106459	A	0.08	-0.326 [-0.445; -0.206]	8.92×10^{-8}
rs7551732	*PKN2*	1	89139041	T	0.42	-0.178 [-0.244; -0.113]	1.06×10^{-7}
rs7466269	*FUBP3*	9	133464084	G	0.36	-0.183 [-0.251; -0.116]	1.07×10^{-7}
rs862034	*LTBP2*	14	74990746	A	0.36	-0.180 [-0.248; -0.113]	1.46×10^{-7}
rs33852	*FBXW11*	5	171189571	G	0.32	0.185 [0.115; 0.254]	1.7×10^{-7}
rs9967417	*DYM*	18	46959500	G	0.44	0.173 [0.108; 0.238]	2.1×10^{-7}
rs720390	*IGF2BP2*	3	185548683	A	0.37	0.175 [0.108; 0.242]	3.05×10^{-7}
rs4973429	*C2orf52*	2	232377818	T	0.32	-0.182 [-0.252; -0.112]	3.08×10^{-7}
rs749052	*NPPC*	2	232796610	C	0.06	-0.350 [-0.486; -0.214]	4.68×10^{-7}
rs12330322	*RYBP*	3	72455355	T	0.22	-0.196 [-0.275; -0.118]	8.69×10^{-7}
rs4141885	*HIST1H1E*	6	26157481	T	0.08	-0.288 [-0.410; -0.166]	3.61×10^{-6}
rs12470505	*CCDC108*	2	219908369	G	0.10	-0.254 [-0.363; -0.145]	5.3×10^{-6}
rs996743	*OTUD4*	4	146128884	G	0.09	-0.262 [-0.376; -0.149]	5.98×10^{-6}
rs648831	*BCKDHB*	6	80956208	C	0.47	-0.151 [-0.216; -0.085]	6.4×10^{-6}
rs5742915	*PML*	15	74336633	C	0.46	0.149 [0.084; 0.214]	6.5×10^{-6}
rs7849585	*QSOX2*	9	139111870	T	0.34	0.157 [0.088; 0.225]	7.28×10^{-6}
rs7027110	*ZNF462*	9	109599046	A	0.23	0.174 [0.097; 0.251]	1.04×10^{-5}
rs12458127	*DYM*	18	46657358	T	0.05	-0.329 [-0.477; -0.181]	1.33×10^{-5}
rs12153391	*FBXW11*	5	171203438	A	0.25	-0.150 [-0.225; -0.075]	8.29×10^{-5}
rs2074977	*NFIC*	19	3434028	C	0.36	0.135 [0.068; 0.203]	8.83×10^{-5}
rs3807931	*ITGB8*	7	20381674	A	0.46	0.125 [0.059; 0.190]	0.000176
rs7534365	*SV2A*	1	149876124	C	0.13	0.194 [0.09; 0.299]	0.000269
rs2573625	*ADAMTS17*	15	100513158	C	0.34	-0.126 [-0.195; -0.057]	0.000363
rs227724	*C17orf67*	17	54778817	T	0.34	0.119 [0.051; 0.187]	0.000636
rs1582931	*CCDC100*	5	122657199	A	0.47	-0.112 [-0.178; -0.046]	0.000868
rs2806561	*LUZP1*	1	23504795	G	0.44	-0.101 [-0.166; -0.036]	0.00226
rs1659127	*MKL2*	16	14388305	A	0.33	0.106 [0.037; 0.175]	0.00245
rs10948222	*SUPT3H*	6	45244415	C	0.40	0.099 [0.033; 0.165]	0.00317
rs7043114	*IPPK*	9	95387983	C	0.44	0.083 [0.018; 0.148]	0.0128

Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *MAF* - minor allele frequency, *Beta* - the effect size in dioptres per copy of the risk allele, *CI* - confidence interval.

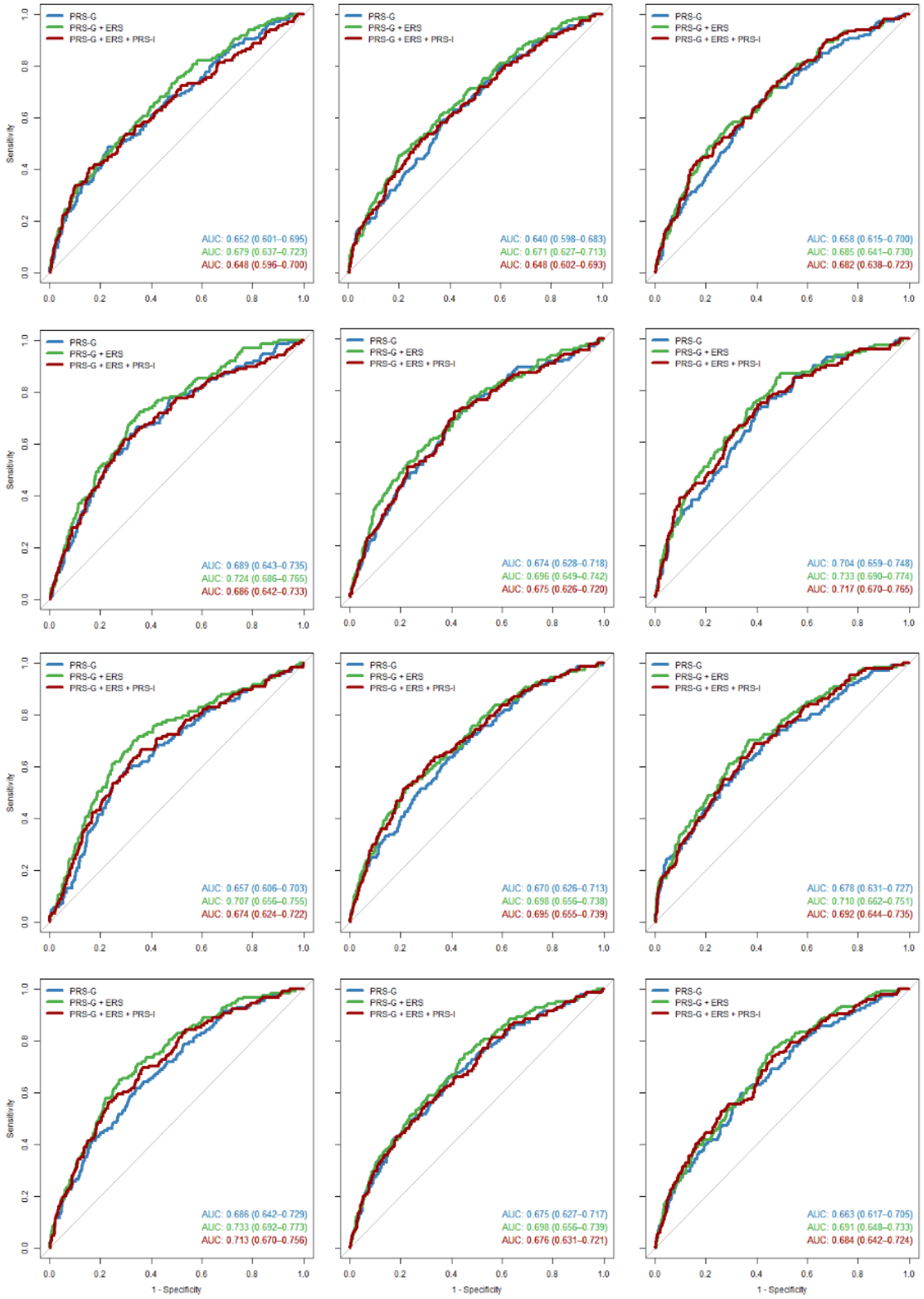
Appendix H: Summary statistics for standard linear regression effect size estimates for association with height.

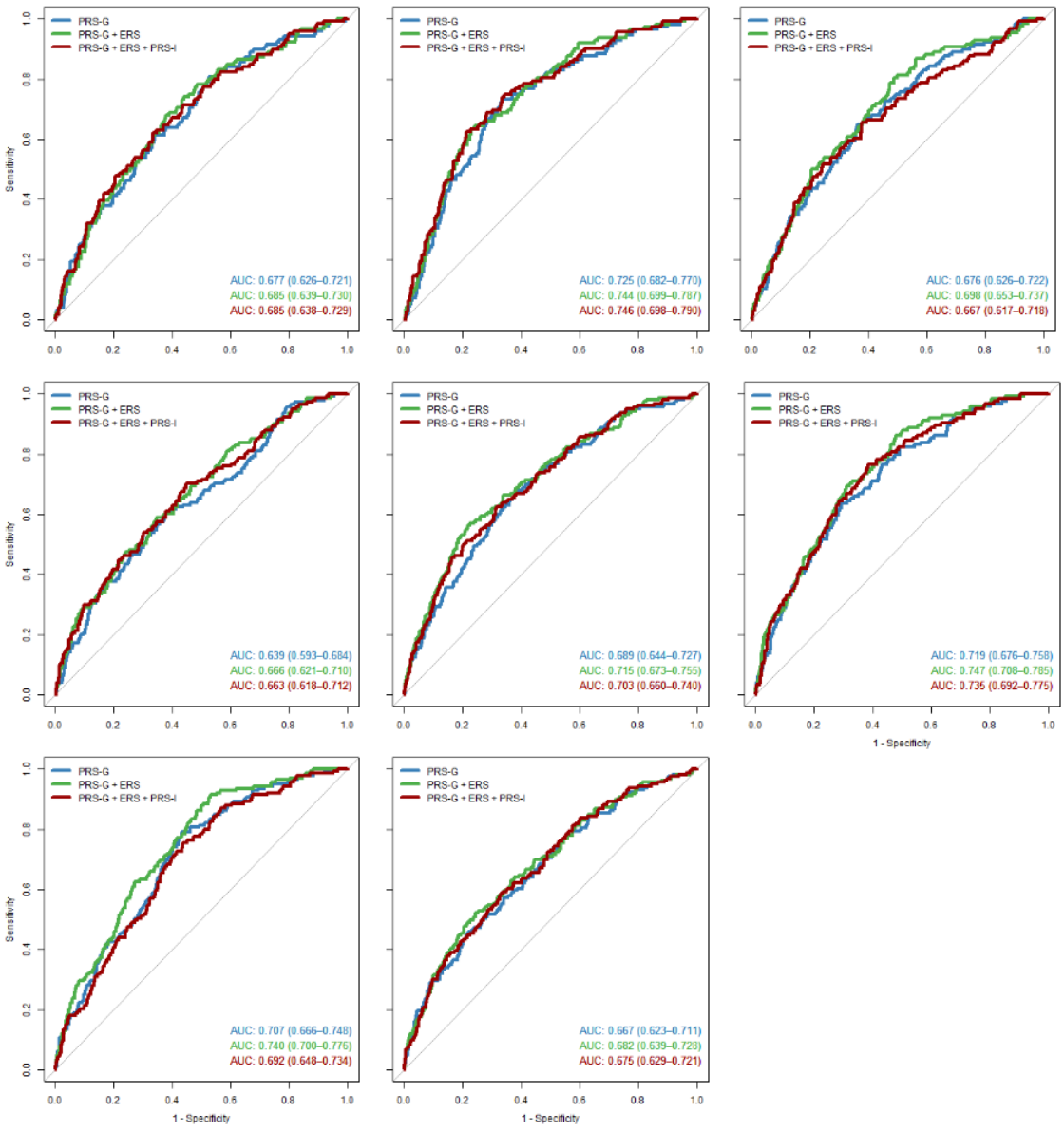
SNP	GENE	β_0		β_1		β_2	
		Beta [95% CI]	P	Beta [95% CI]	P	Beta [95% CI]	P
rs143384	*GDF5*	0.567 [0.441; 0.693]	7.21 x 10-18	0.069 [-0.497; 0.634]	0.812	0.150 [-0.410; 0.710]	0.599
rs724016	*ZBTB38*	0.481 [0.351; 0.611]	1.42 x 10-12	0.291 [-0.284; 0.866]	0.322	-0.274 [-0.839; 0.292]	0.343
rs3825199	*SOCS2*	0.551 [0.401; 0.701]	1.87 x 10-12	-0.682 [-1.351; -0.013]	0.0456	0.715 [0.056; 1.373]	0.0333
rs3791679	*EFEMP1*	-0.536 [-0.685; -0.387]	5.31 x 10-12	0.260 [-0.408; 0.929]	0.445	-0.239 [-0.897; 0.419]	0.476
rs9428104	*SPAG17*	-0.449 [-0.592; -0.306]	1.70 x 10-09	0.345 [-0.291; 0.982]	0.288	-0.272 [-0.898; 0.354]	0.395
rs17556750	*PRKG2*	0.428 [0.289; 0.566]	3.68 x 10-09	-0.062 [-0.683; 0.558]	0.845	-0.106 [-0.721; 0.509]	0.735
rs11144688	*PCSK5*	-0.578 [-0.769; -0.387]	6.65 x 10-09	0.814 [-0.045; 1.673]	0.0633	-0.775 [-1.629; 0.079]	0.0751
rs4896582	*GPR126*	-0.402 [-0.539; -0.266]	1.67 x 10-08	-0.032 [-0.64; 0.576]	0.918	0.088 [-0.511; 0.686]	0.774
rs2871865	*IGF1R*	-0.571 [-0.765; -0.376]	2.00 x 10-08	0.180 [-0.695; 1.056]	0.687	-0.029 [-0.892; 0.835]	0.948
rs7692995	*LCORL*	-0.523 [-0.706; -0.34]	4.00 x 10-08	-0.290 [-1.104; 0.524]	0.486	0.367 [-0.439; 1.173]	0.372
rs806794	*HIST1H2BF*	-0.401 [-0.541; -0.261]	4.00 x 10-08	0.163 [-0.463; 0.788]	0.61	-0.170 [-0.786; 0.446]	0.589
rs42039	*CDK6*	0.402 [0.261; 0.544]	5.23 x 10-08	0.208 [-0.431; 0.848]	0.523	-0.183 [-0.82; 0.454]	0.573
rs3760318	*CENTA2*	-0.363 [-0.493; -0.232]	9.87 x 10-08	0.132 [-0.448; 0.712]	0.655	0.001 [-0.569; 0.571]	0.998
rs2289195	*DNMT3A*	0.348 [0.221; 0.475]	1.60 x 10-07	-0.070 [-0.638; 0.497]	0.809	0.064 [-0.495; 0.622]	0.823
rs4369779	*CABLES1*	-0.420 [-0.574; -0.265]	1.86 x 10-07	-0.120 [-0.809; 0.568]	0.732	0.134 [-0.546; 0.814]	0.699
rs2857693	*BAT2*	-0.34 [-0.465; -0.215]	1.90 x 10-07	0.482 [-0.082; 1.047]	0.0941	-0.395 [-0.956; 0.167]	0.168
rs11049611	*CCDC91*	-0.369 [-0.505; -0.233]	2.13 x 10-07	0.377 [-0.227; 0.981]	0.221	-0.397 [-0.992; 0.197]	0.19
rs798497	*GNA12*	-0.366 [-0.502; -0.23]	2.62 x 10-07	-0.086 [-0.691; 0.520]	0.782	-0.051 [-0.645; 0.544]	0.867
rs7162542	*ADAMTSL3*	-0.327 [-0.449; -0.204]	3.36 x 10-07	-0.151 [-0.700; 0.398]	0.589	0.167 [-0.374; 0.709]	0.545
rs2888877	*CDK6*	0.396 [0.245; 0.547]	5.04 x 10-07	0.070 [-0.613; 0.754]	0.84	-0.191 [-0.874; 0.492]	0.584
rs552707	*JAZF1*	0.359 [0.219; 0.500]	9.69 x 10-07	0.214 [-0.410; 0.839]	0.501	-0.256 [-0.872; 0.361]	0.417
rs11880992	*DOT1L*	0.325 [0.198; 0.453]	9.71 x 10-07	-0.189 [-0.758; 0.380]	0.515	0.071 [-0.490; 0.631]	0.805
rs1812175	*HHIP*	-0.424 [-0.589; -0.258]	9.86 x 10-07	-0.234 [-0.977; 0.509]	0.537	-0.074 [-0.806; 0.657]	0.842
rs16942341	*ACAN*	-0.935 [-1.312; -0.558]	2.03 x 10-06	-0.253 [-1.931; 1.424]	0.767	0.075 [-1.579; 1.729]	0.929
rs12209223	*FILIP1*	0.501 [0.298; 0.704]	2.18 x 10-06	-0.645 [-1.572; 0.281]	0.172	0.475 [-0.454; 1.404]	0.316
rs10780910	*SPIN1*	0.314 [0.187; 0.441]	2.19 x 10-06	-0.520 [-1.087; 0.047]	0.0723	0.531 [-0.027; 1.090]	0.0622
rs9835332	*C3orf63*	-0.305 [-0.429; -0.181]	2.59 x 10-06	0.380 [-0.176; 0.935]	0.181	-0.458 [-1.007; 0.090]	0.102
rs8756	*HMGA2*	0.306 [0.181; 0.431]	2.86 x 10-06	0.169 [-0.389; 0.727]	0.553	-0.052 [-0.601; 0.496]	0.852
rs606452	*SERPINH1*	0.448 [0.264; 0.632]	3.17 x 10-06	-0.282 [-1.105; 0.540]	0.501	0.174 [-0.637; 0.985]	0.675
rs3118905	*DLEU7*	-0.341 [-0.483; -0.2]	3.73 x 10-06	-0.499 [-1.127; 0.128]	0.119	0.594 [-0.021; 1.208]	0.0582
rs3800461	*C6orf106*	0.449 [0.263; 0.635]	3.85 x 10-06	-0.094 [-0.931; 0.743]	0.826	0.034 [-0.792; 0.859]	0.936
rs2425163	*PHF20*	0.402 [0.235; 0.569]	4.04 x 10-06	0.072 [-0.668; 0.812]	0.849	-0.071 [-0.801; 0.659]	0.849
rs12214804	*HMGA1*	0.557 [0.324; 0.79]	4.58 x 10-06	-0.286 [-1.309; 0.737]	0.583	0.294 [-0.709; 1.297]	0.565
rs9292468	*C5orf23*	0.299 [0.173; 0.426]	5.73 x 10-06	-0.464 [-1.028; 0.100]	0.107	0.472 [-0.085; 1.028]	0.0966
rs6845999	*HHIP*	0.298 [0.172; 0.424]	6.13 x 10-06	0.151 [-0.411; 0.713]	0.598	0.092 [-0.460; 0.644]	0.744
rs10748128	*FRS2*	0.302 [0.172; 0.433]	9.08 x 10-06	-0.191 [-0.773; 0.391]	0.521	0.251 [-0.325; 0.826]	0.393
rs2278483	*CENPO*	0.353 [0.199; 0.507]	1.08 x 10-05	0.093 [-0.586; 0.772]	0.789	-0.349 [-1.014; 0.316]	0.303
rs10958476	*PLAG1*	0.364 [0.205; 0.522]	1.09 x 10-05	0.187 [-0.511; 0.884]	0.6	-0.245 [-0.928; 0.439]	0.483
rs12639764	*TET2*	-0.284 [-0.41; -0.158]	1.50 x 10-05	0.083 [-0.481; 0.647]	0.773	-0.048 [-0.605; 0.509]	0.866
rs422421	*FGFR4*	-0.338 [-0.491; -0.186]	2.14 x 10-05	0.023 [-0.658; 0.703]	0.948	0.060 [-0.613; 0.732]	0.862
rs2079795	*C17orf82*	0.283 [0.155; 0.412]	2.27 x 10-05	-0.056 [-0.633; 0.520]	0.848	0.179 [-0.395; 0.752]	0.542
rs7971536	*CCDC53*	-0.270 [-0.394; -0.146]	3.11 x 10-05	0.330 [-0.225; 0.885]	0.244	-0.283 [-0.832; 0.267]	0.313
rs17511102	*CDC42EP3*	0.470 [0.253; 0.687]	3.28 x 10-05	-0.545 [-1.517; 0.426]	0.271	0.687 [-0.268; 1.642]	0.159
rs1036821	*ZFAT*	-0.297 [-0.435; -0.159]	3.75 x 10-05	0.321 [-0.292; 0.933]	0.305	-0.260 [-0.861; 0.342]	0.397
rs1582931	*CCDC100*	-0.269 [-0.395; -0.143]	4.23 x 10-05	0.538 [-0.023; 1.100]	0.0602	-0.336 [-0.891; 0.218]	0.234
rs2687950	*KCNRG*	0.309 [0.164; 0.454]	4.46 x 10-05	0.142 [-0.504; 0.789]	0.666	-0.140 [-0.776; 0.497]	0.667
rs1155939	*C6orf173*	0.262 [0.138; 0.386]	5.19 x 10-05	-0.150 [-0.704; 0.405]	0.597	0.223 [-0.326; 0.772]	0.427
rs991967	*TGFB2*	0.291 [0.153; 0.429]	5.46 x 10-05	-0.056 [-0.67; 0.558]	0.858	-0.082 [-0.687; 0.523]	0.791
rs2856321	*ETV6*	0.269 [0.141; 0.398]	5.76 x 10-05	-0.210 [-0.783; 0.363]	0.473	0.178 [-0.386; 0.743]	0.536
rs2257011	*SH3GL3*	0.261 [0.136; 0.386]	6.25 x 10-05	0.182 [-0.377; 0.740]	0.524	-0.205 [-0.757; 0.346]	0.465
rs1884897	*BMP2*	0.268 [0.139; 0.397]	6.75 x 10-05	0.024 [-0.551; 0.599]	0.935	0.142 [-0.426; 0.710]	0.625
rs17450430	*STAU1*	0.301 [0.156; 0.447]	6.78 x 10-05	-0.048 [-0.694; 0.598]	0.884	-0.061 [-0.699; 0.577]	0.852
rs2284746	*MFAP2*	-0.256 [-0.381; -0.131]	8.34 x 10-05	-0.272 [-0.829; 0.285]	0.339	0.336 [-0.213; 0.886]	0.231
rs6894139	*MEF2C*	-0.257 [-0.382; -0.132]	8.38 x 10-05	-0.066 [-0.621; 0.489]	0.817	0.125 [-0.421; 0.670]	0.654
rs2070776	*CD79B*	-0.305 [-0.454; -0.156]	8.85 x 10-05	-0.027 [-0.698; 0.643]	0.936	-0.035 [-0.698; 0.627]	0.916
rs3116168	*DIS3L2*	-0.289 [-0.431; -0.148]	9.02 x 10-05	-0.005 [-0.637; 0.627]	0.988	0.005 [-0.618; 0.628]	0.988
rs4448343	*PTCH1*	0.272 [0.139; 0.406]	9.13 x 10-05	0.1440 [-0.45; 0.738]	0.635	-0.123 [-0.709; 0.463]	0.681
rs509035	*GHSR*	0.274 [0.140; 0.408]	9.24 x 10-05	-0.371 [-0.971; 0.228]	0.225	0.448 [-0.144; 1.039]	0.138
rs6919534	*ZNF76*	-0.364 [-0.544; -0.185]	9.91 x 10-05	0.260 [-0.539; 1.060]	0.523	-0.320 [-1.11; 0.471]	0.428
rs6457374	*HLA-C*	0.274 [0.139; 0.409]	0.000101	0.303 [-0.298; 0.905]	0.323	-0.364 [-0.958; 0.231]	0.231
rs1074683	*PXMP4*	-0.285 [-0.428; -0.141]	0.000145	0.085 [-0.556; 0.725]	0.795	-0.081 [-0.714; 0.552]	0.802
rs4733724	*MLZE*	-0.297 [-0.449; -0.145]	0.000186	-0.249 [-0.938; 0.440]	0.479	0.006 [-0.679; 0.691]	0.986
rs10859567	*CRADD*	-0.243 [-0.369; -0.118]	0.000208	-0.321 [-0.883; 0.241]	0.263	0.348 [-0.206; 0.902]	0.218
rs9392918	*BMP6*	0.239 [0.115; 0.363]	0.000228	-0.023 [-0.578; 0.532]	0.935	0.190 [-0.359; 0.738]	0.497
rs1415701	*L3MBTL3*	-0.278 [-0.422; -0.133]	0.000242	0.010 [-0.636; 0.655]	0.976	-0.128 [-0.764; 0.508]	0.693
rs12474201	*SOCS5*	0.243 [0.115; 0.372]	0.000285	-0.295 [-0.866; 0.277]	0.312	0.298 [-0.265; 0.862]	0.299
rs7033487	*PAPPA*	-0.296 [-0.453; -0.139]	0.000296	0.233 [-0.469; 0.936]	0.516	-0.354 [-1.051; 0.344]	0.32
rs7849585	*QSOX2*	0.251 [0.117; 0.384]	0.000318	-0.456 [-1.051; 0.139]	0.133	0.405 [-0.181; 0.991]	0.176
rs2280470	*ACAN*	0.246 [0.114; 0.378]	0.000346	0.099 [-0.492; 0.690]	0.743	-0.031 [-0.616; 0.554]	0.917
rs7740107	*L3MBTL3*	0.266 [0.123; 0.408]	0.000371	0.145 [-0.489; 0.778]	0.654	-0.071 [-0.695; 0.552]	0.822
rs3817428	*ACAN*	-0.257 [-0.397; -0.117]	0.000441	0.097 [-0.524; 0.718]	0.76	-0.199 [-0.811; 0.412]	0.523
rs9217	*ZBTB4*	0.238 [0.108; 0.367]	0.00045	-0.078 [-0.657; 0.502]	0.793	0.013 [-0.560; 0.585]	0.965
rs817300	*PTCH1*	-0.398 [-0.616; -0.18]	0.000467	-0.011 [-1.005; 0.982]	0.982	-0.330 [-1.326; 0.666]	0.516
rs7652177	*FNDC3B*	-0.23 [-0.356; -0.104]	0.000476	0.079 [-0.481; 0.638]	0.783	-0.181 [-0.732; 0.370]	0.519
rs2305833	*VIL1*	-0.234 [-0.363; -0.106]	0.000489	0.321 [-0.249; 0.892]	0.27	-0.314 [-0.874; 0.246]	0.272
rs10770705	*SLCO1C1*	0.238 [0.107; 0.369]	0.000515	-0.018 [-0.603; 0.568]	0.953	-0.162 [-0.739; 0.415]	0.582
rs7980687	*SBNO1*	0.284 [0.127; 0.442]	0.000532	-0.164 [-0.860; 0.532]	0.645	0.290 [-0.393; 0.973]	0.406
rs1401795	*C17orf67*	-0.229 [-0.356; -0.102]	0.000536	-0.031 [-0.595; 0.532]	0.913	0.097 [-0.458; 0.652]	0.732
rs3814333	*GLT25D2*	0.247 [0.110; 0.384]	0.000538	0.068 [-0.536; 0.671]	0.826	-0.089 [-0.680; 0.501]	0.767
rs7027110	*ZNF462*	0.279 [0.124; 0.435]	0.000591	-0.236 [-0.921; 0.448]	0.499	0.0450 [-0.626; 0.716]	0.896
rs11175992	*HMGA2*	-0.274 [-0.43; -0.118]	0.000776	0.119 [-0.582; 0.820]	0.739	-0.084 [-0.780; 0.611]	0.812
rs314263	*LIN28B*	0.230 [0.098; 0.361]	0.000828	0.290 [-0.298; 0.877]	0.334	-0.288 [-0.868; 0.292]	0.331

rs4141885	*HIST1H1E*	-0.41 [-0.646; -0.175]	0.000832	0.793 [-0.253; 1.84]	0.137	-0.888 [-1.925; 0.148]	0.093
rs2573625	*ADAMTS17*	-0.228 [-0.362; -0.094]	0.00114	0.428 [-0.170; 1.027]	0.161	-0.366 [-0.956; 0.224]	0.224
rs2074977	*NFIC*	0.211 [0.083; 0.339]	0.00158	-0.307 [-0.878; 0.264]	0.292	0.268 [-0.296; 0.832]	0.351
rs6696239	*ZNF678*	-0.258 [-0.416; -0.100]	0.00174	-0.111 [-0.811; 0.589]	0.757	0.072 [-0.613; 0.756]	0.838
rs6955948	*TMEM176A*	0.227 [0.087; 0.366]	0.00187	-0.193 [-0.815; 0.430]	0.544	0.327 [-0.288; 0.942]	0.297
rs2974438	*SLIT3*	-0.247 [-0.401; -0.093]	0.00207	-0.480 [-1.171; 0.212]	0.174	0.700 [0.016; 1.385]	0.0449
rs2854207	*CSH2*	0.225 [0.085; 0.366]	0.00215	0.400 [-0.225; 1.026]	0.21	-0.291 [-0.908; 0.325]	0.354
rs181338	*ZCCHC6*	-0.197 [-0.320; -0.073]	0.00228	0.011 [-0.537; 0.558]	0.969	-0.034 [-0.570; 0.503]	0.902
rs9993613	*ADAMTS3*	0.199 [0.073; 0.324]	0.00237	-0.049 [-0.605; 0.506]	0.862	0.165 [-0.379; 0.710]	0.552
rs9967417	*DYM*	0.197 [0.070; 0.324]	0.00288	-0.137 [-0.700; 0.425]	0.633	0.155 [-0.397; 0.707]	0.583
rs5742915	*PML*	0.192 [0.068; 0.316]	0.00301	-0.172 [-0.726; 0.382]	0.543	0.120 [-0.428; 0.669]	0.667
rs1265097	*PSORS1C1/PSORS1C2*	-0.352 [-0.581; -0.123]	0.00319	0.466 [-0.554; 1.487]	0.37	-0.628 [-1.63; 0.374]	0.219
rs4803468	*BCKDHA*	0.192 [0.066; 0.318]	0.00352	0.009 [-0.55; 0.569]	0.975	0.100 [-0.449; 0.650]	0.72
rs9650315	*CHCHD7*	-0.267 [-0.444; -0.091]	0.00373	-0.770 [-1.573; 0.032]	0.0598	0.786 [-0.013; 1.586]	0.054
rs10990303	*PTCH1*	0.234 [0.079; 0.390]	0.00395	0.057 [-0.637; 0.751]	0.872	0.078 [-0.606; 0.761]	0.824
rs17721822	*BMP2*	-0.195 [-0.326; -0.065]	0.00405	-0.008 [-0.587; 0.570]	0.978	0.008 [-0.562; 0.577]	0.979
rs1753637	*DLEU7*	0.198 [0.066; 0.330]	0.00417	0.081 [-0.509; 0.671]	0.787	-0.054 [-0.634; 0.525]	0.854
rs310421	*FAM46A*	-0.185 [-0.309; -0.061]	0.00417	0.074 [-0.478; 0.627]	0.792	-0.128 [-0.672; 0.417]	0.646
rs12470505	*CCDC108*	-0.312 [-0.521; -0.102]	0.00431	0.068 [-0.859; 0.995]	0.886	0.121 [-0.789; 1.031]	0.794
rs12538407	*IGF2BP3*	-0.187 [-0.314; -0.059]	0.00515	-0.180 [-0.749; 0.390]	0.536	0.192 [-0.37; 0.754]	0.503
rs4986172	*ACBD4*	-0.192 [-0.325; -0.058]	0.00602	0.338 [-0.253; 0.930]	0.262	-0.537 [-1.120; 0.046]	0.0711
rs4337252	*LOXL1*	0.179 [0.054; 0.304]	0.00615	-0.037 [-0.590; 0.517]	0.896	0.080 [-0.464; 0.624]	0.773
rs862034	*LTBP2*	-0.182 [-0.311; -0.053]	0.00674	0.206 [-0.368; 0.780]	0.482	-0.284 [-0.849; 0.281]	0.325
rs7870753	*HABP4*	0.211 [0.061; 0.361]	0.00719	0.080 [-0.591; 0.752]	0.815	0.205 [-0.456; 0.866]	0.543
rs11677466	*DIS3L2*	0.313 [0.089; 0.537]	0.00741	0.176 [-0.815; 1.166]	0.728	-0.021 [-0.991; 0.948]	0.966
rs7551732	*PKN2*	-0.176 [-0.303; -0.05]	0.00763	0.123 [-0.441; 0.687]	0.669	-0.241 [-0.796; 0.314]	0.395
rs4973429	*C2orf52*	-0.185 [-0.318; -0.051]	0.00816	-0.107 [-0.698; 0.483]	0.722	0.159 [-0.416; 0.734]	0.587
rs4868126	*FBXW11*	-0.189 [-0.326; -0.051]	0.00844	-0.435 [-1.046; 0.176]	0.163	0.433 [-0.169; 1.034]	0.159
rs1809889	*FAM101A*	0.197 [0.053; 0.340]	0.00874	0.019 [-0.617; 0.655]	0.954	0.053 [-0.570; 0.675]	0.868
rs33852	*FBXW11*	0.181 [0.049; 0.314]	0.00884	0.024 [-0.571; 0.619]	0.936	-0.061 [-0.650; 0.528]	0.84
rs3020418	*ESR1*	0.191 [0.050; 0.332]	0.00923	0.259 [-0.363; 0.881]	0.414	-0.293 [-0.903; 0.316]	0.345
rs11684404	*EIF2AK3*	0.174 [0.041; 0.307]	0.012	0.021 [-0.571; 0.613]	0.946	0.051 [-0.534; 0.636]	0.864
rs4548838	*ADAMTS17*	0.167 [0.039; 0.294]	0.0121	0.088 [-0.477; 0.653]	0.759	-0.013 [-0.570; 0.544]	0.963
rs4735677	*PXMP3*	0.181 [0.039; 0.322]	0.0147	0.155 [-0.469; 0.779]	0.627	-0.131 [-0.738; 0.477]	0.674
rs1950500	*NFATC4*	0.169 [0.034; 0.304]	0.0162	-0.004 [-0.607; 0.600]	0.991	0.083 [-0.514; 0.681]	0.784
rs720390	*IGF2BP2*	0.163 [0.033; 0.294]	0.0163	-0.062 [-0.639; 0.516]	0.835	0.115 [-0.453; 0.683]	0.691
rs7716219	*SLC38A9*	0.163 [0.030; 0.296]	0.019	0.300 [-0.294; 0.894]	0.322	-0.373 [-0.957; 0.212]	0.212
rs11633371	*ACAN*	0.151 [0.027; 0.274]	0.0195	0.268 [-0.282; 0.818]	0.339	-0.253 [-0.795; 0.288]	0.359
rs4240326	*ANAPC10*	0.150 [0.024; 0.275]	0.0227	0.383 [-0.175; 0.941]	0.179	-0.409 [-0.957; 0.139]	0.143
rs648831	*BCKDHB*	-0.143 [-0.267; -0.019]	0.0277	-0.100 [-0.658; 0.458]	0.725	0.143 [-0.408; 0.694]	0.611
rs6600365	*SCMH1*	0.143 [0.017; 0.268]	0.029	0.144 [-0.413; 0.700]	0.613	-0.122 [-0.670; 0.426]	0.662
rs996743	*OTUD4*	-0.236 [-0.446; -0.027]	0.0305	0.072 [-0.882; 1.027]	0.882	-0.232 [-1.183; 0.719]	0.633
rs12458127	*DYM*	-0.330 [-0.628; -0.032]	0.0338	0.203 [-1.114; 1.521]	0.762	-0.279 [-1.566; 1.008]	0.671
rs6439168	*H1FX*	-0.164 [-0.314; -0.015]	0.035	-0.344 [-1.008; 0.320]	0.31	0.273 [-0.379; 0.926]	0.411
rs227724	*C17orf67*	0.139 [0.009; 0.270]	0.0401	-0.097 [-0.675; 0.481]	0.743	0.091 [-0.477; 0.658]	0.754
rs12125882	*DNM3*	0.131 [0.006; 0.255]	0.045	0.367 [-0.189; 0.923]	0.195	-0.340 [-0.888; 0.208]	0.224
rs1923367	*ZCCHC24*	-0.127 [-0.252; -0.002]	0.0518	0.071 [-0.483; 0.626]	0.801	-0.313 [-0.857; 0.230]	0.258
rs10948222	*SUPT3H*	0.126 [-0.001; 0.254]	0.0573	-0.072 [-0.637; 0.493]	0.803	0.011 [-0.544; 0.567]	0.968
rs2093210	*C14orf39*	0.122 [-0.005; 0.249]	0.0656	0.213 [-0.356; 0.783]	0.463	-0.075 [-0.639; 0.489]	0.794
rs12330322	*RYBP*	-0.138 [-0.282; 0.006]	0.0662	0.020 [-0.634; 0.673]	0.953	-0.251 [-0.903; 0.401]	0.45
rs7701414	*PDLIM4*	0.111 [-0.015; 0.237]	0.0918	0.339 [-0.224; 0.902]	0.238	-0.300 [-0.856; 0.256]	0.29
rs2806561	*LUZP1*	-0.107 [-0.234; 0.019]	0.103	-0.115 [-0.672; 0.442]	0.686	0.198 [-0.35; 0.746]	0.478
rs7466269	*FUBP3*	-0.110 [-0.242; 0.022]	0.111	-0.371 [-0.954; 0.213]	0.213	0.343 [-0.229; 0.915]	0.24
rs1681630	*PTPRJ*	0.109 [-0.025; 0.243]	0.119	0.594 [0.000; 1.188]	0.05	-0.549 [-1.133; 0.036]	0.0657
rs7154721	*TRIP11*	-0.102 [-0.229; 0.024]	0.121	-0.417 [-0.980; 0.146]	0.147	0.290 [-0.264; 0.845]	0.305
rs6694089	*DNM3*	0.109 [-0.029; 0.246]	0.13	0.677 [0.069; 1.284]	0.029	-0.577 [-1.170; 0.017]	0.0568
rs1368380	*FBXW11*	0.099 [-0.028; 0.226]	0.135	0.349 [-0.214; 0.913]	0.224	-0.213 [-0.767; 0.341]	0.452
rs1659127	*MKL2*	0.098 [-0.032; 0.229]	0.147	-0.272 [-0.854; 0.310]	0.36	0.402 [-0.172; 0.977]	0.17
rs526896	*PITX1*	-0.098 [-0.234; 0.038]	0.168	-0.362 [-0.970; 0.245]	0.243	0.240 [-0.357; 0.837]	0.431
rs6902771	*ESR1*	0.087 [-0.037; 0.212]	0.179	0.370 [-0.186; 0.926]	0.193	-0.227 [-0.777; 0.323]	0.418
rs3807931	*ITGB8*	0.083 [-0.041; 0.207]	0.201	0.298 [-0.254; 0.851]	0.29	-0.332 [-0.875; 0.211]	0.23
rs7043114	*IPPK*	0.073 [-0.052; 0.198]	0.265	0.000 [-0.555; 0.555]	1	0.015 [-0.532; 0.561]	0.958
rs7733195	*FAM44B*	-0.067 [-0.196; 0.062]	0.317	-0.403 [-0.980; 0.175]	0.172	0.265 [-0.305; 0.835]	0.362
rs7534365	*SV2A*	-0.043 [-0.242; 0.157]	0.684	0.922 [0.023; 1.821]	0.0443	-0.748 [-1.645; 0.150]	0.103
rs12153391	*FBXW11*	-0.020 [-0.163; 0.122]	0.786	-0.713 [-1.35; -0.076]	0.0283	0.729 [0.097; 1.360]	0.0237
rs749052	*NPPC*	0.028 [-0.237; 0.292]	0.842	-1.306 [-2.476; -0.136]	0.0287	0.874 [-0.279; 2.027]	0.138

Abbreviations: *SNP* - single nucleotide polymorphism, *CHR* - chromosome, *BP* - base pair, *EA* - effect allele, *CI* - confidence interval, β_0 - meta-regression intercept effect size in dioptres per copy of the risk allele, β_1 - meta-regression coefficients for the linear term and β_2 - meta-regression coefficients for the quadratic term.

Appendix I: A comparison of AUC curves for different classification models across 20-fold cross validations




PRS-G
PRS-G + ERS
PRS-G + ERS + PRS-I
AUC: 0.677 (0.626–0.721)
AUC: 0.685 (0.639–0.730)
AUC: 0.685 (0.638–0.729)
AUC: 0.725 (0.682–0.770)
AUC: 0.744 (0.699–0.787)
AUC: 0.746 (0.698–0.790)
AUC: 0.676 (0.626–0.722)
AUC: 0.698 (0.653–0.737)
AUC: 0.667 (0.617–0.718)
AUC: 0.639 (0.593–0.684)
AUC: 0.666 (0.621–0.710)
AUC: 0.663 (0.618–0.712)
AUC: 0.689 (0.644–0.727)
AUC: 0.715 (0.673–0.755)
AUC: 0.703 (0.660–0.740)
AUC: 0.719 (0.676–0.758)
AUC: 0.747 (0.708–0.785)
AUC: 0.735 (0.692–0.775)
AUC: 0.707 (0.666–0.748)
AUC: 0.740 (0.700–0.776)
AUC: 0.692 (0.648–0.734)
AUC: 0.667 (0.623–0.711)
AUC: 0.682 (0.639–0.728)
AUC: 0.675 (0.629–0.721)
Sensitivity
1 - Specificity

References

Abadi A, Alyass A, Robiou du Pont S, Bolker B, Singh P, Mohan V, Diaz R, Engert JC, Yusuf S, Gerstein HC, Anand SS, Meyre D (2017a). Penetrance of polygenic obesity susceptibility loci across the Body Mass Index Distribution. *The American Journal of Human Genetics* **101**(6):925-938.

Acosta-Pech R, Crossa J, de los Campos G, Teyssèdre S, Claustres B, Pérez-Elizalde S, Pérez-Rodríguez P (2017). Genomic models with genotype × environment interaction for predicting hybrid performance: an application in maize hybrids. *Theoretical and Applied Genetics* **130**(7):1431-1440.

Al Kawam A, Alshawaqfeh M, Cai JJ, Serpedin E, Datta A (2018). Simulating variance heterogeneity in quantitative genome wide association studies. *BMC Bioinformatics* **19**(3):72.

Albert PS, Ratnasinghe D, Tangrea J, Wacholder S (2001). Limitations of the case-only design for identifying gene-environment interactions. *American Journal of Epidemiology* **154**(8):687-693.

Aldahmesh MA, Khan AO, Alkuraya H, Adly N, Anazi S, Al-Saleh AA, Mohamed JY, Hijazi H, Prabakaran S, Tacke M, Al-Khrashi A, Hashem M, Reinheckel T, Assiri A, Alkuraya FS (2013). Mutations in *LRPAP1* are associated with severe myopia in humans. *Am J Hum Genet* **93**(2):313-320.

Arnau-Soler A, Macdonald-Dunlop E, Adams MJ, Clarke T-K, MacIntyre DJ, Milburn K, Navrady L, Hayward C, McIntosh AM, Thomson PA, Generation S, Major Depressive Disorder Working Group of the Psychiatric Genomics C (2019). Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland. *Translational Psychiatry* **9**(1):14.

Aschard H (2016). A perspective on interaction effects in genetic association studies. *Genetic Epidemiology* **40**(8):678-688.

Aschard H, Chen J, Cornelis MC, Chibnik LB, Karlson EW, Kraft P (2012a). Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *American journal of human genetics* **90**(6):962-972.

Aschard H, Hancock DB, London SJ, Kraft P (2010). Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Human heredity* **70**(4):292-300.

Aschard H, Lutz S, Maus B, Duell EJ, Fingerlin TE, Chatterjee N, Kraft P, Van Steen K (2012b). Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Human genetics* **131**(10):1591-1613.

Atkinson J, Anker S, Bobier W, Braddick O, Durden K, Nardini M, Watson P (2000). Normal emmetropization in infants with spectacle correction for hyperopia. *Investigative Ophthalmology and Visual Science* **41**(12):3726-3731.

Au Eong KG, Tay TH, Lim MK (1993). Education and myopia in 110,236 young Singaporean males. *Singapore Med J* **34**(6):489-492.

Avinun R (2019). The E is in the G: Gene–environment–trait correlations and findings from genome-wide association studies. *Perspectives on Psychological Science*:1745691619867107.

Ayala FJ, Fitch WM (1997). Genetics and the origin of species: An introduction. *Proceedings of the National Academy of Sciences* **94**(15):7691-7697.

Bateson W (1907). Facts limiting the theory of heritbility. *Science* **26**(672):649.

Beaty TH, Ruczinski I, Murray JC, Marazita ML, Munger RG, Hetmanski JB, Murray T, Redett RJ, Fallin MD, Liang KY, Wu T, Patel PJ, Jin S-C, Zhang TX, Schwender H, Wu-Chou YH, Chen PK, Chong SS, Cheah F, Yeow V, Ye X, Wang H, Huang S, Jabs EW, Shi B, Wilcox AJ, Lie RT, Jee SH, Christensen K, Doheny KF, Pugh EW, Ling H, Scott AF (2011). Evidence for gene-environment interaction in a genome wide study of nonsyndromic cleft palate. *Genetic epidemiology* **35**(6):469-478.

Beyerlein A, von Kries R, Ness AR, Ong KK (2011). Genetic Markers of Obesity Risk: Stronger Associations with Body Composition in Overweight Compared to Normal-Weight Children. *PLoS ONE* **6**(4):e19057.

Border R, Keller MC (2017). Commentary: Fundamental problems with candidate gene-by-environment interaction studies - reflections on Moore and Thoemmes (2016). *Journal of child psychology and psychiatry, and allied disciplines* **58**(3):328-330.

Borner GHH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, Sahlender DA, Robinson MS (2012). Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *The Journal of cell biology* **197**(1):141-160.

Bowden J, Holmes MV (2019). Meta-analysis and Mendelian randomization: A review. *Research Synthesis Methods* **0**(0).

Breslow NE, Day NE, International Agency for Research on C. 1987. *Statistical methods in cancer research. Volume II, Volume II*. Lyon: International Agency for Research on Cancer.

Brown MB, Forsythe AB (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association* **69**(346):364-367.

Buckle A, Nozawa R-s, Kleinjan DA, Gilbert N (2018). Functional characteristics of novel pancreatic Pax6 regulatory elements. *Human Molecular Genetics* **27**(19):3434-3448.

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**:291.

Bureau A, Croteau J, Couture C, Vohl M-C, Bouchard C, Pérusse L (2015). Estimating genetic effect sizes under joint disease-endophenotype models in presence of gene-environment interactions. *Frontiers in Genetics* **6**(248).

Bush WS, Moore JH (2012). Chapter 11: Genome-wide association studies. *PLoS computational biology* **8**(12):e1002822-e1002822.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**(7726):203-209.

Cai X-B, Shen S-R, Chen D-F, Zhang Q, Jin Z-B (2019). An overview of myopia genetics. *Experimental Eye Research* **188**:107778.

Cao Y, Maxwell TJ, Wei P (2015). A family-based joint test for mean and variance heterogeneity for quantitative traits. *Annals of human genetics* **79**(1):46-56.

Cao Y, Wei P, Bailey M, Kauwe JSK, Maxwell TJ (2014). A versatile omnibus test for detecting mean and variance heterogeneity. *Genetic epidemiology* **38**(1):51-59.

Carey CE, Agrawal A, Bucholz KK, Hartz SM, Lynskey MT, Nelson EC, Bierut LJ, Bogdan R (2016). Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Frontiers in genetics* **7**:149-149.

Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L (2006). Epistasis and the release of genetic variation during long-term selection. *Nature Genetics* **38**(4):418-420.

Carthew RW, Sontheimer EJ (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**(4):642-655.

Cattaert T, Urrea V, Naj AC, De Lobel L, De Wit V, Fu M, Mahachie John JM, Shen H, Calle ML, Ritchie MD, Edwards TL, Van Steen K (2010). FAM-MDR: a flexible family-based multifactor dimensionality reduction technique to detect epistasis using related individuals. *PloS one* **5**(4):e10304-e10304.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**(1).

Chatterjee N, Carroll RJ (2005). Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* **92**(2):399-418.

Chatterjee N, Shi J, García-Closas M (2016). Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature reviews. Genetics* **17**(7):392-406.

Chen G-B, Liu N, Klimentidis YC, Zhu X, Zhi D, Wang X, Lou X-Y (2014). A unified GMDR method for detecting gene-gene interactions in family and unrelated samples with application to nicotine dependence. *Human genetics* **133**(2):139-150.

Chen Y-H, Chatterjee N, Carroll RJ (2009a). Shrinkage Estimators for Robust and Efficient Inference in Haplotype-Based Case-Control Studies. *Journal of the American Statistical Association* **104**(485):220-233.

Chen Y-P, Hocking PM, Wang L, Považay B, Prashar A, To C-H, Erichsen JT, Feldkaemper M, Hofer B, Drexler W, Schaeffel F, Guggenheim JA (2011a). Selective Breeding for Susceptibility to Myopia Reveals a Gene–Environment Interaction. *Investigative Ophthalmology & Visual Science* **52**(7):4003-4011.

Chen Y, Han X, Guo X, Li Y, Lee J, He M (2019). Contribution of Genome-Wide Significant Single Nucleotide Polymorphisms in Myopia Prediction: Findings from a 10-year Cohort of Chinese Twin Children. *Ophthalmology*.

Chen YP, Hocking PM, Wang L, Považay B, Prashar A, To CH, Erichsen JT, Feldkaemper M, Hofer B, Drexler W, Schaeffel F, Guggenheim JA (2011b). Selective breeding for susceptibility to myopia reveals a gene-environment interaction. *Investigative Ophthalmology and Visual Science* **52**:4003-4011.

Chen ZT-Y, Wang IJ, Liao Y-T, Shih Y-F, Lin LL-K (2011c). Polymorphisms in steroidogenesis genes, sex steroid levels, and high myopia in the Taiwanese population. *Molecular vision* **17**:2297-2310.

Chen ZT-Y, Wang IJ, Shih Y-F, Lin LL-K (2009b). The Association of Haplotype at the Lumican Gene with High Myopia Susceptibility in Taiwanese Patients. *Ophthalmology* **116**(10):1920-1927.

Choi J, Park T (2013). Multivariate generalized multifactor dimensionality reduction to detect gene-gene interactions. *BMC systems biology* **7 Suppl 6**(Suppl 6):S15-S15.

Chua S, Foster P. 2019. The Economic and Societal Impact of Myopia and High Myopia. pp. 53-63.

Chung W, Chen J, Turman C, Lindstrom S, Zhu Z, Loh P-R, Kraft P, Liang L (2019). Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nature Communications* **10**(1):569.

Coffey CS, Hebert PR, Ritchie MD, Krumholz HM, Gaziano JM, Ridker PM, Brown NJ, Vaughan DE, Moore JH (2004). An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC bioinformatics* **5**:49-49.

Cook RC, Glasscock RE (1951). Refractive and ocular findings in the newborn. *Am J Ophthalmol* **34**(10):1407-1413.

Cordell HJ, Todd JA, Hill NJ, Lord CJ, Lyons PA, Peterson LB, Wicker LS, Clayton DG (2001). Statistical Modeling of Interlocus Interactions in a Complex Disease: Rejection of the Multiplicative Model of Epistasis in Type 1 Diabetes. *Genetics* **158**(1):357.

Corty RW, Kumar V, Tarantino LM, Takahashi JS, Valdar W (2018). Mean-Variance QTL Mapping Identifies Novel QTL for Circadian Activity and Exploratory Behavior in Mice. *G3: Genes|Genomes|Genetics* **8**(12):3783-3790.

Corty RW, Valdar W (2018). QTL Mapping on a Background of Variance Heterogeneity. *G3: Genes|Genomes|Genetics* **8**(12):3767.

Cuellar-Partida G, Lu Y, Kho PF, Hewitt AW, Wichmann HE, Yazar S, Stambolian D, Bailey-Wilson JE, Wojciechowski R, Wang JJ, Mitchell P, Mackey DA, MacGregor S (2016). Assessing the Genetic Predisposition of Education on Myopia: A Mendelian Randomization Study. *Genetic epidemiology* **40**(1):66-72.

Cumberland PM, Bao Y, Hysi PG, Foster PJ, Hammond CJ, Rahi JS, Eyes UKB, Vision C (2015). Frequency and Distribution of Refractive Error in Adult Life: Methodology and Findings of the UK Biobank Study. *PloS one* **10**(10):e0139780-e0139780.

Cumberland PM, Bountziouka V, Rahi JS (2018). Impact of varying the definition of myopia on estimates of prevalence and associations with risk factors: time for an approach that serves research, practice and policy. *British Journal of Ophthalmology* **102**(10):1407.

Dai JY, Kooperberg C, Leblanc M, Prentice RL (2012a). Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika* **99**(4):929-944.

Dai JY, Logsdon BA, Huang Y, Hsu L, Reiner AP, Prentice RL, Kooperberg C (2012b). Simultaneously testing for marginal genetic association and gene-environment interaction. *American journal of epidemiology* **176**(2):164-173.

Davies NM, Holmes MV, Davey Smith G (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**:k601.

de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**(4):e1004219-e1004219.

de Souza CF, Kalloniatis M, Polkinghorne PJ, McGhee CNJ, Acosta ML (2012). Functional activation of glutamate ionotropic receptors in the human peripheral retina. *Experimental Eye Research* **94**(1):71-84.

Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H (2008). Gene–environment interactions for complex traits: definitions, methodological requirements and challenges. *European Journal of Human Genetics* **16**(10):1164-1172.

Deng WQ, Asma S, Paré G (2013). Meta-analysis of SNPs involved in variance heterogeneity using Levene's test for equal variances. *European Journal Of Human Genetics* **22**:427.

Devlin B, Roeder K (1999). Genomic Control for Association Studies. *Biometrics* **55**(4):997-1004.

Dick DM, Agrawal A, Keller MC, Adkins A, Aliev F, Monroe S, Hewitt JK, Kendler KS, Sher KJ (2015). Candidate gene-environment interaction research: reflections and recommendations. *Perspectives on psychological science : a journal of the Association for Psychological Science* **10**(1):37-59.

Ding J, Zhao D, Du R, Zhang Y, Yang H, Liu J, Yan C, Zhang F, Xiong H (2016). Clinical and molecular genetic analysis of a family with late-onset LAMA2-related muscular dystrophy. *Brain and Development* **38**(2):242-249.

Dirani M, Chamberlain M, Shekar SN, Islam AFM, Garoufalis P, Chen CY, Guymer RH, Baird PN (2006). Heritability of Refractive Error and Ocular Biometrics: The Genes in Myopia (GEM) Twin Study. *Investigative Ophthalmology & Visual Science* **47**(11):4756-4761.

Dolgin E (2015). The myopia boom. *Nature* **519**(7543):276-278.

Dorani F, Hu T eds. 2018. *Feature Selection for Detecting Gene-Gene Interactions in Genome-Wide Association Studies*. Applications of Evolutionary Computation. Cham, 2018//. Springer International Publishing.

Dudbridge F (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genetics* **9**(3):e1003348.

Duncan LE, Keller MC (2011). A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *American Journal of Psychiatry* **168**(10):1041-1049.

Ek WE, Rask-Andersen M, Karlsson T, Enroth S, Gyllensten U, Johansson Å (2018). Genetic variants influencing phenotypic variance heterogeneity. *Human Molecular Genetics* **27**(5):799-810.

El-Rashidy MA (2019). An efficient methodology for discovering both of gene-environment interactions and gene-gene interactions causing genetic diseases. *Egyptian Informatics Journal*.

Enthoven CA, Tideman JWL, Polling JR, Tedja MS, Raat H, Iglesias AI, Verhoeven VJM, Klaver CCW (2019). Interaction between lifestyle and genetic susceptibility in myopia: the Generation R study. *European Journal of Epidemiology* **34**(8):777-784.

Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**(D1):D649-d655.

Fan Q, Guo X, Tideman JW, Williams KM, Yazar S, Hosseini SM, Howe LD, Pourcain BS, Evans DM, Timpson NJ, McMahon G, Hysi PG, Krapohl E, Wang YX, Jonas JB, Baird PN, Wang

JJ, Cheng CY, Teo YY, Wong TY, Ding X, Wojciechowski R, Young TL, Parssinen O, Oexle K, Pfeiffer N, Bailey-Wilson JE, Paterson AD, Klaver CC, Plomin R, Hammond CJ, Mackey DA, He M, Saw SM, Williams C, Guggenheim JA (2016a). Childhood gene-environment interactions and age-dependent effects of genetic variants associated with refractive error and myopia: The CREAM Consortium. *Scientific Reports* **6**:25853.

Fan Q, Guo X, Tideman JWL, Williams KM, Yazar S, Hosseini SM, Howe LD, Pourcain BS, Evans DM, Timpson NJ, McMahon G, Hysi PG, Krapohl E, Wang YX, Jonas JB, Baird PN, Wang JJ, Cheng C-Y, Teo Y-Y, Wong T-Y, Ding X, Wojciechowski R, Young TL, Pärssinen O, Oexle K, Pfeiffer N, Bailey-Wilson JE, Paterson AD, Klaver CCW, Plomin R, Hammond CJ, Mackey DA, He M, Saw S-M, Williams C, Guggenheim JA, The CC, Meguro A, Wright AF, Hewitt AW, Young AL, Veluchamy AB, Metspalu A, Paterson AD, Döring A, Khawaja AP, Klein BE, Pourcain BS, Fleck B, Klaver CCW, Hayward C, Williams C, Delcourt C, Pang CP, Khor C-C, Cheng C-Y, Gieger C, Hammond CJ, Simpson CL, van Duijn CM, Mackey DA, Evans DM, Stambolian D, Chew E, Tai ES, Krapohl E, Mihailov E, Smith GD, McMahon G, Biino G, Campbell H, Rudan I, Seppälä I, Kaprio J, Wilson JF, Craig JE, Tideman JWL, Ried JS, Korobelnik J-F, Guggenheim JA, Fondran JR, Wang JJ, Liao J, Zhao JH, Xie J, Bailey-Wilson JE, Kemp JP, Lass JH, Jonas JB, Rahi JS, Wedenoja J, Mäkelä K-M, Burdon KP, Williams KM, Khaw K-T, Yamashiro K, Oexle K, Howe LD, Chen LJ, Xu L, Farrer L, Ikram MK, Deangelis MM, Morrison M, Schache M, Pirastu M, Miyake M, Yap MKH, Fossarello M, Kähönen M, Tedja MS, He M, Yoshimura N, Martin NG, Timpson NJ, Wareham NJ, Mizuki N, Pfeiffer N, Pärssinen O, Raitakari O, Polasek O, Tam PO, Foster PJ, Mitchell P, Baird PN, Chen P, Hysi PG, Cumberland P, Gharahkhani P, Fan Q, Höhn R, Fogarty RD, Luben RN, Igo Jr RP, Plomin R, Wojciechowski R, Klein R, Mohsen Hosseini S, Janmahasatian S, Saw S-M, Yazar S, Ping Yip S, Feng S, Vaccargiu S, Panda-Jonas S, MacGregor S, Iyengar SK, Rantanen T, Lehtimäki T, Young TL, Meitinger T, Wong T-Y, Aung T, Haller T, Vitart V, Nangia V, Verhoeven VJM, Jhanji V, Zhao W, Chen W, Zhou X, Guo X, Ding X, Wang YX, Lu Y, Teo Y-Y, Vatavuk Z (2016b). Childhood gene-environment interactions and age-dependent effects of genetic variants associated with refractive error and myopia: The CREAM Consortium. *Scientific Reports* **6**:25853.

Fan Q, Wojciechowski R, Ikram MK, Cheng CY, Chen P, Zhou X, Pan CW, Khor CC, Tai ES, Aung T, Wong TY, Teo YY, Saw SM (2014). Education influences the association between genetic variants and refractive error: A meta-analysis of five Singapore studies. *Human Molecular Genetics* **23**(2):546-554.

Fischer AJ, Skorupa D, Schonberg DL, Walton NA (2006). Characterization of glucagon-expressing neurons in the chicken retina. *J Comp Neurol* **496**(4):479-494.

Fisher RA (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**(2):399-433.

Fledelius HC (1985). Adult-onset myopia associated with diabetes, a new phenomenon? *Acta Ophthalmologica* **63**(S173):88-88.

Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, Rahi J, Resnikoff S, Vitale S, Yannuzzi L (2019). IMI – Defining and Classifying Myopia: A Proposed Set of Standards for Clinical and Epidemiologic Studies. *Investigative Ophthalmology & Visual Science* **60**(3):M20-M30.

Forsberg SKG, Andreatta ME, Huang X-Y, Danku J, Salt DE, Carlborg Ö (2015). The Multi-allelic Genetic Architecture of a Variance-Heterogeneity Locus for Molybdenum Concentration in Leaves Acts as a Source of Unexplained Additive Genetic Variance. *PLOS Genetics* **11**(11):e1005648.

Forsberg SKG, Bloom JS, Sadhu MJ, Kruglyak L, Carlborg Ö (2017). Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nature Genetics* **49**(4):497-503.

Forsberg SKG, Carlborg Ö (2017). On the relationship between epistasis and genetic variance heterogeneity. *Journal of Experimental Botany* **68**(20):5431-5438.

Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* **186**(9):1026-1034.

Gail M, Simon R (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41**(2):361-372.

Gall WE, Geething NC, Hua Z, Ingram MF, Liu K, Chen SI, Graham TR (2002). Drs2p-dependent formation of exocytic clathrin-coated vesicles in vivo. *Curr Biol* **12**(18):1623-1627.

Garcia-Closas M, Gunsoy NB, Chatterjee N (2014). Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. *Journal of the National Cancer Institute* **106**(11):dju305.

Garcia-Closas M, Rothman N, Figueroa JD, Prokunina-Olsson L, Han SS, Baris D, Jacobs EJ, Malats N, De Vivo I, Albanes D, Purdue MP, Sharma S, Fu Y-P, Kogevinas M, Wang Z, Tang W, Tardón A, Serra C, Carrato A, García-Closas R, Lloreta J, Johnson A, Schwenn M, Karagas MR, Schned A, Andriole G, Jr., Grubb R, 3rd, Black A, Gapstur SM, Thun M, Diver WR, Weinstein SJ, Virtamo J, Hunter DJ, Caporaso N, Landi MT, Hutchinson A, Burdett L, Jacobs KB, Yeager M, Fraumeni JF, Jr., Chanock SJ, Silverman DT, Chatterjee N (2013). Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer research* **73**(7):2211-2220.

Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, Witte JS, Amos C, Tai CG, Conti D, Torgerson DG, Lee S, Chatterjee N (2017). Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. *American journal of epidemiology* **186**(7):762-770.

Gauderman WJ, Thomas DC, Murcray CE, Conti D, Li D, Lewinger JP (2010). Efficient genome-wide association testing of gene-environment interaction in case-parent trios. *American journal of epidemiology* **172**(1):116-122.

Gauderman WJ, Zhang P, Morrison JL, Lewinger JP (2013). Finding novel genes by testing G × E interactions in a genome-wide association study. *Genetic epidemiology* **37**(6):603-613.

Genetics of personality Consortium (2015). Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry* **72**(7):642-650.

Ghorbani Mojarrad N, Williams C, Guggenheim JA (2018). A genetic risk score and number of myopic parents independently predict myopia. *Ophthalmic and Physiological Optics* **38**(5):492-502.

Gianola D (2013). Priors in whole-genome regression: the bayesian alphabet returns. *Genetics* **194**(3):573-596.

Gianola D, Fernando RL, Stella A (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **173**(3):1761-1776.

Gianola D, van Kaam JB (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **178**(4):2289-2303.

Gibson G (2009). Decanalization and the origin of complex disease. *Nat Rev Genet* **10**(2):134-140.

Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg Ö (2007). Statistical Epistasis Is a Generic Feature of Gene Regulatory Networks. *Genetics* **175**(1):411.

Goebel R, Muckli L, Kim D-S. 2012. Chapter 37 - Visual System. In: Mai JK and Paxinos G eds. *The Human Nervous System (Third Edition)*. San Diego: Academic Press, pp. 1301-1327.

Gola D, Mahachie John JM, van Steen K, König IR (2016). A roadmap to multifactor dimensionality reduction methods. *Briefings in bioinformatics* **17**(2):293-308.

Goldman D, Landi P, Ducci F. 2013. Chapter 113 - Addictive Disorders. In: Rimoin D et al. eds. *Emery and Rimoin's Principles and Practice of Medical Genetics*. Oxford: Academic Press, pp. 1-29.

Gordon-Shaag A, Millodot M, Shneor E, Liu Y (2015). The Genetic and Environmental Factors for Keratoconus. *BioMed Research International* **2015**:19.

Grabe HJ, Schwahn C, Mahler J, Schulz A, Spitzer C, Fenske K, Appel K, Barnow S, Nauck M, Schomerus G, Biffar R, Rosskopf D, John U, Völzke H, Freyberger HJ (2012). Moderation of adult depression by the serotonin transporter promoter variant (5-HTTLPR), childhood abuse and adult traumatic events in a general population sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **159B**(3):298-309.

Greene CS, Penrod NM, Kiralis J, Moore JH (2009). Spatially uniform relieff (SURF) for computationally-efficient filtering of gene-gene interactions. *BioData mining* **2**(1):5-5.

Grosvenor T (1987). A review and a suggested classification system for myopia on the basis of age-related prevalence and age of onset. *Am J Optom Physiol Opt* **64**(7):545-554.

Grover S, Del Greco M F, König IR (2018). Evaluating the current state of Mendelian randomization studies: a protocol for a systematic review on methodological and clinical aspects using neurodegenerative disorders as outcome. *Systematic Reviews* **7**(1):145.

Guggenheim JA, Ghorbani Mojarrad N, Williams C, Flitcroft DI (2017). Genetic prediction of myopia: prospects and challenges. *Ophthalmic and Physiological Optics* **37**(5):549-556.

Guggenheim JA, Kirov G, Hodson SA (2000). The heritability of high myopia: a reanalysis of Goldschmidt's data. *J Med Genet* **37**(3):227-231.

Guggenheim JA, McMahon G, Northstone K, Mandel Y, Kaiserman I, Stone RA, Lin X, Saw SM, Forward H, Mackey DA, Yazar S, Young TL, Williams C (2013a). Birth order and myopia. *Ophthalmic epidemiology* **20**(6):375-384.

Guggenheim JA, Northstone K, McMahon G, Ness AR, Deere K, Mattocks C, Pourcain BS, Williams C (2012). Time outdoors and physical activity as predictors of incident myopia in childhood: a prospective cohort study. *Invest Ophthalmol Vis Sci* **53**(6):2856-2865.

Guggenheim JA, Zhou X, Evans DM, Timpson NJ, McMahon G, Kemp JP, St Pourcain B, Northstone K, Ring SM, Fan Q, Wong T-Y, Cheng CY, Khor CC, Aung T, Saw SM, Williams C (2013b). Coordinated genetic scaling of the human eye: Shared determination of axial eye length and corneal curvature. *Investigative Ophthalmology and Visual Science* **54**(3):1715-1721.

Gui J, Andrew AS, Andrews P, Nelson HM, Kelsey KT, Karagas MR, Moore JH (2011a). A robust multifactor dimensionality reduction method for detecting gene-gene interactions with application to the genetic analysis of bladder cancer susceptibility. *Annals of human genetics* **75**(1):20-28.

Gui J, Moore JH, Kelsey KT, Marsit CJ, Karagas MR, Andrew AS (2011b). A novel survival multifactor dimensionality reduction method for detecting gene-gene interactions with application to bladder cancer prognosis. *Human genetics* **129**(1):101-110.

Gui J, Moore JH, Williams SM, Andrews P, Hillege HL, van der Harst P, Navis G, Van Gilst WH, Asselbergs FW, Gilbert-Diamond D (2013). A Simple and Computationally Efficient

Approach to Multifactor Dimensionality Reduction Analysis of Gene-Gene Interactions for Quantitative Traits. *PloS one* **8**(6):e66545-e66545.

Guoping L, Xiang Y, Jianfeng W, Dadong G, Jie H, Wenjun J, Junguo G, Hongsheng B (2017). Alterations of Glutamate and γ-Aminobutyric Acid Expressions in Normal and Myopic Eye Development in Guinea Pigs. *Investigative Ophthalmology & Visual Science* **58**(2):1256-1265.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**:186.

Hallgrímsdóttir IB, Yuster DS (2008). A complete classification of epistatic two-locus models. *BMC genetics* **9**:17-17.

Hammond CJ, Snieder H, Gilbert CE, Spector TD (2001). Genes and environment in refractive error: the twin eye study. *Invest Ophthalmol Vis Sci* **42**(6):1232-1236.

Hancock DB, Soler Artigas M, Gharib SA, Henry A, Manichaikul A, Ramasamy A, Loth DW, Imboden M, Koch B, McArdle WL, Smith AV, Smolonska J, Sood A, Tang W, Wilk JB, Zhai G, Zhao JH, Aschard H, Burkart KM, Curjuric I, Eijgelsheim M, Elliott P, Gu X, Harris TB, Janson C, Homuth G, Hysi PG, Liu JZ, Loehr LR, Lohman K, Loos RJ, Manning AK, Marciante KD, Obeidat M, Postma DS, Aldrich MC, Brusselle GG, Chen TH, Eiriksdottir G, Franceschini N, Heinrich J, Rotter JI, Wijmenga C, Williams OD, Bentley AR, Hofman A, Laurie CC, Lumley T, Morrison AC, Joubert BR, Rivadeneira F, Couper DJ, Kritchevsky SB, Liu Y, Wjst M, Wain LV, Vonk JM, Uitterlinden AG, Rochat T, Rich SS, Psaty BM, O'Connor GT, North KE, Mirel DB, Meibohm B, Launer LJ, Khaw KT, Hartikainen AL, Hammond CJ, Glaser S, Marchini J, Kraft P, Wareham NJ, Volzke H, Stricker BH, Spector TD, Probst-Hensch NM, Jarvis D, Jarvelin MR, Heckbert SR, Gudnason V, Boezen HM, Barr RG, Cassano PA, Strachan DP, Fornage M, Hall IP, Dupuis J, Tobin MD, London SJ (2012). Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* **8**(12):e1003098.

He M, Xiang F, Zeng Y, Mai J, Chen Q, Zhang J, Smith W, Rose K, Morgan IG (2015). Effect of Time Spent Outdoors at School on the Development of Myopia Among Children in China: A Randomized Clinical Trial. *JAMA* **314**(11):1142-1148.

Hemani G, Shakhbazov K, Westra H-J, Esko T, Henders AK, McRae AF, Yang J, Gibson G, Martin NG, Metspalu A, Franke L, Montgomery GW, Visscher PM, Powell JE (2014). Detection and replication of epistasis influencing transcription in humans. *Nature* **508**:249.

Hemani G, Zheng J, Elsworth B, Wade K, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan V, Yarmolinsky J, Shihab H, Timpson N, Evans D, Relton C, Martin R, Smith G, Gaunt T, Haycock P (2018). The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**.

Henderson CR (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**(2):423-447.

Hendriks M, Verhoeven VJM, Buitendijk GHS, Polling JR, Meester-Smoor MA, Hofman A, van Huet RA, Klevering BJ, Bax NM, Lambertus S, Klaver CCW, Hoyng CB, Oomen CJ, van Zelst-Stams WA, Cremers FP, Plomp AS, van Schooneveld MJ, van Genderen MM, Schuil J, Boonstra FN, Schlingemann RO, Bergen AA, Pierrache L, Meester-Smoor M, van den Born LI, Boon CJ, Pott JWR, van Leeuwen R, Kroes HY, de Jong-Hesse Y, Kamermans M, Ingeborgh van den Born L, Klaver CCW (2017). Development of Refractive Errors—What Can We Learn From Inherited Retinal Dystrophies? *American Journal of Ophthalmology* **182**:81-89.

Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H, Armstrong J, Bairoch A, Cesareni G,

Sherman D, Apweiler R (2004). IntAct: an open source molecular interaction database. *Nucleic acids research* **32**(Database issue):D452-D455.

Hickey JM, Kinghorn BP, Tier B, Clark SA, van der Werf JH, Gorjanc G (2013). Genomic evaluations using similarity between haplotypes. *J Anim Breed Genet* **130**(4):259-269.

Hill WG, Goddard ME, Visscher PM (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLOS Genetics* **4**(2):e1000008.

Hill WG, Mulder HA (2010). Genetic analysis of environmental variation. *Genetics Research* **92**(5-6):381-395.

Hirst J, Sahlender DA, Li S, Lubben NB, Borner GHH, Robinson MS (2008). Auxilin depletion causes self-assembly of clathrin into membraneless cages in vivo. *Traffic (Copenhagen, Denmark)* **9**(8):1354-1371.

Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, Wong TY, Naduvilath TJ, Resnikoff S (2016). Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. *Ophthalmology* **123**(5):1036-1042.

Hornbeak DM, Young TL (2009). Myopia genetics: a review of current research and emerging trends. *Current opinion in ophthalmology* **20**(5):356-362.

Hou T-T, Lin F, Bai S, Cleves MA, Xu H-M, Lou X-Y (2019). Generalized multifactor dimensionality reduction approaches to identification of genetic interactions underlying ordinal traits. *Genetic Epidemiology* **43**(1):24-36.

Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C (2012). Powerful Cocktail Methods for Detecting Genome-Wide Gene-Environment Interaction. *Genetic Epidemiology* **36**(3):183-194.

Hua X, Zhang H, Zhang H, Yang Y, Kuk AY (2010). Testing multiple gene interactions by the ordered combinatorial partitioning method in case-control studies. *Bioinformatics* **26**(15):1871-1878.

Huang W, Mackay TFC (2016). The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS genetics* **12**(11):e1006421-e1006421.

Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM, Tank NCIG-ET (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genetic epidemiology* **37**(7):643-657.

Ilegems E, Dicker A, Speier S, Sharma A, Bahow A, Edlund PK, Leibiger IB, Berggren PO (2013) Reporter islets in the eye reveal the plasticity of the endocrine pancreas. Proceedings of the National Academy of Sciences 110(51):20581-20586.

Iribarren R (2015). Crystalline lens and refractive development. *Progress in Retinal and Eye Research* **47**:86-106.

Iyer JV, Low WCJ, Dirani M, Saw SM (2012). Parental smoking and childhood refractive error: the STARS study. *Eye (London, England)* **26**(10):1324-1328.

VanderWeele JT, Knol JM. 2014. *A Tutorial on Interaction*.

Jaffee SR, Price TS (2007). Gene–environment correlations: a review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry* **12**(5):432-442.

Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins J, Muñoz-Manchado AB, Nagel M, Savage JE, Tiemeier H, White T, Tung JY, Hinds DA, Vacic V, Sullivan PF, van der Sluis S, Polderman TJC, Smit AB, Hjerling-Leffler J, Van Someren

EJW, Posthuma D (2018). Genome-wide Analysis of Insomnia (N=1,331,010) Identifies Novel Loci and Functional Pathways. *bioRxiv*:214973.

Janssens K, ten Dijke P, Ralston SH, Bergmann C, Van Hul W (2003). Transforming growth factor-beta 1 mutations in Camurati-Engelmann disease lead to increased signaling by altering either activation or secretion of the mutant protein. *J Biol Chem* **278**(9):7718-7724.

Jia L-Y, Tam PO-S, Chiang SW-Y, Ding N, Chen LJ, Yam GH-F, Pang C-P, Wang N-L (2009). Multiple gene polymorphisms analysis revealed a different profile of genetic polymorphisms of primary open-angle glaucoma in northern Chinese. *Molecular vision* **15**:89-98.

Jiang D, Li J, Xiao X, Li S, Jia X, Sun W, Guo X, Zhang Q (2015). Detection of Mutations in LRPAP1, CTSH, LEPREL1, ZNF644, SLC39A5, and SCO2 in 298 Families With Early-Onset High Myopia by Exome Sequencing. *Investigative Ophthalmology & Visual Science* **56**(1):339-345.

Jiao S, Hsu L, Bézieau S, Brenner H, Chan AT, Chang-Claude J, Le Marchand L, Lemire M, Newcomb PA, Slattery ML, Peters U (2013). SBERIA: Set-Based Gene-Environment Interaction Test for Rare and Common Variants in Complex Diseases. *Genetic Epidemiology* **37**(5):452-464.

Jiao S, Peters U, Berndt S, Bézieau S, Brenner H, Campbell PT, Chan AT, Chang-Claude J, Lemire M, Newcomb PA, Potter JD, Slattery ML, Woods MO, Hsu L (2015). Powerful Set-Based Gene-Environment Interaction Testing Framework for Complex Diseases. *Genetic epidemiology* **39**(8):609-618.

Jones LA, Sinnott LT, Mutti DO, Mitchell GL, Moeschberger ML, Zadnik K (2007). Parental history of myopia, sports and outdoor activities, and future myopia. *Invest Ophthalmol Vis Sci* **48**(8):3524-3532.

Jung H-Y, Leem S, Lee S, Park T (2016). A novel fuzzy set based multifactor dimensionality reduction method for detecting gene–gene interaction. *Computational Biology and Chemistry* **65**:193-202.

Jung H-Y, Leem S, Park T (2018). Fuzzy set-based generalized multifactor dimensionality reduction analysis of gene-gene interactions. *BMC Medical Genomics* **11**(2):32.

Kalbfleisch J, Prentice R. 2011. The Statistical Analysis of Failure Time Data, Second Edition. pp. 247-277.

Kamaraj B, Purohit R (2014). Mutational analysis of oculocutaneous albinism: a compact review. *BioMed research international* **2014**:905472-905472.

Karg K, Burmeister M, Shedden K, Sen S (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry* **68**(5):444-454.

Kaur H, Kumari V (2018). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics*.

Keller MC (2014). Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biological Psychiatry* **75**(1):18-24.

Kendler KS, Baker JH (2007). Genetic influences on measures of the environment: a systematic review. *Psychol Med* **37**(5):615-626.

Kerin M, Marchini J (2019). Gene-environment interactions using a Bayesian whole genome regression model. *bioRxiv*:797829.

Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S (2018). Genome-wide polygenic scores for common diseases

identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**(9):1219-1224.

Kiefer AK, Tung JY, Do CB, Hinds DA, Mountain JL, Francke U, Eriksson N (2013). Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS genetics* **9**(2):e1003299-e1003299.

Kim H, Jeong H-B, Jung H-Y, Park T, Park M (2019). Multivariate Cluster-Based Multifactor Dimensionality Reduction to Identify Genetic Interactions for Multiple Quantitative Phenotypes. *BioMed Research International* **2019**:10.

Klahr AM, Thomas KM, Hopwood CJ, Klump KL, Burt SA (2013). Evocative gene-environment correlation in the mother-child relationship: a twin study of interpersonal processes. *Development and psychopathology* **25**(1):105-118.

Klengel T, Binder EB (2013). Gene × environment interactions in the prediction of response to antidepressant treatment. *International Journal of Neuropsychopharmacology* **16**(3):701-711.

Knights J, Yang J, Chanda P, Zhang A, Ramanathan M (2013). SYMPHONY, an information-theoretic method for gene-gene and gene-environment interaction analysis of disease syndromes. *Heredity (Edinb)* **110**(6):548-559.

Knol MJ, VanderWeele TJ (2011). Recoding preventive exposures to get valid measures of interaction on an additive scale. *European journal of epidemiology* **26**(10):825-826.

Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE (2011). Estimating measures of interaction on an additive scale for preventive exposures. *European journal of epidemiology* **26**(6):433-438.

Koenker R, Hallock KF (2001). Quantile Regression. *Journal of Economic Perspectives* **15**(4):143-156.

Kondo Y, Koshimizu E, Megarbane A, Hamanoue H, Okada I, Nishiyama K, Kodera H, Miyatake S, Tsurusaki Y, Nakashima M, Doi H, Miyake N, Saitsu H, Matsumoto N (2013). Whole-exome sequencing identified a homozygous FNBP4 mutation in a family with a condition similar to microphthalmia with limb anomalies. *American Journal of Medical Genetics Part A* **161**(7):1543-1546.

Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, Oddsson A, Halldorsson BV, Masson G, Gudbjartsson DF, Helgason A, Bjornsdottir G, Thorsteinsdottir U, Stefansson K (2018). The nature of nurture: Effects of parental genotypes. *Science* **359**(6374):424.

Korte A, Farlow A (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* **9**(1):29.

Kraft P, Aschard H (2015). Finding the missing gene-environment interactions. *European journal of epidemiology* **30**(5):353-355.

Kraft P, Yen Y-C, Stram D, Morrison J, Gauderman W (2007). Exploiting Gene-Environment Interaction to Detect Genetic Associations. *Human heredity* **63**:111-119.

Kurtz D, Hyman L, Gwiazda JE, Manny R, Dong LM, Wang Y, Scheiman M (2007). Role of parental myopia in the progression of myopia and its interaction with treatment in COMET children. *Invest Ophthalmol Vis Sci* **48**(2):562-570.

Le Rouzic A, Carlborg Ö (2008). Evolutionary potential of hidden genetic variation. *Trends in Ecology & Evolution* **23**(1):33-37.

Le Rouzic A, Siegel PB, Carlborg O (2007). Phenotypic evolution from genetic polymorphisms in a radial network architecture. *BMC biology* **5**:50-50.

Lee S, Kim Y, Kwon M-S, Park T (2015). A Comparative Study on Multifactor Dimensionality Reduction Methods for Detecting Gene-Gene Interactions with the Survival Phenotype. *BioMed research international* **2015**:671859-671859.

Lee S, Kwon M-S, Oh JM, Park T (2012). Gene-gene interaction analysis for the survival phenotype based on the Cox model. *Bioinformatics (Oxford, England)* **28**(18):i582-i588.

Lee S, Son D, Kim Y, Yu W, Park T (2018). Unified Cox model based multifactor dimensionality reduction method for gene-gene interaction analysis of the survival phenotype. *BioData Mining* **11**(1):27.

Lee Y, Nelder JA (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **55**(2):139-185.

Levy SE, Myers RM (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics* **17**(1):95-115.

Li D, Conti DV (2009). Detecting gene-environment interactions using a combined case-only and case-control approach. *American journal of epidemiology* **169**(4):497-504.

Li N, Xu Y, Li G, Yu T, Yao R-E, Wang X, Wang J (2017). Exome sequencing identifies a de novo mutation of CTNNB1 gene in a patient mainly presented with retinal detachment, lens and vitreous opacities, microcephaly, and developmental delay: Case report and literature review. *Medicine* **96**(20):e6914-e6914.

Liew SH, Elsner H, Spector TD, Hammond CJ (2005). The first "classical" twin study? Analysis of refractive error using monozygotic and dizygotic twins published in 1922. *Twin Res Hum Genet* **8**(3):198-200.

Lim LS, Gazzard G, Low YL, Choo R, Tan DT, Tong L, Yin Wong T, Saw SM (2010). Dietary factors, myopia, and axial dimensions in children. *Ophthalmology* **117**(5):993-997.e994.

Lin W-Y, Huang C-C, Liu Y-L, Tsai S-J, Kuo P-H (2018). Polygenic approaches to detect gene–environment interactions when external information is unavailable. *Briefings in Bioinformatics*.

Lin W-Y, Huang C-C, Liu Y-L, Tsai S-J, Kuo P-H (2019). Genome-Wide Gene-Environment Interaction Analysis Using Set-Based Association Tests. *Frontiers in Genetics* **9**(715).

Lin X, Lee S, Christiani DC, Lin X (2013). Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics* **14**(4):667-681.

Lin X, Lee S, Wu MC, Wang C, Chen H, Li Z, Lin X (2016). Test for rare variants by environment interactions in sequencing association studies. *Biometrics* **72**(1):156-164.

Lin Y, Jones BW, Liu A, Vazquéz-Chona FR, Lauritzen JS, Ferrell WD, Marc RE (2012). Rapid glutamate receptor 2 trafficking during retinal degeneration. *Molecular neurodegeneration* **7**:7-7.

Liu Q, Chen LS, Nicolae DL, Pierce BL (2016). A unified set-based test with adaptive filtering for gene-environment interaction analyses. *Biometrics* **72**(2):629-638.

Liu Q, Qiu X (1998). [Vitrectomy for treatment of retinal detachment caused by macular hole]. *Zhonghua Yan Ke Za Zhi* **34**(6):418-420.

Lor GCY, Risch HA, Fung WT, Au Yeung SL, Wong IOL, Zheng W, Pang H (2019). Reporting and guidelines for mendelian randomization analysis: A systematic review of oncological studies. *Cancer Epidemiology* **62**:101577.

Lou X-Y, Chen G-B, Yan L, Ma JZ, Mangold JE, Zhu J, Elston RC, Li MD (2008). A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *American journal of human genetics* **83**(4):457-467.

Lou X-Y, Chen G-B, Yan L, Ma JZ, Zhu J, Elston RC, Li MD (2007). A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. *American journal of human genetics* **80**(6):1125-1137.

Lyhne N, Sjølie AK, Kyvik KO, 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. *The British journal of ophthalmology* **85**(12):1470-1476.

Ma S, Xu S (2015). Semiparametric nonlinear regression for detecting gene and environment interactions. *Journal of Statistical Planning and Inference* **156**:31-47.

Ma S, Yang L, Romero R, Cui Y (2011). Varying coefficient model for gene–environment interaction: a non-linear look. *Bioinformatics* **27**(15):2119-2126.

Mackay TFC (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature reviews. Genetics* **15**(1):22-33.

MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**:144.

Manning AK, LaValley M, Liu CT, Rice K, An P, Liu Y, Miljkovic I, Rasmussen-Torvik L, Harris TB, Province MA, Borecki IB, Florez JC, Meigs JB, Cupples LA, Dupuis J (2011). Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. *Genet Epidemiol* **35**(1):11-18.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM (2009). Finding the missing heritability of complex diseases. *Nature* **461**(7265):747-753.

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics* **100**(4):635-649.

Martin ER, Ritchie MD, Hahn L, Kang S, Moore JH (2006). A novel method to identify gene–gene effects in nuclear families: the MDR-PDT. *Genetic Epidemiology* **30**(2):111-123.

Mathur MB, VanderWeele TJ (2018). R Function for Additive Interaction Measures. *Epidemiology* **29**(1):e5-e6.

McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, Conti D, Gauderman WJ, Hsu L, Hutter CM, Jankowska MM, Kerr J, Kraft P, Montgomery SB, Mukherjee B, Papanicolaou GJ, Patel CJ, Ritchie MD, Ritz BR, Thomas DC, Wei P, Witte JS (2017). Current Challenges and New Opportunities for Gene-Environment Interaction Studies of Complex Diseases. *American journal of epidemiology* **186**(7):753-761.

Mechanic LE, Chen H-S, Amos CI, Chatterjee N, Cox NJ, Divi RL, Fan R, Harris EL, Jacobs K, Kraft P, Leal SM, McAllister K, Moore JH, Paltoo DN, Province MA, Ramos EM, Ritchie MD, Roeder K, Schaid DJ, Stephens M, Thomas DC, Weinberg CR, Witte JS, Zhang S, Zöllner S, Feuer EJ, Gillanders EM (2012). Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. *Genetic epidemiology* **36**(1):22-35.

Mei H, Ma D, Ashley-Koch A, Martin ER (2005). Extension of multifactor dimensionality reduction for identifying multilocus effects in the GAW14 simulated data. *BMC genetics* **6 Suppl 1**(Suppl 1):S145-S145.

Mertz JR, Wallman J (2000). Choroidal retinoic acid synthesis: a possible mediator between refractive error and compensatory eye growth. *Exp Eye Res* **70**(4):519-527.

Meuwissen TH, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**(4):1819-1829.

Meuwissen TH, Odegard J, Andersen-Ranberg I, Grindflek E (2014). On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genet Sel Evol* **46**:49.

Mitrea C, Taghavi Z, Bokanizad B, Hanoudi S, Tagett R, Donato M, Voichiţa C, Drăghici S (2013). Methods and approaches in the topology-based analysis of biological pathways. *Frontiers in physiology* **4**:278-278.

Montojo J, Zuberi K, Rodriguez H, Bader GD, Morris Q (2014). GeneMANIA: Fast gene network construction and function prediction for Cytoscape. *F1000Research* **3**:153-153.

Moore JH (2003). The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered* **56**(1-3):73-82.

Moore JH, Ritchie MD (2004). STUDENTJAMA. The challenges of whole-genome approaches to common diseases. *Jama* **291**(13):1642-1643.

Moore JH, Williams SM (2002). New strategies for identifying gene-gene interactions in hypertension. *Ann Med* **34**(2):88-95.

Moore JH, Williams SM (2005). Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *Bioessays* **27**(6):637-646.

Moore JH, Williams SM (2009). Epistasis and its implications for personal genetics. *American journal of human genetics* **85**(3):309-320.

Moore R, Casale FP, Jan Bonder M, Horta D, Heijmans BT, C.'t Hoen PA, van Meurs J, Isaacs A, Jansen R, Franke L, Boomsma DI, Pool R, van Dongen J, Hottenga JJ, van Greevenbroek MMJ, Stehouwer CDA, van der Kallen CJH, Schalkwijk CG, Wijmenga C, Zhernakova A, Tigchelaar EF, Slagboom PE, Beekman M, Deelen J, van Heemst D, Veldink JH, van den Berg LH, van Duijn CM, Hofman BA, Uitterlinden AG, Jhamai PM, Verbiest M, Suchiman HED, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Mei H, van Iterson M, Galen Mv, Bot J, van't Hof P, Deelen P, Nooren I, Moed M, Vermaat M, Zhernakova DV, Luijk R, Jan Bonder M, van Dijk F, Arindrarto W, Kielbasa SM, Swertz MA, van Zwet EW, Franke L, Barroso I, Stegle O, Consortium B (2019). A linear mixed-model approach to study multivariate gene–environment interactions. *Nature Genetics* **51**(1):180-186.

Mordechai S, Gradstein L, Pasanen A, Ofir R, El Amour K, Levy J, Belfair N, Lifshitz T, Joshua S, Narkis G, Elbedour K, Myllyharju J, Birk Ohad S (2011). High Myopia Caused by a Mutation in LEPREL1, Encoding Prolyl 3-Hydroxylase 2. *The American Journal of Human Genetics* **89**(3):438-445.

Morgan IG, Ashby RS, Nickla DL (2013). Form deprivation and lens-induced myopia: are they different? *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)* **33**(3):355-361.

Morgan IG, Ding X, Guo X (2018). An overview of the myopia problem in China. *Annals of Eye Science* **3**(9).

Morgan IG, Ohno-Matsui K, Saw S-M (2012). Myopia. *The Lancet* **379**(9827):1739-1748.

Morgan IG, Rose KA (2013). Myopia and international educational performance. *Ophthalmic and Physiological Optics* **33**(3):329-338.

Morgan IG, Rose KA (2019). Myopia: is the nature-nurture debate finally over? *Clin Exp Optom* **102**(1):3-17.

Morgan JR, Prasad K, Jin S, Augustine GJ, Lafer EM (2001). Uncoating of Clathrin-Coated Vesicles in Presynaptic Terminals: Roles for Hsc70 and Auxilin. *Neuron* **32**(2):289-300.

Morota G, Abdollahi-Arpanahi R, Kranis A, Gianola D (2014). Genome-enabled prediction of quantitative traits in chickens using genomic annotation. *BMC Genomics* **15**:109.

Motsinger AA, Ritchie MD (2006). Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Human genomics* **2**(5):318-328.

Mountjoy E, Davies NM, Plotnikov D, Smith GD, Rodriguez S, Williams CE, Guggenheim JA, Atan D (2018). Education and myopia: assessing the direction of causality by mendelian randomisation. *BMJ* **361**:k2022.

Mukherjee B, Chatterjee N (2008). Exploiting Gene-Environment Independence for Analysis of Case–Control Studies: An Empirical Bayes-Type Shrinkage Estimator to Trade-Off between Bias and Efficiency. *Biometrics* **64**(3):685-694.

Mullins N, Power RA, Fisher HL, Hanscombe KB, Euesden J, Iniesta R, Levinson DF, Weissman MM, Potash JB, Shi J, Uher R, Cohen-Woods S, Rivera M, Jones L, Jones I, Craddock N, Owen MJ, Korszun A, Craig IW, Farmer AE, McGuffin P, Breen G, Lewis CM (2016). Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychological medicine* **46**(4):759-770.

Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ (2011). Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genetic epidemiology* **35**(3):201-210.

Murcray CE, Lewinger JP, Gauderman WJ (2008). Gene-Environment Interaction in Genome-Wide Association Studies. *American Journal of Epidemiology* **169**(2):219-226.

Mutti DO, Mitchell GL, Moeschberger ML, Jones LA, Zadnik K (2002). Parental myopia, near work, school achievement, and children's refractive error. *Invest Ophthalmol Vis Sci* **43**(12):3633-3640.

Mutti DO, Mitchell GL, Sinnott LT, Jones-Jordan LA, Moeschberger ML, Cotter SA, Kleinstein RN, Manny RE, Twelker JD, Zadnik K (2012). Corneal and crystalline lens dimensions before and after myopia onset. *Optometry and Vision Science* **89**(3):251–262.

Myles S, Peiffer J, Brown P, Ersoz E, Zhang Z, Costich D, Buckler E (2009). Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant cell* **21**:2194-2202.

NCBI Resource Coordinators (2016). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **45**(D1):D12-D17.

Nelson RM, Pettersson ME, Carlborg O (2013). A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet* **29**(12):669-676.

Neuman RJ, Rice JP, Chakravarti A (1992). Two-Locus models of disease. *Genetic Epidemiology* **9**(5):347-365.

Nickels S, Hopf S, Pfeiffer N, Schuster AK (2019). Myopia is associated with education: Results from NHANES 1999-2008. *PLOS ONE* **14**(1):e0211196.

Niu A, Zhang S, Sha Q (2011). A novel method to detect gene-gene interactions in structured populations: MDR-SP. *Ann Hum Genet* **75**(6):742-754.

Norton B, Pearson ES (1976). A note on the background to, and refereeing of, R. A. Fisher's 1918 paper 'On the correlation between relatives on the supposition of Mendelian inheritance'. *Notes Rec R Soc Lond* **31**(1):151-162.

Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, Dermitzakis ET, Consortium GT (2017). Estimating the causal tissues for complex traits and diseases. *Nature Genetics* **49**(12):1676-1683.

Ottman R (1990). An epidemiologic approach to gene-environment interaction. *Genetic epidemiology* **7**(3):177-185.

Ottman R (1996). Gene-environment interaction: definitions and study designs. *Preventive medicine* **25**(6):764-770.

Pan C-W, Ramamurthy D, Saw S-M (2012). Worldwide prevalence and risk factors for myopia. *Ophthalmic and Physiological Optics* **32**(1):3-16.

Panagiotou ES, Sanjurjo Soriano C, Poulter JA, Lord EC, Dzulova D, Kondo H, Hiyoshi A, Chung BH-Y, Chu YW-Y, Lai CHY, Tafoya ME, Karjosukarso D, Collin RWJ, Topping J, Downey LM, Ali M, Inglehearn CF, Toomes C (2017). Defects in the Cell Signaling Mediator β-Catenin Cause the Retinal Vascular Condition FEVR. *American journal of human genetics* **100**(6):960-968.

Pare G, Cook NR, Ridker PM, Chasman DI (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet* **6**(6):e1000981.

Park S, Kim W (2017). Multifator dimensionality reduction method based on area under receiver operating characteristic curve. *Model Assisted Statistics and Applications* **12**:245-253.

Parker M, Bull SJ, de Vries J, Agbenyega T, Doumbo OK, Kwiatkowski DP (2009). Ethical data release in genome-wide association studies in developing countries. *PLoS medicine* **6**(11):e1000143-e1000143.

Parmeggiani F, Barbaro V, De Nadai K, Lavezzo E, Toppo S, Chizzolini M, Palù G, Parolin C, Di Iorio E (2016). Identification of novel X-linked gain-of-function RPGR-ORF15 mutation in Italian family with retinitis pigmentosa and pathologic myopia. *Scientific Reports* **6**:39179.

Parssinen TO (1987). Relation between refraction, education, occupation, and age among 26- and 46-year-old Finns. *Am J Optom Physiol Opt* **64**(2):136-143.

Patel CJ, Ioannidis JPA (2014). Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *Journal of epidemiology and community health* **68**(11):1096-1100.

Patel CJ, Manrai AK (2015). Development of exposome correlation globes to map out environment-wide associations. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* **20**:231-242.

Pattin KA, White BC, Barney N, Gui J, Nelson HH, Kelsey KT, Andrew AS, Karagas MR, Moore JH (2009). A computationally efficient hypothesis testing method for epistasis analysis using multifactor dimensionality reduction. *Genetic epidemiology* **33**(1):87-94.

Pelov I, Teltsh O, Greenbaum L, Rigbi A, Kanyas-Sarner K, Lerer B, Lombroso P, Kohn Y (2012). Involvement of PTPN5, the gene encoding the striatal-enriched protein tyrosine phosphatase, in schizophrenia and cognition. *Psychiatr Genet* **22**(4):168-176.

Piegorsch WW, Weinberg CR, Taylor JA (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* **13**(2):153-162.

Pietz J, Kreis R, Rupp A, Mayatepek E, Rating D, Boesch C, Bremer HJ (1999). Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. *The Journal of clinical investigation* **103**(8):1169-1178.

Platt A, Vilhjálmsson BJ, Nordborg M (2010). Conditions Under Which Genome-Wide Association Studies Will be Positively Misleading. *Genetics* **186**(3):1045.

Plotnikov D, Shah RL, Rodrigues JN, Cumberland PM, Rahi JS, Hysi PG, Atan D, Williams C, Guggenheim JA, Eye UKB, Vision C (2019). A commonly occurring genetic variant within the NPLOC4–TSPAN10–PDE6G gene cluster is associated with the risk of strabismus. *Human Genetics*.

Pośpiech E, Draus-Barini J, Kupiec T, Wojas-Pelc A, Branicki W (2011). Gene–gene interactions contribute to eye colour variation in humans. *Journal of Human Genetics* **56**(6):447-455.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**(3):559-575.

Rask-Andersen M, Karlsson T, Ek WE, Johansson Å (2017). Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLOS Genetics* **13**(9):e1006977.

Read SA, Collins MJ, Vincent SJ (2014). Light exposure and physical activity in myopic and emmetropic children. *Optom Vis Sci* **91**(3):330-341.

Read SA, Collins MJ, Vincent SJ (2015). Light Exposure and Eye Growth in ChildhoodLight Exposure and Eye Growth in Childhood. *Investigative Ophthalmology & Visual Science* **56**(11):6779-6787.

Risch N, Herrell R, Lehner T, Liang K-Y, Eaves L, Hoh J, Griem A, Kovacs M, Ott J, Merikangas KR (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* **301**(23):2462-2471.

Ritchie MD, Davis JR, Aschard H, Battle A, Conti D, Du M, Eskin E, Fallin MD, Hsu L, Kraft P, Moore JH, Pierce BL, Bien SA, Thomas DC, Wei P, Montgomery SB (2017). Incorporation of Biological Knowledge Into the Study of Gene-Environment Interactions. *American journal of epidemiology* **186**(7):771-777.

Ritchie MD, Hahn LW, Moore JH (2003). Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genetic Epidemiology* **24**(2):150-157.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American journal of human genetics* **69**(1):138-147.

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D (2015). Methods of integrating data to uncover genotype–phenotype interactions. *Nature Reviews Genetics* **16**:85.

Ritz BR, Chatterjee N, Garcia-Closas M, Gauderman WJ, Pierce BL, Kraft P, Tanner CM, Mechanic LE, McAllister K (2017). Lessons Learned From Past Gene-Environment Interaction Successes. *American journal of epidemiology* **186**(7):778-786.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**(1):77.

Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, Peyrot WJ, Abdellaoui A, Zietsch BP, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, The LifeLines Cohort S, Alizadeh BZ, Boezen HM, Franke L, Harst Pvd, Navis G, Rots M, Snieder H, Swertz M, Wolffenbuttel BHR, Wijmenga C, Genetic Investigation of Anthropometric Traits c, Abecasis GR, Absher D, Alavere H, Albrecht E, Allen HL, Almgren P, Amin N, Amouyel P, Anderson D, Arnold AM, Arveiler D, Aspelund T, Asselbergs FW, Assimes TL, Atalay M, Attwood AP, Atwood LD, Bakker SJL, Balkau B, Balmforth AJ, Barlassina C, Barroso I, Basart H, Bauer S, Beckmann JS, Beilby JP, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Berndt SI, Biffar R, Di Blasio AM, Boehm BO, Boehnke M, Boeing H, Boerwinkle E, Bolton JL, Bonnefond A, Bonnycastle LL, Boomsma DI, Borecki IB, Bornstein SR, Bouatia-Naji N, Boucher G, Bragg-Gresham JL, Brambilla P, Bruinenberg M, Buchanan TA, Buechler C, Cadby G, Campbell H, Caulfield MJ, Cavalcanti-Proença C, Cesana G, Chanock SJ, Chasman DI, Chen Y-DI, Chines PS, Clegg DJ, Coin L, Collins FS, Connell JM, Cookson W, Cooper MN, Croteau-Chonka DC, Cupples LA, Cusi D, Day FR, Day INM, Dedoussis GV, Dei M, Deloukas P, Dermitzakis ET, Dimas AS, Dimitriou M, Dixon AL, Dörr M, van Duijn CM, Ebrahim S, Edkins S, Eiriksdottir G, Eisinger K, Eklund N, Elliott P, Erbel R, Erdmann J, Erdos MR, Eriksson JG, Esko T, Estrada K, Evans DM, de Faire U, Fall T, Farrall M, Feitosa MF, Ferrario MM, Ferreira T, Ferrières J, Fischer K, Fisher E, Fowkes G, Fox CS, Franke L, Franks PW, Fraser RM, Frau F, Frayling T, Freimer NB, Froguel P, Fu M, Gaget S, Ganna A, Gejman PV, Gentilini D, Geus EJC, Gieger C, Gigante B, Gjesing AP, Glazer NL, Goddard ME, Goel A, Grallert H, Gräßler J, Grönberg H, Groop LC, Groves CJ, Gudnason V, Guiducci C, Gustafsson S, Gyllensten U, Hall AS, Hall P, Hallmans G, Hamsten A, Hansen T, Haritunians T, Harris TB, van der Harst P, Hartikainen A-L, Hassanali N, Hattersley AT, Havulinna AS, Hayward C, Heard-Costa NL, Heath AC, Hebebrand J, Heid IM, den Heijer M, Hengstenberg C, Herzig K-H, Hicks AA, Hingorani A, Hinney A, Hirschhorn JN, Hofman A, Holmes CC, Homuth G, Hottenga J-J, Hovingh KG, Hu FB, Hu Y-J, Huffman JE, Hui J, Huikuri H, Humphries SE, Hung J, Hunt SE, Hunter D, Hveem K, Hyppönen E, Igl W, Illig T, Ingelsson E, Iribarren C, Isomaa B, Jackson AU, Jacobs KB, James AL, Jansson J-O, Jarick I, Jarvelin M-R, Jöckel K-H, Johansson Å, Johnson T, Jolley J, Jørgensen T, Jousilahti P, Jula A, Justice AE, Kaakinen M, Kähönen M, Kajantie E, Kanoni S, Kao WHL, Kaplan LM, Kaplan RC, Kaprio J, Kapur K, Karpe F, Kathiresan S, Kee F, Keinanen-Kiukaanniemi SM, Ketkar S, Kettunen J, Khaw K-T, Kiemeney LA, Kilpeläinen TO, Kinnunen L, Kivimaki M, Kivmaki M, Van der Klauw MM, Kleber ME, Knowles JW, Koenig W, Kolcic I, Kolovou G, König IR, Koskinen S, Kovacs P, Kraft P, Kraja AT, Kristiansson K, Krjutškov K, Kroemer HK, Krohn JP, Krzelj V, Kuh D, Kulzer JR, Kumari M, Kutalik Z, Kuulasmaa K, Kuusisto J, Kvaloy K, Laakso M, Laitinen JH, Lakka TA, Lamina C, Langenberg C, Lantieri O, Lathrop GM, Launer LJ, Lawlor DA, Lawrence RW, Leach IM, Lecoeur C, Lee SH, Lehtimäki T, Leitzmann MF, Lettre G, Levinson DF, Li G, Li S, Liang L, Lin D-Y, Lind L, Lindgren CM, Lindström J, Liu J, Liuzzi A, Locke AE, Lokki M-L, Loley C, Loos RJF, Lorentzon M, Luan Ja, Luben RN, Ludwig B, Madden PA, Mägi R, Magnusson PKE, Mangino M, Manunta P, Marek D, Marre M, Martin NG, März W, Maschio A, Mathieson I, McArdle WL, McCaroll SA, McCarthy A, McCarthy MI, McKnight B, Medina-Gomez C, Medland SE, Meitinger T, Metspalu A, van Meurs JBJ, Meyre D, Midthjell K, Mihailov E, Milani L, Min JL, Moebus S, Moffatt MF, Mohlke KL, Molony C, Monda KL, Montgomery GW, Mooser V, Morken MA, Morris AD, Morris AP, Mìhleisen TW, Mìller-Nurasyid M, Munroe PB, Musk AW, Narisu N, Navis G, Neale BM, Nelis M, Nemesh J, Neville MJ, Ngwa JS, Nicholson G, Nieminen MS, Njølstad I, Nohr EA, Nolte IM, North KE, Nöthen MM, Nyholt DR, O'Connell JR, Ohlsson C, Oldehinkel AJ, van Ommen G-J, Ong KK, Oostra BA, Ouwehand WH, Palmer CNA, Palmer LJ, Palotie A, Paré G, Parker AN, Paternoster L, Pawitan Y, Pechlivanis S, Peden JF, Pedersen NL, Pedersen O, Pellikka N, Peltonen L, Penninx B, Perola M, Perry JRB, Person T, Peters A, Peters MJ, Pichler I, Pietiläinen KH, Platou CGP, Polasek O, Pouta A, Power C, Pramstaller PP, Preuss M, Price JF, Prokopenko I, Province MA, Psaty BM, Purcell S, Pìtter C, Qi L, Quertermous T, Radhakrishnan A, Raitakari O, Randall JC, Rauramaa R, Rayner NW,

Rehnberg E, Rendon A, Ridderstråle M, Ridker PM, Ripatti S, Rissanen A, Rivadeneira F, Rivolta C, Robertson NR, Rose LM, Rudan I, Saaristo TE, Sager H, Salomaa V, Samani NJ, Sambrook JG, Sanders AR, Sandholt C, Sanna S, Saramies J, Schadt EE, Scherag A, Schipf S, Schlessinger D, Schreiber S, Schunkert H, Schwarz PEH, Scott LJ, Shi J, Shin S-Y, Shuldiner AR, Shungin D, Signorini S, Silander K, Sinisalo J, Skrobek B, Smit JH, Smith AV, Smith GD, Snieder H, Soranzo N, Sørensen TIA, Sovio U, Spector TD, Speliotes EK, Stančáková A, Stark K, Stefansson K, Steinthorsdottir V, Stephens JC, Stirrups K, Stolk RP, Strachan DP, Strawbridge RJ, Stringham HM, Stumvoll M, Surakka I, Swift AJ, Syvanen A-C, Tammesoo M-L, Teder-Laving M, Teslovich TM, Teumer A, Theodoraki EV, Thomson B, Thorand B, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tönjes A, Tregouet D-A, Tremoli E, Trip MD, Tuomi T, Tuomilehto J, Tyrer J, Uda M, Uitterlinden AG, Usala G, Uusitupa M, Valle TT, Vandenput L, Vatin V, Vedantam S, de Vegt F, Vermeulen SH, Viikari J, Virtamo J, Visscher PM, Vitart V, Van Vliet-Ostaptchouk JV, Voight BF, Vollenweider P, Volpato CB, Völzke H, Waeber G, Waite LL, Wallaschofski H, Walters GB, Wang Z, Wareham NJ, Watanabe RM, Watkins H, Weedon MN, Welch R, Weyant RJ, Wheeler E, White CC, Wichmann HE, Widen E, Wild SH, Willemsen G, Willer CJ, Wilsgaard T, Wilson JF, van Wingerden S, Winkelmann BR, Winkler TW, Witte DR, Witteman JCM, Wolffenbuttel BHR, Wong A, Wood AR, Workalemahu T, Wright AF, Yang J, Yarnell JWG, Zgaga L, Zhao JH, Zillikens MC, Zitting P, Zondervan KT, Medland SE, Martin NG, Magnusson PKE, Iacono WG, McGue M, North KE, Yang J, Visscher PM (2017). Genetic evidence of assortative mating in humans. *Nature Human Behaviour* **1**:0016.

Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E (2010). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genetics, selection, evolution : GSE* **42**(1):8-8.

Rose K, Harper R, Tromans C, Waterman C, Goldberg D, Haggerty C, Tullo A (2000). Quality of life in myopia. *The British journal of ophthalmology* **84**(9):1031-1034.

Rosner M, Belkin M (1987). Intelligence, Education, and Myopia in Males. *JAMA Ophthalmology* **105**(11):1508-1511.

Rothman KJ (1974). Synergy and antagonism in cause-effect relationships. *Am J Epidemiol* **99**(6):385-388.

Rothman KJ (1976). The estimation of synergy or antagonism. *Am J Epidemiol* **103**(5):506-511.

Rudnicka AR, Kapetanakis VV, Wathern AK, Logan NS, Gilmartin B, Whincup PH, Cook DG, Owen CG (2016). Global variations and time trends in the prevalence of childhood myopia, a systematic review and quantitative meta-analysis: implications for aetiology and early prevention. *The British journal of ophthalmology* **100**(7):882-890.

Sa J, Liu X, He T, Liu G, Cui Y (2016). A Nonlinear Model for Gene-Based Gene-Environment Interaction. *International journal of molecular sciences* **17**(6):882.

Saw S-M, Gazzard G, Shih-Yen EC, Chua W-H (2005). Myopia and associated pathological complications. *Ophthalmic and Physiological Optics* **25**(5):381-391.

Saw S-M, Katz J, Schein O, Chew SJ, Chan TK (1996). Epidemiology of Myopia. *Epidemiologic reviews* **18**:175-187.

Saw S-M, Nieto FJ, Katz J, Schein OD, Levy B, Chew S-J (2000). Factors Related to the Progression of Myopia in Singaporean Children. *Optometry and Vision Science* **77**(10):549-554.

Schwartz M, Haim M, Skarsholm D (1990). X-linked myopia: Bornholm Eye Disease. *Clinical Genetics* **38**(4):281-286.

Schwarz DF, Konig IR, Ziegler A (2010). On safari to Random Jungle: a fast implementation of Random Forests for high-dimensional data. *Bioinformatics* **26**(14):1752-1758.

Seitz C, Huang J, Geiselhöringer A-L, Galbani-Bianchi P, Michalek S, Phan TS, Reinhold C, Dietrich L, Schmidt C, Corazza N, Delgado ME, Schnalzger T, Schoonjans K, Brunner T (2019). The orphan nuclear receptor LRH-1/NR5a2 critically regulates T cell functions. *Science Advances* **5**(7):eaav9732.

Shah RL, Huang Y, Guggenheim JA, Williams C (2017). Time Outdoors at Specific Ages During Early Childhood and the Risk of Incident MyopiaEarly Time Outdoors and Risk of Incident Myopia. *Investigative Ophthalmology & Visual Science* **58**(2):1158-1166.

Shen X, Pettersson M, Rönnegård L, Carlborg Ö (2012). Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. *PLoS genetics* **8**(8):e1002839-e1002839.

Si JK, Tang K, Bi HS, Guo DD, Guo JG, Wang XR (2015). Orthokeratology for myopia control: a meta-analysis. *Optom Vis Sci* **92**(3):252-257.

Sitlani CM, Dupuis J, Rice KM, Sun F, Pitsillides AN, Cupples LA, Psaty BM (2015). Genome-wide gene–environment interactions on quantitative traits using family data. *European Journal Of Human Genetics* **24**:1022.

Slatkin M (2009). Epigenetic inheritance and the missing heritability problem. *Genetics* **182**(3):845-850.

Snead MP, Yates JR (1999). Clinical and Molecular genetics of Stickler syndrome. *Journal of medical genetics* **36**(5):353-359.

Sorsby (1962). Refraction of the Eye and its Components in Twins. *Nature* **195**(4843):758-759.

Spiller W, Slichter D, Bowden J, Davey Smith G (2019). Detecting and correcting for bias in Mendelian randomization analyses using Gene-by-Environment interactions. *Int J Epidemiol* **48**(3):702-712.

Steen KV (2012). Travelling the world of gene-gene interactions. *Brief Bioinform* **13**(1):1-19.

Stenzel SL, Ahn J, Boonstra PS, Gruber SB, Mukherjee B (2015). The impact of exposure-biased sampling designs on detection of gene-environment interactions in case-control studies with potential exposure misclassification. *Eur J Epidemiol* **30**(5):413-423.

Struchalin MV, Dehghan A, Witteman JC, van Duijn C, Aulchenko YS (2010). Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. *BMC genetics* **11**:92-92.

Su M-W, Tung K-Y, Liang P-H, Tsai C-H, Kuo N-W, Lee YL (2012). Gene-gene and gene-environmental interactions of childhood asthma: a multifactor dimension reduction approach. *PloS one* **7**(2):e30694-e30694.

Su Y-R, Di C-Z, Hsu L, Genetics, Epidemiology of Colorectal Cancer C (2016). A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics* **18**(1):119-131.

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (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**(3):e1001779-e1001779.

Sun R, Carroll RJ, Christiani DC, Lin X (2018). Testing for gene-environment interaction under exposure misspecification. *Biometrics* **74**(2):653-662.

Sun X, Elston R, Morris N, Zhu X (2013). What Is the Significance of Difference in Phenotypic Variability across SNP Genotypes? *American Journal of Human Genetics* **93**(2):390-397.

Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* **20**(8):467-484.

Tang W, Wu X, Jiang R, Li Y (2009). Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. *PLoS genetics* **5**(5):e1000464-e1000464.

Tano Y (2002). Pathologic myopia: where are we now? *American journal of ophthalmology* **134**:645-660.

Tay MT, Au Eong KG, Ng CY, Lim MK (1992). Myopia and educational attainment in 421,116 young Singaporean males. *Ann Acad Med Singapore* **21**(6):785-791.

Tchetgen Tchetgen EJ, Kraft P (2011). On the Robustness of Tests of Genetic Associations Incorporating Gene-environment Interaction When the Environmental Exposure is Misspecified. *Epidemiology* **22**(2):257-261.

Teasdale TW, Goldschmidt E (1988). Myopia and its relationship to education, intelligence and height. Preliminary results from an on-going study of Danish draftees. *Acta Ophthalmol Suppl* **185**:41-43.

Tedja MS, Wojciechowski R, Hysi PG, Eriksson N, Furlotte NA, Verhoeven VJM, Iglesias AI, Meester-Smoor MA, Tompson SW, Fan Q, Khawaja AP, Cheng C-Y, Höhn R, Yamashiro K, Wenocur A, Grazal C, Haller T, Metspalu A, Wedenoja J, Jonas JB, Wang YX, Xie J, Mitchell P, Foster PJ, Klein BEK, Klein R, Paterson AD, Hosseini SM, Shah RL, Williams C, Teo YY, Tham YC, Gupta P, Zhao W, Shi Y, Saw W-Y, Tai ES, Sim XL, Huffman JE, Polašek O, Hayward C, Bencic G, Rudan I, Wilson JF, Aung T, Veluchamy AB, Burdon KP, Campbell H, Chen LJ, Chen P, Chen W, Chew E, Deangelis MM, Ding X, Döring A, Evans DM, Feng S, Fleck B, Fogarty RD, Fondran JR, Fossarello M, Guo X, Haarman AEG, He M, Howe LD, Janmahasatian S, Jhanji V, Kähönen M, Kaprio J, Kemp JP, Khaw K-T, Khor C-C, Krapohl E, Korobelnik J-F, Lee K, Li S-M, Lu Y, Luben RN, Mäkelä K-M, McMahon G, Meguro A, Mihailov E, Miyake M, Mizuki N, Morrison M, Nangia V, Oexle K, Panda-Jonas S, Pang CP, Pirastu M, Plomin R, Rantanen T, Schache M, Seppälä I, Smith GD, Pourcain BS, Tam PO, Tideman JWL, Timpson NJ, Vaccargiu S, Vatavuk Z, Wang JJ, Wang N, Wareham NJ, Wright AF, Xu L, Yap MKH, Yazar S, Yip SP, Yoshimura N, Young AL, Zhao JH, Zhou X, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Hinds DA, McCreight JC, Huber KE, Kleinman A, Litterman NK, McIntyre MH, Mountain JL, Noblin ES, Northover CAM, Pitts SJ, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Vacic V, Wilson CH, Aslam TM, Barman SA, Barrett JH, Bishop PN, Blows P, Bunce C, Carare RO, Chakravarthy U, Chan M, Chua S, Crabb D, Day A, Desai P, Dhillon B, Dick AD, Egan CA, Ennis S, Fruttiger M, Gallacher J, Garway-Heath DF, Gibson J, Gore DM, Hardcastle A, Harding SP, Hogg RE, Keane PA, Khaw PT, Lascaratos G, Lotery A, Luthert PJ, MacGillivray TJ, Mackie SL, Martin KR, McGaughey M, McGuinness B, McKay GJ, McKibbin M, Mitry D, Moore T, Morgan JE, Muthy ZA, O'Sullivan E, Owen C, Patel PJ, Paterson EN, Peto T, Petzold A, Rudnicka AR, Self JE, Sivaprasad S, Steel DHW, Stratton IM, Strouthidis N, Sudlow CLM, Thaung C, Thomas D, Trucco E, Tufail A, Vernon SA, Viswanathan AC, Woodside JV, Yates M, Yip JLY, Zheng Y, Joshi PK, Tsujikawa A, Matsuda F, Whisenhunt KN, Zeller T, van der Spek PJ, Haak R, Meijers-Heijboer H, van Leeuwen EM, Iyengar SK, Lass JH, Hofman A, Rivadeneira F, Uitterlinden AG, Vingerling JR, Lehtimäki T, Raitakari OT, Biino G, Concas MP, Schwantes-An T-H, Igo RP, Cuellar-Partida G, Martin NG, Craig JE, Gharahkhani P, Williams KM, Nag A, Rahi JS, Cumberland PM, Delcourt C, Bellenguez C, Ried JS, Bergen AA, Meitinger T, Gieger C, Wong TY, Hewitt AW, Mackey DA, Simpson CL, Pfeiffer N, Pärssinen O, Baird PN, Vitart V, Amin N, van Duijn CM, Bailey-Wilson

JE, Young TL, Saw S-M, Stambolian D, MacGregor S, Guggenheim JA, Tung JY, Hammond CJ, Klaver CCW, The CC, andMe Research T, Eye UKB, Vision C (2018). Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nature Genetics* **50**(6):834-848.

the Haplotype Reference C, McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki A-E, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PIW, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R, Abecasis G, Marchini J (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**:1279.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**(9):2129-2141.

Tkatchenko AV, Tkatchenko TV, Guggenheim JA, Verhoeven VJ, Hysi PG, Wojciechowski R, Singh PK, Kumar A, Thinakaran G, Williams C (2015). *APLP2* Regulates Refractive Error and Myopia Development in Mice and Humans. *PLoS Genetics* **11**(8):e1005432.

Troilo D, Nickla DL, Mertz JR, Summers Rada JA (2006). Change in the synthesis rates of ocular retinoic acid and scleral glycosaminoglycan during experimentally altered eye growth in marmosets. *Invest Ophthalmol Vis Sci* **47**(5):1768-1777.

Tsai C-T, Lai L-P, Lin J-L, Chiang F-T, Hwang J-J, Ritchie Marylyn D, Moore Jason H, Hsu K-L, Tseng C-D, Liau C-S, Tseng Y-Z (2004). Renin-Angiotensin System Gene Polymorphisms and Atrial Fibrillation. *Circulation* **109**(13):1640-1646.

Tucci V, Kleefstra T, Hardy A, Heise I, Maggi S, Willemsen MH, Hilton H, Esapa C, Simon M, Buenavista M-T, McGuffin LJ, Vizor L, Dodero L, Tsaftaris S, Romero R, Nillesen WN, Vissers LELM, Kempers MJ, Vulto-van Silfhout AT, Iqbal Z, Orlando M, Maccione A, Lassi G, Farisello P, Contestabile A, Tinarelli F, Nieus T, Raimondi A, Greco B, Cantatore D, Gasparini L, Berdondini L, Bifone A, Gozzi A, Wells S, Nolan PM (2014). Dominant β-catenin mutations cause intellectual disability with recognizable syndromic features. *The Journal of Clinical Investigation* **124**(4):1468-1482.

Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, Agee M, Alipanahi B, Auton A, Bell RK, Bryc K, Elson SL, Fontanillas P, Furlotte NA, Hinds DA, Hromatka BS, Huber KE, Kleinman A, Litterman NK, McIntyre MH, Mountain JL, Northover CAM, Sathirapongsasuti JF, Sazonova OV, Shelton JF, Shringarpure S, Tian C, Tung JY, Vacic V, Wilson CH, Pitts SJ, Magnusson P, Oskarsson S, Johannesson M, Visscher PM, Laibson D, Cesarini D, Neale BM, Benjamin DJ, andMe Research T, Social Science Genetic Association C (2018). Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics* **50**(2):229-237.

Tyler AL, Asselbergs FW, Williams SM, Moore JH (2009). Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays* **31**(2):220-227.

Tzeng J-Y, Zhang D, Pongpanich M, Smith C, McCarthy MI, Sale MM, Worrall BB, Hsu F-C, Thomas DC, Sullivan PF (2011). Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *American journal of human genetics* **89**(2):277-288.

Ueki M, Fujii M, Tamiya G, for Alzheimer's Disease Neuroimaging I, the Alzheimer's Disease Metabolomics C (2019). Quick assessment for systematic test statistic inflation/deflation due to null model misspecifications in genome-wide environment interaction studies. *PLOS ONE* **14**(7):e0219825.

Umbach DM, Weinberg CR (1997). Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* **16**(15):1731-1743.

VanderWeele TJ (2013). Reconsidering the denominator of the attributable proportion for interaction. *European journal of epidemiology* **28**(10):779-784.

Verhoeven VJ, Hysi PG, Wojciechowski R, Fan Q, Guggenheim JA, Hohn R, MacGregor S, Hewitt AW, Nag A, Cheng CY, Yonova-Doing E, Zhou X, Ikram MK, Buitendijk GH, McMahon G, Kemp JP, Pourcain BS, Simpson CL, Makela KM, Lehtimaki T, Kahonen M, Paterson AD, Hosseini SM, Wong HS, Xu L, Jonas JB, Parssinen O, Wedenoja J, Yip SP, Ho DW, Pang CP, Chen LJ, Burdon KP, Craig JE, Klein BE, Klein R, Haller T, Metspalu A, Khor CC, Tai ES, Aung T, Vithana E, Tay WT, Barathi VA, Chen P, Li R, Liao J, Zheng Y, Ong RT, Doring A, Evans DM, Timpson NJ, Verkerk AJ, Meitinger T, Raitakari O, Hawthorne F, Spector TD, Karssen LC, Pirastu M, Murgia F, Ang W, Mishra A, Montgomery GW, Pennell CE, Cumberland PM, Cotlarciuc I, Mitchell P, Wang JJ, Schache M, Janmahasatian S, Igo RP, Jr., Lass JH, Chew E, Iyengar SK, Gorgels TG, Rudan I, Hayward C, Wright AF, Polasek O, Vatavuk Z, Wilson JF, Fleck B, Zeller T, Mirshahi A, Muller C, Uitterlinden AG, Rivadeneira F, Vingerling JR, Hofman A, Oostra BA, Amin N, Bergen AA, Teo YY, Rahi JS, Vitart V, Williams C, Baird PN, Wong TY, Oexle K, Pfeiffer N, Mackey DA, Young TL, van Duijn CM, Saw SM, Bailey-Wilson JE, Stambolian D, Klaver CC, Hammond CJ (2013). Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat Genet* **45**(3):314-318.

Verkicharla PK, Ramamurthy D, Nguyen QD, Zhang X, Pu S-H, Malhotra R, Ostbye T, Lamoureux EL, Saw S-M (2017). Development of the FitSight Fitness Tracker to Increase Time Outdoors to Prevent Myopia. *Translational Vision Science & Technology* **6**(3):20-20.

Verstraeten A, Alaerts M, Van Laer L, Loeys B (2016). Marfan Syndrome and Related Disorders: 25 Years of Gene Discovery. *Human mutation* **37**.

Vessey KA, Lencses KA, Rushforth DA, Hruby VJ, Stell WK, (2005) Glucagon receptor agonists and antagonists affect the growth of the chick eye: a role for glucagonergic regulation of emmetropization? Investigative ophthalmology & visual science 46(11)3922-3931.

Viechtbauer W (2010). Conducting Meta-Analyses in R with the metafor package. *Journal of Statistical Software* **36**(3):1-48.

Vilhjálmsson Bjarni J, Yang J, Finucane Hilary K, Gusev A, Lindström S, Ripke S, Genovese G, Loh P-R, Bhatia G, Do R, Hayeck T, Won H-H, Ripke S, Neale Benjamin M, Corvin A, Walters James TR, Farh K-H, Holmans Peter A, Lee P, Bulik-Sullivan B, Collier David A, Huang H, Pers Tune H, Agartz I, Agerbo E, Albus M, Alexander M, Amin F, Bacanu Silviu A, Begemann M, Belliveau Richard A, Jr., Bene J, Bergen Sarah E, Bevilacqua E, Bigdeli Tim B, Black Donald W, Bruggeman R, Buccola Nancy G, Buckner Randy L, Byerley W, Cahn W, Cai G, Campion D, Cantor Rita M, Carr Vaughan J, Carrera N, Catts Stanley V, Chambert Kimberly D, Chan Raymond CK, Chen Ronald YL, Chen Eric YH, Cheng W, Cheung Eric FC, Chong Siow A, Cloninger CR, Cohen D, Cohen N, Cormican P, Craddock N, Crowley James J, Curtis D, Davidson M, Davis Kenneth L, Degenhardt F, Del Favero J, DeLisi Lynn E, Demontis D, Dikeos D, Dinan T, Djurovic S, Donohoe G, Drapeau E, Duan J, Dudbridge F, Durmishi N, Eichhammer P, Eriksson J, Escott-

Price V, Essioux L, Fanous Ayman H, Farrell Martilias S, Frank J, Franke L, Freedman R, Freimer Nelson B, Friedl M, Friedman Joseph I, Fromer M, Genovese G, Georgieva L, Gershon Elliot S, Giegling I, Giusti-Rodrguez P, Godard S, Goldstein Jacqueline I, Golimbet V, Gopal S, Gratten J, Grove J, de Haan L, Hammer C, Hamshere Marian L, Hansen M, Hansen T, Haroutunian V, Hartmann Annette M, Henskens Frans A, Herms S, Hirschhorn Joel N, Hoffmann P, Hofman A, Hollegaard Mads V, Hougaard David M, Ikeda M, Joa I, Julia A, Kahn Rene S, Kalaydjieva L, Karachanak-Yankova S, Karjalainen J, Kavanagh D, Keller Matthew C, Kelly Brian J, Kennedy James L, Khrunin A, Kim Y, Klovins J, Knowles James A, Konte B, Kucinskas V, Kucinskiene Zita A, Kuzelova-Ptackova H, Kahler Anna K, Laurent C, Keong Jimmy Lee C, Lee SH, Legge Sophie E, Lerer B, Li M, Li T, Liang K-Y, Lieberman J, Limborska S, Loughland Carmel M, Lubinski J, Lnnqvist J, Macek M, Jr., Magnusson Patrik KE, Maher Brion S, Maier W, Mallet J, Marsal S, Mattheisen M, Mattingsdal M, McCarley Robert W, McDonald C, McIntosh Andrew M, Meier S, Meijer Carin J, Melegh B, Melle I, Mesholam-Gately Raquelle I, Metspalu A, Michie Patricia T, Milani L, Milanova V, Mokrab Y, Morris Derek W, Mors O, Mortensen Preben B, Murphy Kieran C, Murray Robin M, Myin-Germeys I, Mller-Myhsok B, Nelis M, Nenadic I, Nertney Deborah A, Nestadt G, Nicodemus Kristin K, Nikitina-Zake L, Nisenbaum L, Nordin A, O'Callaghan E, O'Dushlaine C, O'Neill FA, Oh S-Y, Olincy A, Olsen L, Van Os J, Pantelis C, Papadimitriou George N, Papiol S, Parkhomenko E, Pato Michele T, Paunio T, Pejovic-Milovancevic M, Perkins Diana O, Pietilinen O, Pimm J, Pocklington Andrew J, Powell J, Price A, Pulver Ann E, Purcell Shaun M, Quested D, Rasmussen Henrik B, Reichenberg A, Reimers Mark A, Richards Alexander L, Roffman Joshua L, Roussos P, Ruderfer Douglas M, Salomaa V, Sanders Alan R, Schall U, Schubert Christian R, Schulze Thomas G, Schwab Sibylle G, Scolnick Edward M, Scott Rodney J, Seidman Larry J, Shi J, Sigurdsson E, Silagadze T, Silverman Jeremy M, Sim K, Slominsky P, Smoller Jordan W, So H-C, Spencer Chris CA, Stahl Eli A, Stefansson H, Steinberg S, Stogmann E, Straub Richard E, Strengman E, Strohmaier J, Stroup TS, Subramaniam M, Suvisaari J, Svrakic Dragan M, Szatkiewicz Jin P, Sderman E, Thirumalai S, Toncheva D, Tooney Paul A, Tosato S, Veijola J, Waddington J, Walsh D, Wang D, Wang Q, Webb Bradley T, Weiser M, Wildenauer Dieter B, Williams Nigel M, Williams S, Witt Stephanie H, Wolen Aaron R, Wong Emily HM, Wormley Brandon K, Wu Jing Q, Xi Hualin S, Zai Clement C, Zheng X, Zimprich F, Wray Naomi R, Stefansson K, Visscher Peter M, Adolfsson R, Andreassen Ole A, Blackwood Douglas HR, Bramon E, Buxbaum Joseph D, Børglum Anders D, Cichon S, Darvasi A, Domenici E, Ehrenreich H, Esko T, Gejman Pablo V, Gill M, Gurling H, Hultman Christina M, Iwata N, Jablensky Assen V, Jonsson Erik G, Kendler Kenneth S, Kirov G, Knight J, Lencz T, Levinson Douglas F, Li Qingqin S, Liu J, Malhotra Anil K, McCarroll Steven A, McQuillin A, Moran Jennifer L, Mortensen Preben B, Mowry Bryan J, Nthen Markus M, Ophoff Roel A, Owen Michael J, Palotie A, Pato Carlos N, Petryshen Tracey L, Posthuma D, Rietschel M, Riley Brien P, Rujescu D, Sham Pak C, Sklar P, St. Clair D, Weinberger Daniel R, Wendland Jens R, Werge T, Daly Mark J, Sullivan Patrick F, O'Donovan Michael C, Kraft P, Hunter DJ, Adank M, Ahsan H, Aittomäki K, Baglietto L, Berndt S, Blomquist C, Canzian F, Chang-Claude J, Chanock SJ, Crisponi L, Czene K, Dahmen N, Silva IdS, Easton D, Eliassegn AH, Figueroa J, Fletcher O, Garcia-Closas M, Gaudet MM, Gibson L, Haiman CA, Hall P, Hazra A, Hein R, Henderson BE, Hofman A, Hopper JL, Irwanto A, Johansson M, Kaaks R, Kibriya MG, Lichtner P, Lindström S, Liu J, Lund E, Makalic E, Meindl A, Meijers-Heijboer H, Müller-Myhsok B, Muranen TA, Nevanlinna H, Peeters PH, Peto J, Prentice RL, Rahman N, Sánchez MJ, Schmidt DF, Schmutzler RK, Southey MC, Tamimi R, Travis R, Turnbull C, Uitterlinden AG, van der Luijt RB, Waisfisz Q, Wang Z, Whittemore AS, Yang R, Zheng W, Kathiresan S, Pato M, Pato C, Tamimi R, Stahl E, Zaitlen N, Pasaniuc B, Belbin G, Kenny EE, Schierup MH, De Jager P, Patsopoulos NA, McCarroll S, Daly M, Purcell S, Chasman D, Neale B, Goddard M, Visscher PM, Kraft P, Patterson N, Price AL (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics* **97**(4):576-592.

Visscher PM, Brown MA, McCarthy MI, Yang J (2012). Five years of GWAS discovery. *American journal of human genetics* **90**(1):7-24.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* **101**(1):5-22.

Vongphanit J, Mitchell P, Wang JJ (2002). Population prevalence of tilted optic disks and the relationship of this sign to refractive error. *Am J Ophthalmol* **133**(5):679-685.

Waddington CH (1942). CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS. *Nature* **150**(3811):563-565.

Wainschtein P, Jain DP, Yengo L, Zheng Z, Cupples LA, Shadyab AH, McKnight B, Shoemaker BM, Mitchell BD, Psaty BM, Kooperberg C, Roden D, Darbar D, Arnett DK, Regan EA, Boerwinkle E, Rotter JI, Allison MA, McDonald M-LN, Chung MK, Smith NL, Ellinor PT, Vasan RS, Mathias RA, Rich SS, Heckbert SR, Redline S, Guo X, Chen YDI, Liu C-T, de Andrade M, Yanek LR, Albert CM, Hernandez RD, McGarvey ST, North KE, Lange LA, Weir BS, Laurie CC, Yang J, Visscher PM (2019). Recovery of trait heritability from whole genome sequence data. *bioRxiv*:588020.

Wan L, Deng B, Wu Z, Chen X. 2018. Exome sequencing study of 20 patients with high myopia. *PeerJ* [Online] 6

Wang H, Zhang F, Zeng J, Wu Y, Kemper KE, Xue A, Zhang M, Powell JE, Goddard ME, Wray NR, Visscher PM, McRae AF, Yang J (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Science Advances* **5**(8):eaaw3538.

Wang J, Duncan D, Shi Z, Zhang B (2013). WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Research* **41**(W1):W77-W83.

Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017). FUMA: Functional mapping and annotation of genetic associations. *bioRxiv*.

Wei W-H, Hemani G, Haley CS (2014). Detecting epistasis in human complex traits. *Nature Reviews Genetics* **15**:722.

Willer CJ, Li Y, Abecasis GR (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**(17):2190-2191.

Williams KM, Hammond CJ (2014). Prevalence of myopia and association with education in Europe. *The Lancet* **383**:S109.

Williams KM, Hysi PG, Nag A, Yonova-Doing E, Venturini C, Hammond CJ (2013). Age of myopia onset in a British population-based twin cohort. *Ophthalmic and Physiological Optics* **33**(3):339-345.

Williams PT (2012). Quantile-Specific Penetrance of Genes Affecting Lipoproteins, Adiposity and Height. *PLoS ONE* **7**(1):e28764.

Winham SJ, Slater AJ, Motsinger-Reif AA (2010). A comparison of internal validation techniques for multifactor dimensionality reduction. *BMC bioinformatics* **11**:394-394.

Wirapati P, Forner K, Delgado-Vega A, Alarcón-Riquelme M, Delorenzi M, Wojcik J (2011). Detecting Epistasis with Restricted Response Patterns in Pairs of Biallelic Loci. *Annals of Human Genetics* **75**(1):133-145.

Wojciechowski R (2011). Nature and nurture: the complex genetics of myopia and refractive error. *Clinical Genetics* **79**(4):301-320.

Wojciechowski R, Hysi PG (2013). Focusing in on the complex genetics of myopia. *PLoS genetics* **9**(4):e1003442-e1003442.

Wolf BJ, Hill EG, Slate EH (2010). Logic Forest: an ensemble classifier for discovering logical combinations of binary markers. *Bioinformatics* **26**(17):2183-2189.

Wolffsohn JS, Calossi A, Cho P, Gifford K, Jones L, Li M, Lipener C, Logan NS, Malet F, Matos S, Meijome JMG, Nichols JJ, Orr JB, Santodomingo-Rubido J, Schaefer T, Thite N, van der Worp E, Zvirgzdina M (2016). Global trends in myopia management attitudes and strategies in clinical practice. *Contact Lens and Anterior Eye* **39**(2):106-116.

Wong MY, Day NE, Luan JA, Chan KP, Wareham NJ (2003). The detection of gene–environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *International Journal of Epidemiology* **32**(1):51-57.

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Ja, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Magi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Leach IM, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancakova A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlov J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Bluher M, Bolton JL, Bottcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dorr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Graszler J, Gronberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson A, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindstrom J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Muller G, Muller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nothen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundstrom J, Swertz MA, Syvanen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmann S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimaki T, Lupoli S, Madden PAF, Mannisto S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tonjes

A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, The Electronic Medical R, Genomics C, The MC, The PC, The LifeLines Cohort S, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Volzke H, Walker M, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46**(11):1173-1186.

Wray NR (2005). Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res Hum Genet* **8**(2):87-94.

Wu C, Cui Y (2013). A novel method for identifying nonlinear gene-environment interactions in case-control association studies. *Human genetics* **132**.

Wu C, Cui Y, Ma S (2014). Integrative analysis of gene-environment interactions under a multi-response partially linear varying coefficient model. *Statistics in medicine* **33**(28):4988-4998.

Wu P-C, Chuang M-N, Choi J, Chen H, Wu G, Ohno-Matsui K, Jonas JB, Cheung CMG (2019). Update in myopia and treatment strategy of atropine use in myopia control. *Eye* **33**(1):3-13.

Wu PC, Tsai CL, Wu HL, Yang YH, Kuo HK (2013). Outdoor activity during class recess reduces myopia onset and progression in school children. *Ophthalmology* **120**(5):1080-1085.

Xu H-M, Sun X-W, Qi T, Lin W-Y, Liu N, Lou X-Y (2014). Multivariate dimensionality reduction approaches to identify gene-gene and gene-environment interactions underlying multiple complex traits. *PloS one* **9**(9):e108103-e108103.

Yang C-H, Chuang L-Y, Lin Y-D (2017). Multiobjective differential evolution-based multifactor dimensionality reduction for detecting gene–gene interactions. *Scientific Reports* **7**(1):12869.

Yang C-H, Yang H-S, Chuang L-Y (2019). PBMDR: A particle swarm optimization-based multifactor dimensionality reduction for the detection of multilocus interactions. *Journal of Theoretical Biology* **461**:68-75.

Yang CH, Huang CC, Hsu KS (2012a). A critical role for protein tyrosine phosphatase nonreceptor type 5 in determining individual susceptibility to develop stress-related cognitive and morphological changes. *J Neurosci* **32**(22):7550-7562.

Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Mägi R, Waite L, Vernon Smith A, Yerges-Armstrong LM, Monda KL, Hadley D, Mahajan A, Li G, Kapur K, Vitart V, Huffman JE, Wang SR, Palmer C, Esko T, Fischer K, Hua Zhao J, Demirkan A, Isaacs A, Feitosa MF, Luan Ja, Heard-Costa NL, White C, Jackson AU, Preuss M, Ziegler A, Eriksson J, Kutalik Z, Frau F, Nolte IM, Van Vliet-Ostaptchouk JV, Hottenga J-J, Jacobs KB, Verweij N, Goel A, Medina-Gomez C, Estrada K, Lynn Bragg-Gresham J, Sanna S, Sidore C, Tyrer J, Teumer A, Prokopenko I, Mangino M, Lindgren CM, Assimes TL, Shuldiner AR, Hui J, Beilby JP, McArdle WL, Hall P, Haritunians T, Zgaga L, Kolcic I, Polasek O, Zemunik T,

Oostra BA, Juhani Junttila M, Grönberg H, Schreiber S, Peters A, Hicks AA, Stephens J, Foad NS, Laitinen J, Pouta A, Kaakinen M, Willemsen G, Vink JM, Wild SH, Navis G, Asselbergs FW, Homuth G, John U, Iribarren C, Harris T, Launer L, Gudnason V, O'Connell JR, Boerwinkle E, Cadby G, Palmer LJ, James AL, Musk AW, Ingelsson E, Psaty BM, Beckmann JS, Waeber G, Vollenweider P, Hayward C, Wright AF, Rudan I, Groop LC, Metspalu A, Khaw K-T, van Duijn CM, Borecki IB, Province MA, Wareham NJ, Tardif J-C, Huikuri HV, Adrienne Cupples L, Atwood LD, Fox CS, Boehnke M, Collins FS, Mohlke KL, Erdmann J, Schunkert H, Hengstenberg C, Stark K, Lorentzon M, Ohlsson C, Cusi D, Staessen JA, Van der Klauw MM, Pramstaller PP, Kathiresan S, Jolley JD, Ripatti S, Jarvelin M-R, de Geus EJC, Boomsma DI, Penninx B, Wilson JF, Campbell H, Chanock SJ, van der Harst P, Hamsten A, Watkins H, Hofman A, Witteman JC, Zillikens MC, Uitterlinden AG, Rivadeneira F, Carola Zillikens M, Kiemeney LA, Vermeulen SH, Abecasis GR, Schlessinger D, Schipf S, Stumvoll M, Tönjes A, Spector TD, North KE, Lettre G, McCarthy MI, Berndt SI, Heath AC, Madden PAF, Nyholt DR, Montgomery GW, Martin NG, McKnight B, Strachan DP, Hill WG, Snieder H, Ridker PM, Thorsteinsdottir U, Stefansson K, Frayling TM, Hirschhorn JN, Goddard ME, Visscher PM (2012b). FTO genotype is associated with phenotypic variability of body mass index. *Nature* **490**:267.

Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, the GC (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics* **27**(20):3641-3649.

Young AI, Wauthier F, Donnelly P (2016). Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nature Communications* **7**:12724.

Young AI, Wauthier FL, Donnelly P (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nature Genetics* **50**(11):1608-1614.

Yu L, Li Z-K, Gao J-R, Liu J-R, Xu C-T (2011). Epidemiology, genetics and treatments for myopia. *International journal of ophthalmology* **4**(6):658-669.

Yu W, Lee S, Park T (2016). A unified model based multifactor dimensionality reduction framework for detecting gene–gene interactions. *Bioinformatics* **32**(17):i605-i610.

Zeitz C, Robson AG, Audo I (2015). Congenital stationary night blindness: an analysis and update of genotype-phenotype correlations and pathogenic mechanisms. *Prog Retin Eye Res* **45**:58-110.

Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR (2015). The Ensembl Regulatory Build. *Genome Biology* **16**(1):56.

Zhang F, Chen W, Zhu Z, Zhang Q, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J (2019). OSCA: a tool for omic-data-based complex trait analysis. *bioRxiv*:445163.

Zhang J, Hur YM, Huang W, Ding X, Feng K, He M (2011a). Shared genetic determinants of axial length and height in children: the Guangzhou twin eye study. *Archives of Ophthalmology* **129**(1):63-68.

Zhang P, Lewinger JP, Conti D, Morrison JL, Gauderman WJ (2016). Detecting Gene-Environment Interactions for a Quantitative Trait in a Genome-Wide Association Study. *Genetic Epidemiology* **40**(5):394-403.

Zhang Q, Long Q, Ott J (2014). AprioriGWAS, a new pattern mining strategy for detecting genetic variants associated with disease through interaction effects. *PLoS Comput Biol* **10**(6):e1003627.

Zhang Y, Ding C, Li T (2008). Gene selection algorithm by combining reliefF and mRMR. *BMC genomics* **9 Suppl 2**(Suppl 2):S27-S27.

Zhang Y, Jiang B, Zhu J, Liu J (2011b). Bayesian Models for Detecting Epistatic Interactions from Genetic Data. *Annals of human genetics* **75**:183-193.

Zhang Y, Liu JS (2007). Bayesian inference of epistatic interactions in case-control studies. *Nature Genetics* **39**(9):1167-1173.

Zhao G, Marceau R, Zhang D, Tzeng J-Y (2015). Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. *Genetics* **199**(3):695-710.

Zidkova A, Horinek A, Stenzl V, Korabecna M (2013). Application of multifactor dimensionality reduction analysis and Bayesian networks for eye color and ancestry prediction for forensic purposes in the Czech Republic. *Forensic Science International: Genetics Supplement Series* **4**(1):e322-e323.