Linkage Analysis in Highly Myopic Families

Rosalind Creer

School of Optometry & Vision Sciences
Cardiff University

Thesis submitted for the degree of Doctor of Philosophy

January 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I am extremely grateful to my supervisor, Dr. Jeremy Guggenheim, for his unfailing patience, support, help and advice throughout the past few years. Also thanks to the Department of Psychological Medicine, Cardiff University, particularly Hywel Williams for his expertise and help with the genotyping for both the linkage and association studies. I would like to acknowledge the contribution made by Chris Haley and Ricardo Pong-Wong for the MCMC analysis. I am also grateful to Dr. Jon Erichsen for the use of his computer, without which there would have been very few pages to this thesis.

The completion of this section of the Family Study of Myopia has relied on the participation and co-operation of the subjects involved. For this I would like to express my thanks to Adrian Murphy for the initial introduction to his extended family and the following help and organisation of the trip to Ireland and all that surrounds the investigation. Also to the family, not only for their contribution to the project but for the warm reception we received. Thanks also to Jane and Sean Farbrother and Toni Darby for their help with the data collection and with making sure the family get-together ran smoothly and to Dr. Ian Flitcroft for the provision of his time and the Zeiss IOLMaster.

On that same theme, thanks must go to the current participants of the Family Study of Myopia, to the optometrists championing the cause and specifically to John Welsby and Conlons Opticians for their ongoing support. Also to the high myopes seen at the Eye Clinic at Cardiff University, who were willing to add the Myopia Study to their list of research involvements.

The College of Optometrists have provided the grant allowing me to undertake this project for which I am grateful. I am also grateful for the grant from the Sir Jules Thorn Trust which allowed us to go to Kilkenny to collect the vital data for the analysis. Thanks must also go to the Wellcome Trust for enabling me to attend the Human Genome Analysis course- an invaluable insight into the possible genetic analysis methods.

Finally I would like to mention my family and friends for their support in the many forms it took, and my fellow myopia group members, including Jane for her initial encouragement, Caro, Snow, Tetyana and Ankush. Also thanks to the Biophysics group for welcoming me, as an honorary Biophysicist, into their tea room.

Abstract

The Family Study of Myopia is a research project aiming to discover genetic loci causing susceptibility to high myopia. As part of this investigation, the heritability of refractive error and other ocular components was estimated for a large Irish-Welsh multi-generational pedigree using variance components analysis software, SOLAR. Heritabilities of 0.39 ($p=6.92 \times 10^{-5}$, S.E.=0.14), 1.00 ($p=1.84 \times 10^{-4}$, S.E.=0.22) and 0.30 (p=0.13, S.E.=0.33) were found for refractive error, mean corneal curvature and axial length, respectively. Heritability of refractive error was also calculated using within-family regression and a Markov Chain Monte Carlo method producing estimates of between 0.12 and 0.73. This high heritability of ocular refraction suggests the potential for finding susceptibility loci for the control and development of refractive error. To pinpoint these loci, areas of chromosomes which have previously been suggested to harbour genes controlling refractive error were investigated to determine whether linkage was present within this family. DNA was extracted from mouthwashes and linkage analysis was performed. SOLAR revealed no significant linkage to the loci tested, reinforcing the theory that myopia is a highly heterogenous disease. The maximum twopoint LOD score was within the MYP6 locus at marker D22S1176 (LOD=1.19). Multipoint analysis showed the maximum LOD score at the MYP3 locus, between markers D12S1605 and D12S354 (LOD =1.37).

In a separate study of 96 families containing a highly myopic child and the two parents, the involvement of a candidate gene encoding the protein myocilin (MYOC) was examined using an association analysis. There was weak evidence of over-transmission of allele 3 of MYOC1, a marker in the 5' untranslated region of the gene, and under-transmission of allele 4 of that same marker (both $p<0.05$). This suggests MYOC may play a small role in the causation of high myopia development but a larger sample is needed to establish this conclusively.

Table of Contents

Figures

Tables

Equations

Abbreviations

A	adenosine
ASP	affected sibling pair
Bp	base pairs
C	cytosine
CLAMP	contact lens and myopia progression trial
cM	centi-morgan
COMET	correction of myopia evaluation trial
D	dioptres
DC	cylindrical dioptres
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide-triphosphate bases
DS	spherical dioptres
DZ	dizygotic twins
EDTA	ethylenediaminetetraacetic acid
ETDT	Extended Transmission/Disequilibrium Test
FBAT	Family Based Association Test
G	guanine
GC	guanine-cytosine content
gDNA	genomic deoxyribonucleic acid
H^2	broad sense heritability
h^2	narrow sense heritability
H2r	residual heritability
H2q1	heritability due to the specified quantitative trait locus
HBAT	(haplotype) Family Based Association Test
HCl	hydrochloric acid
hLOD	heterogeneity logarithm of odds ratio
HW	Hardy Weinberg
HWE	Hardy Weinberg equilibrium
IBD	identity by descent
IBS	identity by state
K	keratometry reading of corneal curvature
KCl	potassium chloride
KOH	potassium hydroxide
L.E.	left eye
LINKAGE	Linkage analysis software
LMM	least minus meridian
LOD	logarithm of odds ratio
Mar	The Centre for Medical Genetics, Marshfield Clinic
MCMC	Markov Chain Monte Carlo
MDA	multiple displacement amplification
MG	multi-generational family
$MgCl_2$	magnesium chloride
MLE	maximum likelihood estimate
MMM	most minus meridian
mRNA	messenger ribonucleic acid
MSE	mean spherical equivalent

MZ	monozygotic twins
NaCl	sodium chloride
$(NH_4)_2SO_4$	ammonium sulphate
OMIM	online Mendelian inheritance in man
PAL	progressive addition lenses
PBAT	Power calculation of Family-Based Association Tests
PCR	polymerase chain reaction
PEDCHECK	Pedigree error checking software
QTL	quantitative trait locus
R.E.	right eye
RGP	rigid gas permeable
RNA	ribonucleic acid
Rx	refractive error
S.A.G.E.	Statistical Analysis for Genetic Epidemiology
SDS	sodium dodecyl sulphate
Sib	sibling
SLINK	Linkage simulation software
SLRP	small leucine-rich proteoglycans
SNP	single nucleotide polymorphism
SOLAR	Sequential Oligogenic Linkage Analysis Routines (variance components linkage analysis)
T	thymine
TDT	Transmission/Disequilibrium Test
TDTEX	TDT software
TE	tris-ethylenediaminetetraacetic acid buffer
T_m	melting temperature
TNE	tris-sodium- ethylenediaminetetraacetic acid buffer
tSNP	haplotype tagging single nucleotide polymorphism
U	uracil
V_A	additive genetic variance
V_P	phenotypic variance
WGA	whole genome amplification
WTAC	42^{nd} Wellcome Trust Advanced Course. Human Genetic Analysis: Genetic Analysis of Multifactorial Disease.
λ_s	sibling recurrence risk ratio
θ	recombination fraction

Chapter summary and aims

Chapter one provides an overview of the basic principles involved in this thesis. The background describes the trait under investigation- refractive error- concentrating on one extreme of the condition; high myopia. An overview of inheritance and the tools used to construct inheritance patterns is provided and a discussion of the alternative methodologies is presented.

For any genetic study, a cohort of subjects needs to be recruited. The methods and results of the recruitment process used are presented in chapter two, along with a description and summary of the cohort and their trait characteristics.

Chapter three investigates the trait characteristics of the large Irish-Welsh pedigree and the heritabilities thereof using several different methods. The higher the heritability of the trait, the stronger the genetic basis, so providing a measure of potential success of locus mapping for that trait.

Genotypes of the subjects' samples provide the raw data for the genetic analysis. In order to avoid the limitations present for epidemiological studies requiring DNA samples in the form of blood, mouthwashes have been used to increase sample size and also ease of collection. Chapter four presents arguments in favour of this method of DNA sample collection and also investigates the benefits of the method of whole genome amplification to overcome the limitations brought about by sample depletion. Quality and quantity of DNA from the Irish-Welsh family samples are investigated.

Chapter five presents a background of different genetic analysis methods along with the basics of molecular biology related to genotyping and DNA. Genotyping results at loci previously linked to myopia provide the raw data to perform the linkage analysis. The presented results of the analysis will indicate the presence or absence of linkage to previously identified high myopia loci for this Irish-Welsh pedigree.

Using the cohort of trios recruited, an association study was performed for the MYOC gene, which encodes the protein myocilin and has previously been found to be associated with high myopia in a Hong Kong population. Chapter six presents a description of the methodological basis provided and the results of three association study strategies.

Chapter seven is a summary of the findings and general discussion of the work undertaken.

1. Linkage Analysis in High Myopia-An Introduction.

1.1. Myopia

1.1.1. Refractive Error

Refractive error is a quantitative trait and has a continuous distribution within the population (Bear, 1991). The frequency distribution of ocular refraction is leptokurtotic rather than normal and is skewed towards myopia (Sorsby et al., 1957). Figure 1.1. shows the frequency distribution of refractive error in a population attending optometric practice, n=88048. This data is taken from a database of patients attending optometric practice in the North-West of England and Scotland, courtesy of Conlons Opticians Ltd.

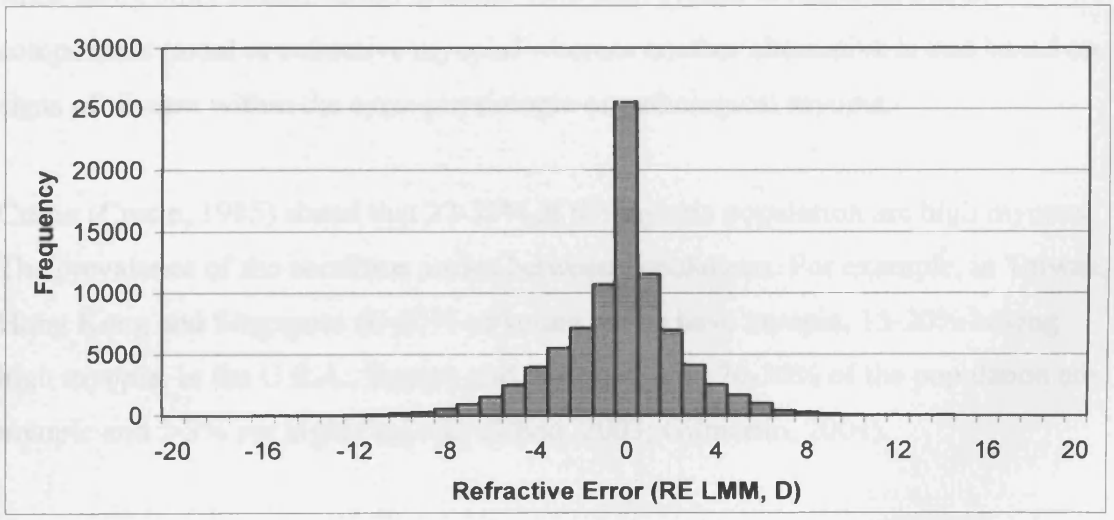


Figure 1.1. Frequency distribution of refractive error of the right eye (RE) in the least minus meridian (LMM) in dioptres (D).

Myopia is defined as a refractive condition in which parallel rays of light are focused in front of the retina with relaxed accommodation. This occurs due to the eye having either a vitreous chamber depth that is too great for the refractive power of the eye or

an ocular refractive power that is too great for the vitreous depth of the eye (Grosvenor and Goss, 1999) (figure 1.2.).

High myopia (generally classified as a refractive error of –6.00D or over) is typically the result of abnormal lengthening of the vitreous chamber. It is associated with a higher risk of glaucoma, cataracts, macular degeneration and retinal detachments (Saw et al., 2005) and is, therefore, one of the major causes of blindness in the U.K. working age population.

Severity is often used as a threshold for myopia classification in myopia studies, either by dichotomising the trait (section 1.2.6.1.1) or dividing it into several sub-categories of magnitude i.e. mild, moderate and marked.

Alternative ways to classify myopia include age of onset (Grosvenor, 1987). The suggestion that all myopia can be divided into: congenital, youth onset, early adult onset and late adult onset removes any assumptions of causation which can be implied when using other classification systems. One such system is based on ocular components (axial or refractive myopia) whereas another alternative is that based on signs of disease within the eyes: physiologic or pathological myopia.

Curtin (Curtin, 1985) stated that 27-33% of the myopic population are high myopes. The prevalence of the condition varies between populations. For example, in Taiwan, Hong Kong and Singapore 60-80% of young adults have myopia, 15-20% having high myopia. In the U.S.A., Europe and Australia only 20-30% of the population are myopic and 2-3% are highly myopic (Choo, 2003; Gilmartin, 2004).

The clinical implications and potential consequences of high myopia and the increasing prevalence within the population (Gilmartin, 2004) emphasises the need to identify the underlying mechanisms of this disease.

1.1.2. Ocular Components

The refractive system of the human eye is made up of several individual refractive components (figure 1.2). The first surface, the cornea, is a convex refracting surface with a power of approximately +42.00D. The light then passes through the anterior chamber to reach the crystalline lens, which has an accommodative facility allowing a changeable focus. The equivalent power is approximately +20.00D in its weakest form. The light is focused onto the retina after passing through this optical system, the retina being on average 24mm behind the cornea, this distance is known as the axial length (Tunnacliffe, 1993). In emmetropic eyes, the corneal power and axial lengths are well correlated and it is the breakdown of the correlation between these refractive components which occurs in an ametropic eye (Sorsby et al., 1957). With increasing ametropia, extreme values are apparent in the axial length rather than with the corneal power. This, along with the observations that lens power varies little and anterior chamber depth appears to be only loosely related to axial length and refractive error, suggests that axial length is the primary determining component of refractive error.

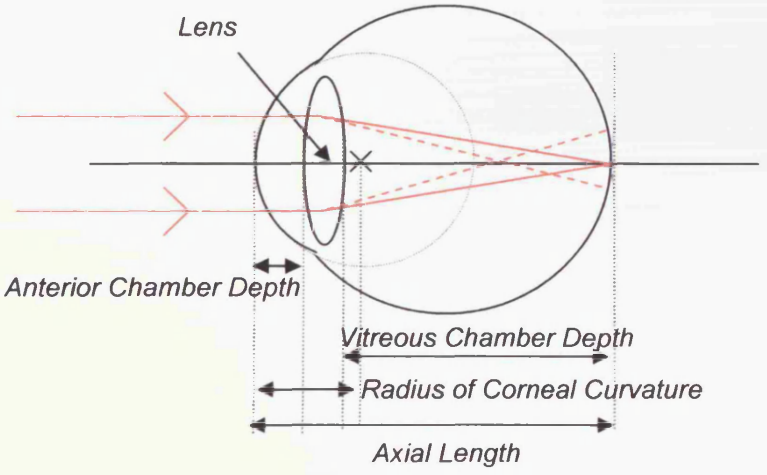


Figure 1.2. Ocular components of refractive error. In an emmetropic eye the light is focused onto the retina (solid line). In a myopic eye, the light entering the refractive system will be focused in front of the retina (dashed line).

1.1.3. Aetiology and associations with myopia

1.1.3.1. Myopia and near work

There is strong evidence for the contribution of non-genetic factors to myopia development including age, occupation, ethnicity, intelligence and near work. Studies have reported that associations can be found with near work but not necessarily causation (Rosenfield and Gilmartin, 1998).

Evidence for this comes from examining the change in prevalence of myopia within different populations with changing or different near work demands. In populations such as precontact Africans, Eskimos and Australian Aborigines where the mode of life involves hunting and gathering, reduced visual acuity was very uncommon as this would put individuals at a selective disadvantage (Bear, 1991). In Native American populations in Alaska and northern Canada, low myopia has increased in prevalence since World War Two, coinciding with the introduction of compulsory formal education and therefore an increase in exposure to prolonged periods of near work. It has also been found that myopia is much rarer in illiterate compared to literate populations (Lithander, 1999).

Studies have also shown that within groups of individuals who have occupations or lifestyles in which a lot of near work is required, the prevalence of myopia is higher (Goldschmidt, 1968; Curtin, 1985; Bear, 1991; Shimizu et al., 2003).

There are several ways in which near work may induce myopia. One such theory comes from animal studies which have shown that defocus of the retinal image due to light focusing behind the retina induces myopia (section 1.1.4.). On correcting myopia in the distance with spectacle or contact lenses, the eye is required to focus on near objects by altering the curvature of the eye's natural lens, known as accommodation. Generally, individuals do not accommodate accurately, producing less accommodative effort than is required, this is known as the accommodative lag and is quantified by measuring the difference between the required accommodative demand and the actual accommodative response (Gwiazda et al., 2003). This accommodative

lag causes the image to be positioned behind the retina. Myopes particularly have been found to accommodate significantly less than emmetropes (Gwiazda et al., 1993), especially early onset myopes (Nakatsuka et al., 2005).

Clinical studies have been based on these findings and have attempted to slow myopia progression by the use of multifocal spectacle lenses to ensure that the image is focused clearly on the retina at all viewing distances and so compensates for the lag of accommodation. Lenses used are either: 1. a bifocal lens with the top portion correcting the vision in the distance and the lower portion with a shorter focal length to correct vision for near distances, such as for reading, or

2. a progressive addition lens (PAL)

which has a smoother progression from the distance correction to the near. These multifocal lenses reduce the amount of accommodation required to place the image at near onto the retina.

Reduced accommodation has been found to be associated with a shift in heterophoria towards esophoria (Fulk et al., 2000; Gwiazda et al., 2003). That is, when the eyes are dissociated, they move out of alignment to rest at their natural position and realign when viewing objects binocularly. In esophoria the natural resting position of the eyes is slightly converged. Therefore in order for the esophore to maintain binocular single vision at near, the accommodation must be relaxed to reduce the accommodative-convergence to overcome the esophoria and so increasing the lag of accommodation. The treatment with multifocal lenses is, therefore, more applicable to this sub-section of myopes. In a single vision versus PAL study, the myopia in esophoric subjects corrected with single vision lenses progressed more than twice as much as the non esophores (Brown et al., 2002). The Correction of Myopia Evaluation Trial (COMET) (Gwiazda et al., 2003) found that the children with larger accommodative lags developed significantly more myopia than those with smaller lags. Those children with larger lags and increased near work had a larger treatment effect of the myopia progression with PALs than the children with smaller lags or more myopia. Fulk and colleagues (Fulk et al., 2000) also found a reduction in myopia progression in myopic esophores who were treated with bifocal lenses compared to single vision lenses of 0.25D over 30 months.

Other clinical studies have found treatment effects, including a reduction of myopia progression of 0.50D in a cohort of Chinese children wearing PALs compared to single vision lenses over two years (Leung and Brown, 1999). The results cannot, however, support or reject the accommodative theory if the accommodative functions of the cohort are not measured but the results are nevertheless useful in gaining a more informative picture of the possible aetiology of myopia progression. Other studies have found similar effects, but not necessarily statistically significant results.

An alternative trial method includes the attempt to slow progression of myopia by the use of contact lenses. Theories behind using this method of refractive correction to reduce myopia progression include the possibility that the image on the retina is of higher quality and so reducing any blur stimulus for myopia inducement. Also, due to its mechanical properties, particularly for rigid gas permeable lenses, the contact lens may flatten the cornea slightly and therefore cause a temporary reduction in myopia. A recent finding showed that in a study population of school age Singapore children, no evidence was found that wearing rigid gas permeable contact lenses significantly reduced the progression of myopia (Katz et al., 2003). Changing from a correction of spectacles to soft contact lenses was associated with an increase in myopia progression in a study which initially corrected the subjects with either bifocals or single vision lenses, the rate of progression being slower in the bifocal wearer's group in the first year but remaining the same for both groups in the following two years (Fulk et al., 2003).

The Contact Lens and Myopia Progression study (CLAMP) compared the progression of myopia in children wearing soft and rigid gas permeable contact lenses over a three year period. Initially the children went through a preparatory stage of contact lens wear. Those completing this stage with no aversions to the lens wear were randomised into two groups of the different lens types. Baseline measurements of refraction and ocular component dimensions were taken and the children were followed up at regular intervals. Over the three years the RGP wearers' myopia progressed less than the soft lens wearers' (-1.56 D compared to -2.19D respectively). This reduction of myopia progression was limited to only the first year of the trial, possibly due to the change of corneal curvature induced by the lens. The corneal curvature steepened more in soft

lens wearers by 0.26D but there was no statistically significant difference in axial length between the two groups (Walline et al., 2004).

Under correction of myopia, that is the prescribing of spectacle lenses that are less than the power required for a clear retinal image at distance, but closer to that required for near with a non-accommodating eye, was assessed in a two year study of Malaysian myopes aged between 9-14 years. The children were randomised into one group of fully corrected spectacle wearers and one wearing spectacles with an undercorrection of +0.75D. This resulted in the under-corrected group showing a greater axial length elongation than the fully corrected group, suggesting under-correction as an ineffective and unwarranted method of myopia control (Chung et al., 2002).

A recent study looked at the effect of correcting children by monovision in an attempt to reduce accommodative lag and therefore myopia progression (Phillips, 2005). A cohort of 18 children was prescribed a distance prescription for their dominant eye and a near correction of approximately +2.00D difference for their non-dominant eye. The results of which found that the children were accommodating over their distance prescriptions and using their dominant eyes to read. However, there was a significant difference in myopia progression (0.36D per year) and vitreous chamber depth (0.13mm per year) between the two eyes suggesting differing optical consequences for uni- and bilateral undercorrection and that the myopic retinal defocus experienced by the undercorrected eye counteracted the abnormal axial elongation.

In the same way that clinical studies have used multifocal lenses to reduce the accommodation required, other studies have trialled pharmaceutical agents to do the same in an attempt to reduce myopia progression. Accommodation is controlled by the activity of the ciliary muscle which is connected to the lens by zonules. As the ciliary muscle contracts, the tension in the zonules is released and the lens changes shape to focus the image at near. When the eye is relaxed, the zonules resume their tension and the lens returns to its thinner, less curved shape. Paralysing the ciliary muscle will prevent accommodation and is known as cycloplegia. This is done by targeting the muscarinic receptors of the ciliary muscle with pharmaceutical agents. In one double-masked randomized trial, a progression of only 0.42D of myopia over 18

months was found in a group of children treated with the non-specific anti-muscarinic agent atropine with multi-focal lenses compared to 1.19D in the multi-focal lens wearers without atropine and 1.40D in the single-vision lens, placebo control group (Shih et al., 2001). Pirenzipine has a similar muscarinic antagonistic action to atropine, but acts selectively on the M1 receptors thereby reducing effects of cycloplegia and mydriasis (dilation of the pupil). It also has fewer side effects and is a less toxic drug. A 2% gel formulation was found to reduce myopia progression by approximately 50% over a one year period (Tan et al., 2005). Due to the specificity of the target receptor sites of this drug and its reduced cycloplegic action, these findings support that of animal work suggesting that myopia development is independent of the accommodation mechanism (Tan et al., 2005).

So far these studies have not been able to find a system of clinical management of myopia which will reduce its progression (Table 1.1). Whilst some studies show significant treatment effects (Leung and Brown, 1999; Fulk et al., 2000; Shih et al., 2001; Fulk et al., 2003; Gwiazda et al., 2003; Walline et al., 2004; Tan et al., 2005), especially for near-esophoric myopes, a clinically significant effect to change the prescribing guidelines for clinicians has yet to be produced. If an effective treatment is proposed there are many effects that could have a major influence on the reliability and success of this treatment. When prescribing multifocal lenses to children there is the increased cost of the lenses and the potential problems of adaptation to consider as well as the social implications for the child. Contact lenses also have a factor of non-compliance due to hygiene and care issues. The increased number of after-care visits as well as increased cost and increased risk of health complications would be issues to tackle. Side effects of pirenzipine would pose problems such as a degree of mydriasis and cycloplegia, although these would be reduced compared to atropine (Tan et al., 2005). Compliance would also be a major difficulty to this treatment. This option does seem, however, the most likely source of a potential treatment due to the success of the trials (Tan et al., 2005). Table 1.1 shows a summary of some of these and oher trials, the strong evidence and potential of treatment appears to be with the pharmaceutical agents.

Table 1.1. Patient, study characteristics and outcomes of randomized clinical trials taken from (Saw et al., 2002)

Study	Country	Methods	Masking	Total Randomized	Intervention	Follow up (years)	Group rates (Dioptre/year)	Evidence
Yen 1989	Taiwan	Parallel RCT Baseline similar	No	247	1% atropine 1%cyclopentolate Normal saline*	1	-0.22 (0.54) -0.58 (0.49) -0.91 (0.58)	B,I
Shih 1999	Taiwan	Parallel RCT Baseline similar	Single	200	0.5% atropine 0.25% atropine 0.1% atropine 0.5% tropicamide*	2	-0.04 (0.63) -0.45 (0.55) -0.47 (0.91) -1.06 (0.61)	B,I
Shih 2000	Taiwan	Parallel RCT (central office) Baseline similar	Double	227	0.5% atropine Multifocals Single vision*	1.5	-0.28 (0.05) -0.79 (0.05) -0.93 (0.06)	B,I
Schwartz 1981	USA	Parallel RCT in twins Baseline similar	Single (assessor)	26 pairs	1% tropicamide Single vision*	3.5	Paired analysis	C,I
Grosvenor 1987	USA	Parallel RCT (random number table). Baseline similarity unknown	No	207	Bifocals +2.00D Bifocals +1.00D Single vision*	3	-0.32 -0.34 -0.32	C,I
Parssinen 1989	Finland	Parallel RCT (sealed envelopes)	No	240	Bifocals +1.75D Single vision (distance only) Single vision*	3	-0.56 (0.3) -0.59 (0.3) -0.49 (0.3)	C,I
Jensen 1991	Denmark	Parallel RCT Baseline similar	No	150	Bifocals +2.00D Single vision + 0.25% timolol Single vision*	2	-0.48 (0.28) -0.59 (0.30) -.57 (0.36)	C,I
Fulk 1996	USA	Parallel RCT (sealed envelopes) Baseline similarity unknown	Single	32	Bifocals +1.25D Single vision*	1.5	-0.39 (0.12) -0.57 (0.11)	C,I
Fulk 2000	USA	Parallel RCT Baseline similar	Double	82	Bifocals +1.50D Single vision*	2.5	-0.40 (0.27) -0.50 (0.26)	C,I
Horner 1999	USA	Parallel RCT Baseline similar	Single	175	Soft contact lenses Single vision*	3	-0.36 (0.03) -0.30 (0.03)	C,II

Eye drops — *Bifocal lenses* — *CL*

* Control group; D Dioptres; RCT Randomized Clinical Trial

B = moderately important information; C = relevant but not critical recommendation; I = strong evidence supporting recommendation; II = substantial evidence supporting recommendation but lacking some qualities required for strong support.

1.1.3.2. Myopia and other associations

Myopia has been found to have an association with a number of other factors, for example ethnicity (section 1.1.1.) and IQ (Saw et al., 2004b). A retrospective study by Chong and colleagues investigated the prevalence of myopia in children who were breastfed and in those who were not. The results showed that 62.0% of breast fed children were myopic, compared to 69.2% of children who were not breast fed, giving an odds ratio of 0.73 and therefore a protective effect. The authors speculated that this was due to the increased levels of docosahexaenoic acid passed to the child in the mother's milk, which is beneficial for neurodevelopment (Chong et al., 2005).

A study of a Japanese population found an association of myopia with higher education and management based occupations in men. For women, an association with high income and clerical based occupations was found (Shimizu et al., 2003). The clerical work and education associations support the hypothesis that increased near work may be associated with myopia.

Subjects born prematurely are at a higher risk of developing myopia than babies born at full term, the incidence increasing with decreasing gestational age and incidence of retinopathy of prematurity (Fledelius, 1996; Snir et al., 2004).

Other studies have investigated such factors as diet (Edwards, 1996), night lighting (Gwiazda et al., 2000; Zadnik et al., 2000) and parental smoking (Saw et al., 2004a) but found no clinically significant association.

1.1.3.3. Myopia and association with systemic and ocular disease.

Myopia is manifested in a number of rare, monogenic disorders (reviewed by (Curtin, 1985)).

Albinism is caused by a pigment deficiency and mainly affects the skin and hair, but primarily structures of the eye. A study by Edmunds and colleagues in 1949 found a high myopia prevalence of 12% in a group of albino subjects (Edmunds, 1949).

Stickler syndrome causes skeletal abnormalities producing premature degeneration of the joints. Ocular associations of Stickler's syndrome include high myopia, anterior chamber abnormalities and secondary glaucoma. It can be defined as type I (Online Mendelian Inheritance in Man (OMIM) #120140) or type II (OMIM #604841), this classification depending on the gene within which the mutation occurs which, in turn, gives rise to a different vitreous phenotype. A mutation in COL2A1, which codes for the α_1 chain of type II collagen, produces a membranous vitreous phenotype. Approximately two thirds of families affected by Stickler syndrome are linked to this gene (Richards et al., 1996). Less common is a mutation in COL11A1, which codes for the α_1 chain of the collagen type XI molecule, producing a type II vitreous phenotype described as "beaded" (Martin et al., 1999). There are other phenotypic differences between the two classifications of individuals. Subjects with the type I Stickler syndrome commonly have the additional ocular complications of retinal detachment and vitreous degeneration whereas these signs are seldom seen in subjects with type II. However, hearing loss is quite common in type II Stickler syndrome but not in type I. An alternative mutation has also been shown to be linked to Stickler syndrome with the gene COL11A2 but this shows no ocular involvement (Richards et al., 1996).

In Marfan syndrome (OMIM #154700) a prevalence of myopia of 83% has been found (Curtin, 1985). The main characteristics of this disease include elongation and thinning of the long bones. Cardiac, pulmonary and renal problems are also often associated. Ocular abnormalities include upward subluxation of the lens, a higher incidence of cataract and abnormalities of the trabecular meshwork and outflow system. All cases of Marfan syndrome appear to be due to a mutation in the gene FBN1 which codes for fibrillin which is found in the connective tissues. Mutations of this gene also cause Weill-Marchesani syndrome (OMIM #608328), another connective tissue disorder associated with high myopia.

Homocystinuria (OMIM +236200) is a metabolic disorder and is mainly associated with ocular complications of ectopia lentis as well as myopia.

The prevalence of high myopia is approximately 27% in subjects with Down Syndrome (OMIM #190685) but there is also a high prevalence of keratoconus (OMIM #605020, #608932, #608586, #609271) of approximately 15% (Shapiro and France, 1985; Van Allen, Fung and Jurenka, 1999) which is also associated with myopia.

Other ocular complications associated with myopia include congenital stationary night blindness, nystagmus, retinitis pigmentosa and cataracts. Cataracts particularly have a high association, especially with high myopia. The association with cataracts has also been found in a number of studies based on different types of cataracts although care has to be taken to avoid including those individuals with index myopia; the cataract therefore causing the myopia (Weale, 2003). A study carried out in the Blue Mountains, Australia revealed a high risk of developing posterior sub-capsular cataract for high myopes (OR=5.5) compared to an inversely related association with hyperopes (OR=0.6). The risk was greater with increasing myopic refraction (Lim et al., 1999). This same Blue Mountains' study population was used to investigate the associations of myopia with glaucoma. The results showed that myopic subjects had a two-to-threefold greater risk of developing glaucoma than emmetropes (Mitchell et al., 1999). This result was supported by the Beijing Eye Study, which found an odds risk ratio of 7.56 of high myopes developing glaucoma compared to non-myopes (Xu et al., 2006).

The association of high myopia with systemic diseases was investigated in a community based population of children. It suggested 56% of highly myopic children had no systemic disease association but 19% had associations with Stickler's syndrome, Weill-Marchesani syndrome or homocystinuria. Ocular abnormalities such as amblyopia or retinal dystrophies were found in 25% of the children (Logan et al., 2004). A separate study based in a hospital ophthalmology department found only 8% of highly myopic children had no underlying systemic or ocular association, the difference in prevalence of this association with the previous study may be due to the population base, the hospital department being a more likely situation in which to see

a high concentration of complex cases than in community based clinics. Systemic associations as above including Down's syndrome and prematurity were found in 54% of the children and the remaining 38% had ocular problems, similar to those mentioned and including lens subluxation and coloboma (Marr et al., 2001).

Even though the associations are common within these cohorts, high myopia is still found in only a small section of the population, approximately 2%. In children this number is much smaller. A cohort of 7600 school children aged seven had a myopia prevalence (more than -1.00D) of only 1.1% (Barnes et al., 2001). Conditions such as Stickler and Marfan syndrome are even rarer having a prevalence of only 0.0001% and 0.005% respectively in the general population and as there is an age dependency, this will be even smaller in children.

These syndromic disorders are relatively rare conditions and, although a genetic study of a cohort of families with myopia associated with one of these will provide greater power than non-syndromic myopia due to the increased genetic homogeneity, it will not identify those genetic risk factors applicable to the general population. Therefore any condition which causes or is associated with myopia acts as an exclusion criterion for contributing to a genetic study that aims to investigate genetic risk factors for a general population. These ocular or systemic associations can be identified either by observing clinical signs or by genetic screening and will ensure the basis of the disease is not confounded by these factors.

1.1.4. Animal Models of Myopia

Inducing myopia in animal models such as tree shrews, chickens and monkeys enables investigations into the mechanisms controlling the development of myopia. Form deprivation is a commonly used method, degrading the retinal image by adding a frosted lens in front of the eye. Many studies found this causes axial elongation, primarily of the vitreous chamber (Irving et al., 1992).

Eye length is controlled during post-natal development, in order to keep the focused image on the retina by using visual feedback. During a critical period of development, myopia can be induced by simulating an innate refractive error, e.g. by fitting a lens of known power in front of the eye. When an eye is fitted with a positive lens this will induce myopic blur, causing distant objects to be blurred. Over a period of days or weeks, the retina will be moved towards the focal point of the image by reducing the rate of axial elongation and increasing the choroidal thickness. The opposite will happen when the eye is fitted with a negative lens, the hyperopic blur causing the rate of axial elongation to increase and the choroid to reduce in thickness (Zhu et al., 2003).

Chick studies have shown that not only will the eye attempt to emmetropise, but suggest that the sign of blur can be distinguished and therefore guide the growth towards emmetropia (Park et al., 2003). More investigations have looked at the effect of blur exposure time on lens compensation (Winawer and Wallman, 2002; Zhu et al., 2003) which found that brief periods of myopic defocus may contribute to a protective effect, reducing the rate of myopia progression which implies the amount of defocus is not summed linearly for myopic or hyperopic defocus. Whether these principles may be useful for human eyes remains to be seen. Winawer (Winawer et al., 2005) has suggested that the total time spent on near work not taking into account the time spent looking into the distance and the defocus produced, cannot predict the degree of myopia.

The precise aetiology of myopia progression is, as yet, unknown. It is likely that this trait is under the control of a number of factors, including environmental and genetic influences. Clinical and experimental studies along with genetic studies will provide an insight into possible mechanisms of refractive error development.

1.2. Genetic Analysis

Linkage analysis is a tool used to identify areas of a chromosome which cosegregate along with a trait and therefore may contain a causative gene. A background to linkage analysis and its underlying principles are described below.

1.2.1. DNA sequences, constituents and function with relation to genetic analysis

An organism's complement of deoxyribonucleic acid (DNA) is called the genome. The human genome is made up of 23 paired chromosomes, each of which consists of a chain of genes. Genes are long sequences of DNA, variants of which are called alleles. The human genome contains around three billion base pairs of DNA which codes for approximately 30 000 genes.

Genes have several functions. Firstly, replication is necessary in order to maintain the continuation of the species. Secondly, genes contain information regarding a specific protein and therefore a specific function. This function being either as a structural component of the cell or involvement in a chemical process within the cell, enabling the cell, and therefore the organism, to function.

Considering these 30 000 genes and approximately 100 000 proteins that are present within the human body, it is extremely unlikely that each gene is responsible for only one function alone. Each gene will code for a different number of proteins due to alternative splicing or post-translational modifications. It is the interaction between these different proteins produced and expressed along with environmental factors that determine the trait rather than there being a direct relationship between DNA sequence, protein expression and functionality.

1.2.1.1. DNA

DNA consists of two intertwining chains of nucleotides, the nucleotides themselves comprising a phosphate, a pentose sugar and one of four bases: guanine, cytosine (the purines), adenine or thymine (the pyrimidines). The nucleotides are joined together to form a polynucleotide chain with a phosphate backbone. The two polynucleotide chains are held together by hydrogen bonds joining complementary bases together. Specifically adenine will bind exclusively to thymine and cytosine only to guanine. The complementary nature of the base pairs and the hydrogen bonds between them maintains the double helical structure and therefore the stability of the molecule.

1.2.1.2. Protein Synthesis

The synthesis of protein is required to maintain the constituents and the viability of a cell during its renewal and the growth and development of an organism. There are three steps in this process: replication, transcription and translation. If the protein synthesis is inaccurate, changes occur within the sequence and may alter the function.

1.2.1.2.1. Replication

During replication of the DNA the two polynucleotide chains will separate and each will act as a template for new strands composed of complementary bases. Once replicated each strand can, once again, separate from its template and so act as a new template for further replication.

1.2.1.2.2. Transcription

The sequence of DNA bases represents a code for synthesising proteins, the order and content of the gene in turn defining the function (of which there may be one or more within the system). During the process of transcription, the DNA sequence is

replicated with complementary bases and a ribose sugar in place of the deoxyribose sugar to produce a single strand of ribonucleic acid (RNA). This is similar in structure to DNA, the difference, along with the alternative sugar, is the replacement of thymine with a different nitrogenous base, uracil. RNA molecules which code for proteins- messenger RNA (mRNA) - leave the nucleus of the cell and transfer to the cytoplasm where protein synthesis occurs.

1.2.1.2.3. Translation

The next step of this process is translation, in which the sequence of mRNA bases is decoded. The mRNA sequence is read three bases (a codon) at a time. Each codon binds to a specific amino acid, which is attached by means of a transfer RNA molecule, a complementary three base sequence attached to the amino acid. As the complementary bases attach to the RNA, an amino acid chain is built up to form the protein, the structure of which is regulated by its sequence. If this sequence is altered, the structure of the protein is altered which can potentially affect its function. Malfunctioning genes may be unable to maintain the organism's chemical processes and so can lead to the expression of a disease phenotype.

1.2.1.3. Genes and alleles

A gene is made up a sequence of DNA which comprises of areas which will transcribe proteins, exons, and interspersed sequences which do not perform this function, introns. Marking the gene's boundaries are a promoter region to which proteins bind and initiate the transcription and, at the other end, an area marking the end point of transcription.

Genes are connected in a linear fashion to produce chromosomes of which humans have 23 pairs. The nucleus of most cells of the human body will contain all of these chromosomes. Gametes will, however, only have one of each pair of chromosomes and therefore offspring only inherit half of their genetic information from one parent

and half from the other, described as Mendel's first law. A point on a chromosome, a locus, will be paired with its homologous chromosome, the other one in the pair, and the two gene variants (alleles, see below) at that locus will describe that individual's genotype.

Microsatellite markers are tandem repeats of nucleic acid bases mostly found within introns, for example, a dinucleotide repeat of CA. Each allele (i.e. a variant form of a gene or specific polymorphic site (Terwilliger and Goring, 2000)) is represented by a different number of repeats, i.e. a sequence of three repeats may be allele 1, and four repeats may be allele 2. Considering a pair of homologous chromosomes, there will be two alleles for each marker, i.e. a genotype. Linkage analysis makes use of these variations in DNA sequences within genes, to investigate the cosegregation of a trait and specific areas of a chromosome.

Each marker locus has a specific number of alleles. The higher the possible number of alleles of a marker, the more likely it is that an individual will be heterozygous at that locus: that is, they will have a different allele on each of their homologous chromosomes. When tracing the inheritance of an allele through a pedigree, a more definite pattern of which parent the allele has been inherited from can be made if the individuals are heterozygous and therefore the more informative the genotypes are for linkage analysis. This measure of informativeness is known as the heterozygosity (equation 1.1.) and should be considered when choosing the most efficient and informative markers to use when genotyping individuals within a genetic study.

$$H=1-\sum_{i=1}^{n} p_i^2$$

Equation 1.1. Heterozygosity. Where H is the heterozygosity, p is the frequency of the *ith* allele at the locus and n is the number of alleles

Another consideration is the spacing of the markers and the genetic distance the analysis will cover. A genome scan will space markers, on average, approximately 10cM apart whereas investigating a targeted region of chromosome requires a denser distribution. A centimorgan is the genetic distance based on the recombination rate, θ. Two points on a chromosome 1cM apart will have a 1% chance of having a recombination occurring between them. The denser the spacing of markers across a chromosomal region, the more information is obtained. However, for linkage analysis, spacing of markers of 1cM will give 100% information content and increasing the density will not significantly increase the power to detect linkage (Ulgen and Li, 2005).

A vital tool for genetic analysis is the knowledge of recombinations within chromosomes. During meiosis, at the prophase I stage in cell division, the chromosomes may come together and exchange homologous parts. This recombination or "crossing over" results in a different sequence of DNA on the offspring's chromosome compared to that of the parents. Recombination will happen, on average, once between each chromosome pair per meiosis. The further apart any two loci are on a chromosome, the more likely a crossing over will occur between them and so it is less likely that alleles at these loci will be inherited together. The recombination fraction, θ, defines the genetic distance between two loci and is a function of the probability that a crossing over will occur between the two loci. It can, therefore, be used to map the position of a locus on a chromosome. The recombination fraction will not exceed 0.5, this occurring when there is no linkage since the probability that an allele will be inherited from either parent by chance is 50%.

For a genetically based trait, the genotype at one or more loci will determine the phenotype of the individual. For a dichotomous trait, a particular allele at a locus may determine whether that individual is affected with the disease of interest. For a quantitative trait it may determine the value or extent to which the individual is affected.

Molecular biology methods are used to identify genotypes of individuals at specific pre-designated loci for use in linkage analysis.

1.2.1.4. Microsatellite Genotyping

This process involves identifying the genotype at each locus under investigation by differentiating between the size of the fluorescently labelled polymerase chain reaction products and therefore between alleles. The techniques involved are described below.

1.2.1.4.1. Polymerase Chain Reaction

This is a method of amplifying a specific sequence of DNA. A "master mix" is prepared which contains all of the required ingredients for the amplification. The master-mix is added to the DNA template and includes a buffer to optimise the chemical conditions for the reaction. Magnesium chloride ions are used as a cofactor for the enzyme, *Taq* DNA polymerase which is stable at high temperatures and initiates the amplification. Deoxynucleotide-triphosphate bases are used to extend from the primer strand and create the new sequence producing a new complementary strand of DNA. The primers are short sequences of DNA designed to be complementary to sequences surrounding the target section of DNA. They are designed to work optimally under the chosen PCR thermal cycling parameters in that they should have a specific melting temperature (T_m, approximately 50-65°C) and guanine and cytosine content (GC content between 40-50%). The initial stage of the PCR denatures the template DNA by separating the double stranded helix. Due to their complementarity, primers attach to sites on the sequence strands surrounding the section of DNA to be amplified, the amplimer. This process is known as annealing and the temperature that this occurs at during the cycling procedure can be tailored to optimise the efficiency of the reaction. The dNTP bases are taken from the mastermix and are joined to the primer creating a new complementary strand and so a new double stranded DNA molecule. In the second cycle of the reaction, this double

stranded molecule is again denatured and a new primer attaches at the complementary strand. This procedure continues until all of the ingredients are depleted and the reaction is saturated, or the number of cycles requested has been completed, resulting in PCR product of the specified sequence and length. Genotyping subjects at specific loci where microsatellite markers exist within the amplimer requires use of this technique, specifically using fluorescently labelled primers in the PCR to recognise alternatively sized molecules and so differentiate between alleles.

1.2.1.4.2. Primer design

Primers are specifically designed to be compatible with the PCR cycling parameters, by considering T_m and GC content (section 1.2.1.4.1.). Due to the required specificity of genotyping the procedure must be exactly complementary and amplimers must be amplified accurately. Certain primers may encourage the addition of an extra adenosine base to the product. In order to eliminate the potential genotyping errors produced by the uncertainty of the presence of this extra base, a "PIGtail" can be added to the 5' end of the primer which will encourage the addition of an adenosine molecule. A GTTTCT PIGtail has been suggested which results in nearly complete adenylation (Brownstein et al., 1996) and so eliminates this inconsistency and this source of potential genotyping error.

1.2.1.4.3. Identification of genotypes.

Once the PCRs have been performed, the reaction products are run on a DNA sequencer. This machine uses an electrophoresis based method to separate the differentially sized PCR products and is able to specify exact sizes of each of the products and therefore both alleles, in base pairs, using comparisons to a size standard and the detection of the fluorescence given out by the labelled forward primer. Each sequence length is assigned an allele number and so genotypes are produced.

When genotyping several loci, to reduce the number of genotype reactions run on the sequencer, the markers can be pooled into sets. These should be designed so that each set includes markers which have differing product sizes which are not over-lapping. Those in close proximity to each other should use a different label from their neighbours, these labels being determined by that attached to the forward primer.

1.2.2. Inheritance

How the alleles inherited affect the phenotype of the offspring is dependent on the mode of inheritance of the trait.

Mendelian diseases are inherited dominantly or recessively. In dominant inheritance only one copy of a disease allele is required in order to express the phenotype in an individual, whereas two copies are needed for a recessive autosomal disease to be expressed. A typical, rare dominantly inherited disease will show affected offspring at a ratio of 2:1 when one parent is affected. If both parents are unaffected and there is an affected offspring the disease must not be inherited dominantly. A recessive inheritance pattern shows unaffected heterozygous parents with only one of four offspring to be affected (figure 1.3.).

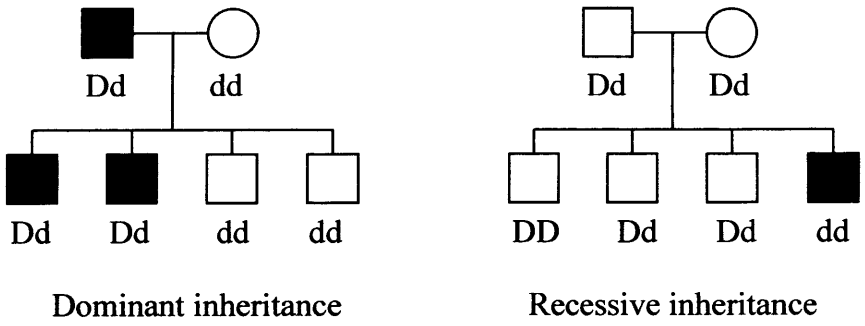


Figure 1.3. Example of typical recessive and dominant inheritance patterns where D is the dominant allele and d is the recessive allele. In the example for dominant inheritance the individual requires only one dominant disease allele (D) for the individual to be affected. For recessive inheritance the individual is considered affected if the disease allele (d) appears twice in the genotype.

A phenotype can exist which is different from the homozygous phenotypes and occurs for a heterozygote. Here, the locus is said to be codominant and the inheritance model additive. This is an important type of inheritance when investigating quantitative traits and is based on the biometrical model (figure 1.4.). This model defines a quantitative trait in terms of the allele frequencies and genotypic values of each causative locus (Camp and Cox, 2002) and describes the effect produced due to a quantitative trait locus (QTL). For example, the additive effect of the alleles may produce a phenotype equivalent to an additive value, *a*. Homozygous genotypes will produce the minimum and maximum phenotypes of $-a$ and *a* and the heterozygote will produce a phenotype somewhere between these, dependent on the dominance effects; the closer the heterozygous trait value is to the midpoint of the homozygous values, the smaller the dominance effects of either allele present in the homozygote.

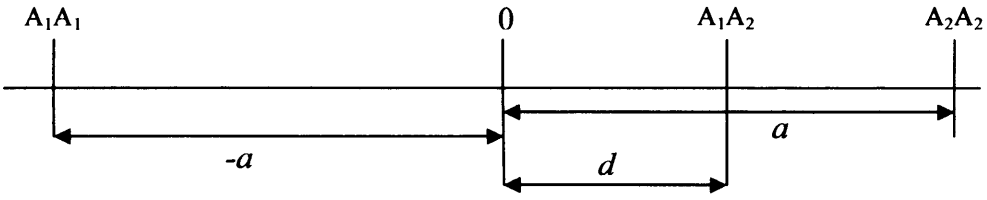


Figure 1.4. The additive genetic biometrical model (taken from (Camp and Cox, 2002)) where A_1 and A_2 are the alleles, a is the additive genetic value and d is the dominance deviation.

An additional form of non-Mendelian inheritance occurs through mitochondria. In this case the mutation is generally passed to the offspring by the mother. Mutations in the DNA may cause functional problems which most commonly manifest as muscle disorders (Weaver and Hedrick, 1997). A strong pattern of offspring affected with the disease when the mother is affected is shown due to only a negligible amount of mitochondrial-DNA being passed on by the father.

Relating the inheritance of the disease to the actual inheritance of marker alleles is achieved by the application of linkage analysis methods.

1.2.3. Linkage Analysis

Linkage is present when two loci on the same chromosome are inherited together more often than expected by chance. For linked loci the two syntenic loci are close enough to each other that recombination during meiosis is uncommon enough for their segregation to be detectable within families (Camp and Cox, 2002). Linkage analysis is a method which allows the examination of how a particular trait is transmitted through a family. If a particular marker allele is transmitted along with the phenotype it may be that this allele and the disease pre-disposing allele are linked. The likelihood that these loci are linked can be calculated by the ratio of the likelihood that they are linked compared to that of not being linked, known as the odds ratio (equation 1.2.). For simple pedigrees, this can be calculated by hand, identifying the number of recombinations occurring within the pedigree, having the knowledge of the phenotypes and the genotypes at the locus. The logarithm of this odds ratio calculation is taken as the standard measure (LOD score, equation 1.3.) of the evidence in favour of linkage. As the pedigree structure increases in complexity or the identification of the recombinants becomes less apparent due to unknown phase of the parental chromosomes (i.e. being unable to identify which of the offspring's chromosomes were inherited from which parent), the calculation becomes more complex and computer based calculations are required.

$$\text{Odds Ratio} = \frac{\text{likelihood the disease locus and marker are linked}}{\text{likelihood the disease locus and marker are not linked}}$$

Equation 1.2. Odds ratio

$$\text{LOD score} = \log_{10} [(1-\theta)^{n-r}.\theta^{r}/(1/2)^{n}]$$

where n = number of offspring
r = number of recombinant offspring
θ = recombination fraction = r/n
where parental phase is known

Equation 1.3. Logarithm of odds ratio.

A LOD score of 3 or more is usually taken as significant evidence for linkage and a value of over 2 is suggestive of linkage. This former value has been shown to be roughly equivalent to a conventional significance level of $p=1 \times 10^{-3}$ and a LOD score of less than -2 excludes linkage from that locus (Sham, 1998).

A "twopoint analysis" is performed when linkage is assessed between a trait and each marker locus, separately. In general, the higher the LOD score, the closer the marker is located to the disease locus. Multipoint analysis uses all of the markers genotyped along a section of chromosome and the position of the disease locus is varied. The probability of the locus segregating with the marker taking into account the location of the other markers is examined producing multipoint LOD scores. The advantage of using multipoint analysis over twopoint is the increase in information gained by use of haplotypes for linkage analysis. Haplotypes are the representation of the genotypes along the chromosome. Tracing the inheritance of a haplotype as opposed to an allele allows fewer assumptions to be made as any missing data can be allowed for within the section, conditional on whether the neighbouring genotypes are present.

1.2.3.1. Assumptions made for linkage analysis

Linkage analysis involves a number of assumptions:

1. The assumption of Hardy-Weinberg equilibrium states that the genotype frequencies between generations are not changed (see also chapter 6).

2. Linkage equilibrium is also assumed; that is, that genotype frequencies at one locus are independent of those at a second. In reality this is not the case, some alleles on the same chromosome are frequently inherited together, i.e. they are in

linkage disequilibrium. This occurs when the allele is so close to the disease allele on the chromosome that there has not been a recombination between them to separate the inheritance of the two. The stronger the linkage disequilibrium the more likely the two loci are linked (Terwilliger and Goring, 2000).

3. Random mating is assumed. Non-random mating, the mating of two phenotypically similar individuals, (chapter 2) will alter the genotype frequencies.

4. No epistasis, i.e. no interaction between alleles at different loci is assumed.

1.2.3.2. Methods of linkage analysis.

There are many different methods involved in genetic analysis. For linkage analysis those methods can be classified into two main groups, parametric and non-parametric linkage analysis. The preferred method depends on a number of factors, including those based on the definition of the phenotype, the ability to define parameters associated with the phenotype and the structure of the pedigrees collected or to be recruited.

1.2.3.2.1. Parametric Linkage Analysis.

Parametric linkage analysis requires a model to be described, so defining the relationship between the phenotype and the genotype. Pedigree and genotypic data are then fitted to it. In general, the model requires the specification of the following parameters:

- disease gene frequency,
- the mode of inheritance, for example, dominant or recessive,
- the disease penetrance: the probability of an individual being affected with the disease given the genotype,
- the phenocopy rate: the probability of being affected by the disease but having an alternative genotype,
- marker allele frequencies and
- the rate of mutation.

The actual observed data, i.e. the genotypes at a specific locus within a pedigree, can then be compared to the model and the likelihood of linkage is calculated, assuming the model is correct. The process relies on a low level of recombination between the marker and the trait loci. It also assumes that the locus has a major effect on the phenotype and that a defined Mendelian pattern of inheritance is segregating within families (Camp and Cox, 2002).

As described above, the model used for the analysis is chosen based on a number of factors including the disease allele frequency, the transmission probability and the relationship between the unobserved genotypes and phenotype (Camp and Cox, 2002). Providing the model specified is correct, this method is more powerful than non parametric analysis, resulting in a lower type I error rate. It will also provide an estimate of the recombination fraction, therefore information about the location of the disease locus. Mis-specification of the model, however, will lead to a loss in statistical power to detect linkage. Specifically, mis-specifying the disease mode of inheritance will have a large impact on the LOD score as will incorrect marker allele frequencies especially when a lot of genotypic data is missing. Mis-specifying the penetrance and disease allele frequency has little effect on the LOD score (Haines and Pericak-Vance, 1998).

Although parametric analysis is extremely powerful when correctly specifying the model, the method is sensitive to bilineality. When the "affected" phenotype occurs in the family in both parents, or in two or more founders of a pedigree, different genes may act to produce the same phenotype therefore confounding the analysis. This arouses concern because the analysis occurs within pedigrees, not across pedigrees, therefore not allowing for this potential heterogeneity.

In general, the chromosomal area linked to the trait may be mapped efficiently, but the localized interval will usually be greater than 1 Mb wide, relying on recombination events having occurred within the pedigree. Therefore further analysis of additional pedigrees is usually required to refine the interval. However, as described above, with the correct model specified, the tests are powerful and can provide an accurate significance level. There are also many well developed marker sets for this purpose.

1.2.3.2.2. Non parametric linkage analysis

The alternative to parametric linkage analysis is non parametric linkage analysis which does not require the specification of model parameters. It is based on examining which alleles or sections of chromosomes are shared by affected individuals within pedigrees. The concept behind this being that for a genetically controlled trait, relatives that closely resemble each other phenotypically, will resemble each other genotypically. By sharing alleles or sections of chromosomes more often than expected by chance, it is possible that these regions will have some effect on the phenotype. The shared alleles can be classed as identical by descent (IBD), where the alleles are inherited from a common ancestor, or identity by state (IBS), where the alleles are the same but have not necessarily been transmitted from a common ancestor.

Affected sib pair analysis uses families with at least two affected siblings and examines the number of alleles shared IBD. By chance they will share 0, 1 or 2 alleles at a locus 25, 50 and 25% of the time respectively.

For example, in figure 1.5., sibling 1 has the genotype A C. Considering the parental genotypes, sibling 2 has four possible genotypes, one which shares both alleles and is therefore genotypically identical to sibling 1, two out of the four possible genotypes (i.e. 50% of the time) show the siblings having only one allele in common. The final possible genotype shows the siblings sharing no alleles, sibling 2 inheriting the two alternative alleles from the parents. If the proportion of shared alleles is significantly larger or smaller than the 25, 50 and 25% proportions in affected sib pairs, it indicates that a gene at that locus is linked to the trait. This approach has been used in many genetic analyses including the investigation of insulin dependent diabetes mellitus (Davies et al., 1994) as a dichotomous trait. Success for QTL mapping using this method is not as common. Other affected relative methods can be used: the likelihood of excess sharing between more distant affected family members by chance is decreased and therefore, if excess sharing does occur, it is more likely to be the cause of the phenotype.

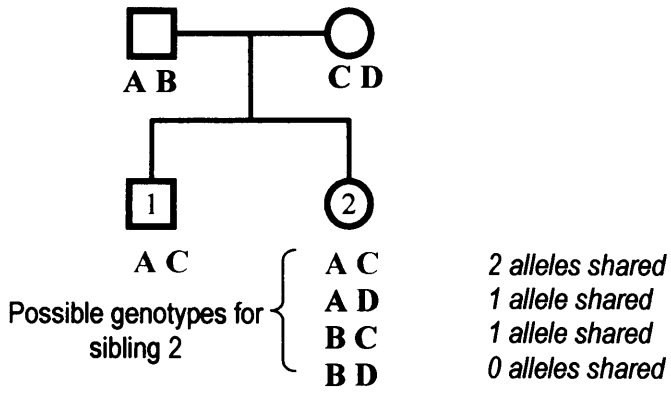


Figure 1.5. Allele sharing, adapted from (Strachan and Read, 1999)

Variance components methods provide a mechanism for partitioning observed phenotypic variance into component parts (see also chapters 4 and 5) and is particularly successful for the analysis of quantitative traits. IBD probabilities, i.e. the probabilities that alleles are shared IBD between relatives, are calculated from marker data of a pedigree. In a similar way to parametric LOD score calculations, two alternative models are then fitted, the first where the major gene effect is hypothesised to be 0 (i.e. null hypothesis of no linkage to that locus) and the alternative model where an estimation of the size of the gene effect is made (i.e. alternative hypothesis of linkage). The two models are fitted at regular intervals along the chromosome and compared so that the log likelihood ratio is calculated at each point. The most likely position of the quantitative trait locus can then be found based on the significance of the LOD scores (Camp and Cox, 2002).

1.2.4. Association

Association investigates the over-representation or under-representation of a specific allele in diseased individuals compared to control subjects (Baur and Knapp, 1997) (chapter 6). Association tests are commonly used after linkage analysis has identified a region harbouring a possible disease gene. The probability of causal action by a candidate gene within the linked region is investigated by studying the association

between the frequency of alleles at polymorphic loci and the presence or absence of the phenotype.

Two major strategies can be used in association studies. In case-control studies, the allele frequencies of a population affected with the trait under investigation are compared to those of a matched control population. In family based tests the transmission of parental alleles to affected offspring is compared to those alleles not transmitted. The preferential transmission of an allele to the affected offspring is suggestive of association.

1.2.5. Heterogeneity

It is possible that several different loci control a particular phenotype, known as locus heterogeneity. For example, one locus could control a trait in a particular population but in another population, that locus may be unrelated to the trait and an alternative locus is responsible for that phenotype. This could be due to differences in the genetic background or environmental conditions between populations. Locus heterogeneity can confound genetic analyses especially in the case of bilineality.

To include possible locus heterogeneity into the analysis, heterogenous LOD scores (hLOD) can be implemented which test the hypothesis of allowing a certain proportion of pedigrees within the study population to be linked. This proportion, α, is varied, producing a range of possible LOD scores. This increases the power of the study as it attempts to eliminate the possibility of, for example, excluding a locus which may be linked in some of the families and not in the others. This is because some of the families may produce a high LOD score and those not linked will have a negative value. The summation of these values may decrease the LOD score so that it is below the significant value.

In contrast to locus heterogeneity, in allelic heterogeneity two different alleles at the same point cause the same phenotype. Linkage analysis is insensitive to this

phenomenon but it can have a major influence in association studies, generally causing a reduction in power.

1.2.6. Experimental Design

There are a number of factors to consider when designing a genetic study: the definition of the phenotype under investigation, its likely genetic basis and the overall power of the study including consideration of the number and structure of pedigrees to collect.

1.2.6.1. Phenotype classification

The phenotype is the observed condition of a subject. For example, if the trait under observation is a disease, the phenotype would be either the extent of that disease, or simply if that individual has the disease or not. Therefore, there are two different ways in which a phenotype can be described: as a dichotomous (binary) value, or as a quantitative measure. In binary traits there are only two options and so the phenotype is either "affected" or "not affected". However, continuous traits such as refractive error may provide the investigator with more information. A quantitative measure potentially provides a greater power, especially if the genotype-to-phenotype relationship is direct.

1.2.6.1.1. Quantitative and dichotomous traits

As mentioned above, a dichotomous trait is one in which there are only two possibilities for the phenotype. This affectation status defines whether the individual is affected or unaffected with the disease. If the measured trait is quantitative, a threshold approach can be applied (figure 1.6.). A certain value, x, is defined as the threshold value, any value above this stated phenotype is classified as affected, below this as unaffected. This is the case in some studies of high myopia. For example,

taking a refractive error of -6.00D as the threshold value and classifying those individuals more myopic than this as affected. Although this approach can be successful, it has limitations. For example, it assumes that those individuals who have a refractive error of -5.75D are unaffected and therefore in the same group as those who are hyperopic.

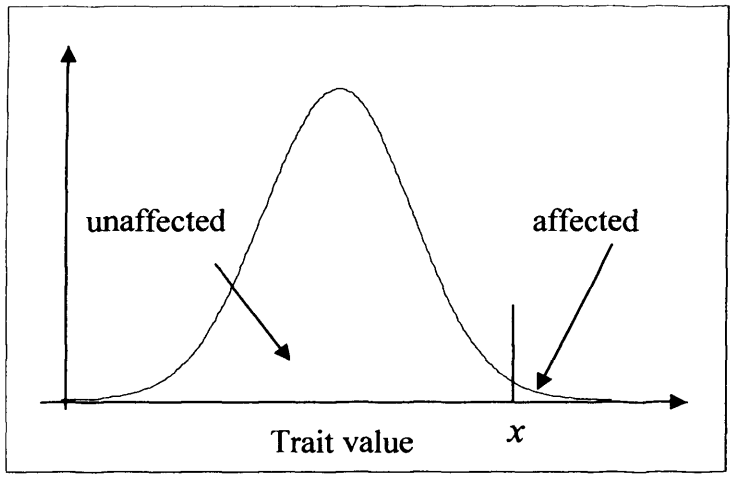


Figure 1.6. Applying an affection threshold to a quantitative trait. Any individual with a trait value falling in the area defined by that threshold value x and the area under the curve is classed as affected.

Some clinical phenotypes have a dichotomous basis; however they may have a strongly related quantitative risk factor. Although this may not describe the disease accurately, it may provide more information by quantifying it. If the disease under investigation is age related, it may allow earlier identification of the proband, even allowing for parental information to be included. However, this would require a direct genotype-to-phenotype relationship and would be limited to specific diseases.

Quantitative traits have a continuous distribution and suggest polygenic inheritance. Such traits could be controlled by a few quantitative trait loci each with a large effect or more QTLs each with a small effect. The individual genes will generally act in an additive manner, but there may be interactions between them, known as epistasis.

Clinical phenotypes are often in the form of quantitative measures and so give a more informative measure of the trait. For example, the additive effects of a number of "high risk alleles" may increase the quantitative trait value. For a dichotomous measure, it may require a certain number of these alleles to push the trait value above the affection threshold mark.

1.2.6.2. Heritability

Heritability estimates will give an indication of phenotypic variation due to genetic factors (chapter 3). The lower the heritability value, the less genetic factors will affect the trait and therefore the larger the size of the cohort of subjects needed to detect linkage.

1.2.6.3. Power

The power of a linkage study is its likelihood to detect linkage, given that it is present. A type I error occurs when the null hypothesis is incorrectly rejected. A type II error occurs when a false null hypothesis is not rejected. Power is therefore defined as $(1-\beta)$ where β is the type II error rate. Power depends on a number of factors including the required significance level, the specification of the alternative hypothesis and the sample size.

Based on these factors and the assumptions made about the mode of inheritance and the model's parameters, power calculations can be performed to calculate the sample size required to produce power to detect linkage at the required level. When inheritance patterns are known, power calculations are generally reliable. However, the results can be misleading if the mode of inheritance is specified incorrectly. This is particularly true for complex diseases, where assumptions about locus heterogeneity, phenocopy rate, recombination between marker and disease locus and marker allele frequencies have to be made.

For complex diseases, an indication of the power to detect disease loci can be obtained by considering the relative risk to relatives, λ (equation 1.4.). This describes the genetic contribution to a trait. For siblings, the λ value (λ_s) is calculated as follows:

$$\lambda_s = \frac{\text{trait risk for siblings of an affected individual}}{\text{trait prevalence of the general population}}$$

Equation 1.4. Relative risk to siblings.

In general, the higher the λ_s value, the greater the power to detect disease-carrying loci. Unfortunately, even diseases with a high λ_s may not be amenable to linkage analysis if numerous susceptibility genes contribute to the phenotype. Each locus (i) will import its own risk (λ_{si}) with their sum ($\sum\lambda_{si}$) being equal to the overall λ_s, assuming additive effects.

1.2.6.4. Populations and Pedigree Structure

Pedigree structure also plays a part in the sample size calculations. The power of a study is increased if the sample population is enriched with affected individuals. This is why the recruitment of subjects for a linkage analysis study targets affected probands and their families.

Many linkage analysis studies concentrate on the affected sib pair design, the methodology of which is based on the ability to identify those alleles inherited IBD compared to that expected by chance (section 1.2.3.2.2.). This is partly due to the relative ease of ascertainment of this pedigree type; often small nuclear families are more in abundance than larger extended families. The ideal type of pedigree to ascertain depends on the underlying genetic basis of the disease. For example, for a disease controlled by a single gene, extended pedigrees may be more useful but for a

more complex disease, sib pairs are often more efficient (Haines and Pericak-Vance, 1998).

If the study is using an association study design, alternative pedigree structures are required (chapter 6) dependent on the exact strategy to be used. A case-control study, comparing the allele frequencies of affected individuals to a matched population of controls, requires only individual subjects. A family based design requires a population enriched for the trait, generally small families with at least one affected offspring. In some strategies families with or without parents and including additional siblings, affected or not, are collected.

Due to the possibility of heterogeneity, mentioned previously, ideally the subjects recruited for the study would be of similar genetic background. This would reduce the possibility of locus and allelic heterogeneity. Bilinear families would also be best avoided in case the disease susceptibility loci are different in the two sides of the family although this may reduce the sample size and therefore the power. This is, however, difficult to detect unless it occurs directly from members of the phenotyped pedigree. It is entirely possible that bilineality goes undetected if the phenotypes of extended family members are not collected.

1.3. Myopia and Linkage Analysis.

Despite the ongoing nature versus nurture debates for myopia, as well as associations with near work and environmental factors, linkage studies have found a number of loci linked to refractive error and myopia (summarised in Table 1.2. up until summer 2006).

The first locus, MYP1, was mapped as Bornholm Eye Disease to Xq28 in 1990 (Schwartz et al., 1990). Analysis of an additional family resulted in a narrowing of this region from 34.4cM to 6.8cM (Young et al., 2004a). However, differing colour vision anomalies between the two families suggests separate mutations for the observed phenotypes.

The MYP2 locus was described in 1998 following a genome scan in four multigenerational families (Young et al., 1998b). By defining high myopia as a dichotomous trait, the cut-off point being -6.00D in the most minus meridian (MMM) of the right eye, and using a model based method of linkage analysis, an area of linkage was highlighted within 18p. By fine mapping an additional four families and using the same model of autosomally dominant inherited high myopia (susceptibility gene frequency 0.0133, 100% penetrance) a LOD score of 8.30 (θ=0.001) at a marker D18S63 was produced, suggesting strong evidence of linkage. The critical interval of 7.6cM at the location 18p11.31 was between markers D18S59 and D18S1138. This was later refined to a much smaller interval of only 0.8cM, flanked by markers D18S63 and D18S52 (Young et al., 2001a).

Following on from this study, MYP3 was mapped by the same laboratory using the same analysis method but in one large Italian-German family (Young et al., 1998a). The critical interval of 30.1cM was found between markers D12S1684 and D12S1603.

Twenty-one French and two Algerian families, each containing at least one high myope, were recruited by Naiglin and colleagues in 2002 (Naiglin et al., 2002). Negative LOD scores were found for the loci on 12q and 18p therefore, excluding

linkage to these regions and indicating the heterogeneity of this disease. A genome scan provided suggestive evidence of linkage on chromosome seven from marker D7S798 to D7S2423 with a multipoint LOD score of 2.81 (two-point LOD score of 1.87 at θ=0). A recombination event aided the narrowing of this interval to 11.7cM, from D7S798 to the telomeric end of the chromosome. This analysis was based on an autosomal dominant model of inheritance (susceptibility gene frequency 0.013, 58.4% penetrance) and was named MYP4.

A large English-Canadian family was analysed by Paluru and colleagues (2003) and the MYP2-4 loci were excluded. A linkage peak with a LOD score of 3 was found, however, on chromosome 17 (Paluru et al., 2003). The critical interval was narrowed to 7.71cM between markers D17S787 and D17S1811 and named MYP5.

A study of an isolated population of Ashkenazi Jews led to the linkage of a locus on chromosome 22 to common myopia (Stambolian et al., 2004) (MYP6). Forty families were phenotyped for common myopia. Those subjects with a refractive error of -1.00D or more were classed as affected and those with myopia of less than this, emmetropes and hypermetropes were classed as unaffected, with some allowances for age also involved in the phenotypic classification. Linkage analysis was performed using a range of models based on autosomal dominant inheritance with a gene susceptibility frequency of 0.0133 (penetrances 90, 80 and 58.4%, phenocopy rates 0, 5, 10 and 15%). The 12q and 18p loci were excluded for linkage, but the genome scan produced a linkage peak localised to 22q12 (hLOD=3.56, NPL=4.62). The same cohort also produced a high linkage peak from a genome scan when considering the trait quantitatively, on chromosome 1 (LOD=9.5, p=0.065 for ocular refraction and LOD=8.7, p<0.005 for log transformed refraction) (Wojciechowski et al., 2006). This QTL for refractive error lies in an 11Mb region flanked by markers D1S552 and D1S1622.

MYP7-10 were found in a twin based study (Hammond et al., 2004). A cohort of unselected twin pairs underwent a genome scan using refractive error as a continuous, quantitative trait. This indicated four areas which had evidence of linkage: 11p13 (LOD=6.1, MYP7), 3q26 (LOD=3.7, MYP8), 4q12 (LOD=3.3, MYP9) and 8p23 (LOD=4.1, MYP10). It was noted that the gene PAX6, having a major contributory

role in eye development, is mapped to within the 11p13 region. An association study was then performed with SNPs adjacent to the PAX6 gene but, despite the strong linkage signal, no association was found.

More recently, another new locus has been found within a 2.22cM region at chromosome 2q37.1 (Paluru et al., 2005). This study involved a large pedigree consisting of 32 family members in four generations, 14 of whom were highly myopic. The family were based in the U.S.A. and were of Northern European extraction. Using a model of autosomally dominant inherited high myopia (myopia gene frequency of 0.01, 90% penetrance and 10% phenocopy rate) and phenotype classification -6.00D or more, RE MMM as affected, a genome scan with two-point linkage analysis highlighted three areas as providing suggestive evidence of linkage. Further multipoint analysis excluded two of the loci but supported the evidence for linkage with a maximum multipoint LOD score of 4.75 for the area 2q37. Haplotype analysis further refined the interval to between markers D2S1279 and D2S2205.

On chromosome four, a Chinese family was found to be linked to a 20.4cM region surrounding marker D4S1564 giving a two-point LOD score of 3.11 at $\theta=0$ (Zhang et al., 2005).

Table 1.2. Summary of mapped MYP loci.

Name	Chromosome and locus	Trait used	Population mapped in
MYP1	Xq28	High myopia	1 MG Danish pedigree
MYP2	18p11.31	High myopia	4 MG pedigrees
MYP3	12q21-q23	High myopia	1 MG Italian-German pedigree
MYP4	7q36	High myopia	21 French & 1 Algerian pedigree
MYP5	17q21-q22	High myopia	1 large English-Canadian pedigree
MYP6	22q12	Common myopia	40 MG Ashkenazi Jewish pedigrees
MYP7	11p13	Refractive error (quant)	Unselected twin pairs
MYP8	3q26	Refractive error (quant)	Unselected twin pairs
MYP9	4q12	Refractive error (quant)	Unselected twin pairs
MYP10	8p23	Refractive error (quant)	Unselected twin pairs
MYP11	4q22-q27	High myopia	1 MG Chinese pedigree

MG= multigenerational, quant = quantitative value

Considering the loci found to date, this provides strong evidence of major heterogeneity of this disease. Farbrother and colleagues attempted to replicate linkage to MYP2, 3 and 4 (Farbrother et al., 2004b). They found in excess of 25% of the families in their cohort were linked to MYP3 therefore further suggesting locus heterogeneity. Ibay and colleagues also excluded MYP2 and MYP3 as being linked to common myopia using the Ashkenazi Jew cohort (Ibay et al., 2004).

As yet, no candidate genes have been found to be causative of high or common myopia within these regions. Several, including lumican, fibromodulin and TGIF-β have been excluded along with others (Paluru et al., 2004; Scavello et al., 2004; Scavello et al., 2005). To actually name a causative gene for myopia is proving difficult due to a number of factors. Firstly, the disease is known to be highly heterogenous. Once linkage is found, the candidate genes under test have to be picked accurately from those known genes within the locus. They may be chosen due to their biological function but this also relies on knowing that function and relating it to the trait. The disease causing polymorphism or a polymorphism in linkage disequilibrium also has to be identified and could be within sections of the gene or that of the promoter or another controlling area of the gene domain. Myopia is often described as a polygenic disease and therefore it may be that these candidate genes are providing only a small and therefore difficult to detect effect on the phenotype. As with other diseases, the quest for genes and QTL is ongoing and benefits from specific adaptations to study designs and analysis methods with regard to the exact trait characteristics.

2. Family Study of Myopia: Subject Recruitment

2.1. Introduction

The Family Study of Myopia was set up in 1999, aiming to recruit families for genetic analysis of high myopia. Initially, only families containing affected sibling pairs were sought, but this criteria was expanded to include trios, a highly myopic proband (the member of the family who is the primary contact) with two parents, and multigenerational families containing at least one high myope (MG families).

For the purposes of the study high myopia was defined as having a refractive error of at least -6.00D in both eyes in the least minus meridian. Taking the least minus meridian allows for the most conservative estimate of refractive error thereby ensuring a lower false positive error rate.

Probands also had to meet additional criteria. They were excluded if they were born prematurely as these subjects are more at risk of having significant refractive errors (Larsson et al., 2003; Snir et al., 2004). Any subject who has a systemic or ocular disease associated with myopia was excluded from the study; for example, Stickler's syndrome and Marfan's syndrome (Chapter 1). Any subjects with lens opacities were not excluded, but when obtaining phenotypic details, a pre-cataract spectacle prescription was specifically requested. As 80% of 75 to 85 year olds show some signs of age-related cataracts (Tunnacliffe, 1993), excluding those subjects with lens opacities would decrease the information about the pedigrees, the benefits of excluding any subjects with cataracts and therefore possible related myopia will be overshadowed by the major decrease in power shown by the reduction in pedigree size.

To reduce heterogeneity, as discussed in the previous chapter, the ideal cohort would be of the same ethnic background. At this stage of recruiting the ethnicity of a pedigree is not a limiting factor in the inclusion criteria because the aim is to obtain a

large number of highly myopic families and to build a database of willing participants for possible future analysis.

Family studies can use varying methods to ascertain their cohort of families. These include referrals from support groups and organisations or health care professionals, or through public databases and the internet (Haines and Pericak-Vance, 1998). Referral from specialists will ensure the participants are within the clinically defined study criteria but rely heavily on the participation of the professional. Using an advertising campaign may increase the response rate but may also include a small number of those subjects outside the criteria. Considering these advantages and disadvantages, the subjects for the Family Study of Myopia were recruited in a number of ways (section 2.2.2.).

It is important to recruit families with specific structures for genetic analysis. Specifically, the type of pedigree to ascertain is based on previous knowledge of the disease inheritance and the phenotype. For example, if it is suspected that one major gene controls the phenotype, extended multigenerational families are more powerful for linkage whereas affected sibling pairs may be more advantageous for a highly complex, polygenic disease (Haines and Pericak-Vance, 1998).

Other myopia studies have varying approaches to their pedigree ascertainment. Selecting for two affected siblings gives an advantage of using allele sharing linkage analysis methods, but it can be difficult to recruit the sample required. Collecting a number of large multigenerational families with multiple affected family members allows for various other methods of linkage analysis but the higher the number of affected family members and the more generations included will further increase the power. MYP1, MYP3 and MYP5 were discovered by linkage analysis using only one large multigenerational family (Schwartz et al., 1990; Young et al., 1998a; Young et al., 2001b). MYP2 used a cohort of four multigenerational families (Young et al., 1998b) whereas MYP4 and MYP6 had much larger cohorts consisting of 24 and 40 smaller families, respectively, containing one or more affected subjects (Naiglin et al., 2002; Stambolian et al., 2004). MYP7-10 were discovered based on a population sample of twins (Hammond et al., 2004). There are a number of advantages and disadvantages to recruiting the various types of families which is discussed further in

Chapter 5. The ascertainment criteria were extended to any family type in this study so that a database of myopia pedigrees in the British Isles can be maintained and subjects included for future analyses.

2.1.1. Assortative Mating

Random mating within a population is generally assumed and is essential for the validity of many types of genetic analysis. Assortative mating suggests that more often than expected, mating occurs with like-for-like individuals. Assortative mating can occur in several ways and can be influenced by factors such as location, familial expectations and social implications. One type of assortative mating occurs between relatives. Consanguineous matings may influence the incidence of inherited disease based on the fact that relatives will already share a proportion of their genes. Another type occurs within subgroups of populations e.g. ethnic groups. This will, eventually, increase the genetic differences between these subgroups (Vogel and Motulsky, 1996).

In this study, the information required to test the presence of assortative mating is how often is the spouse of a high myope also a high myope and does this occur more often than expected based on the phenotype prevalence. If it does, this situation within a population selected for genetic analysis will increase the heterogeneity of the sample particularly within a specific pedigree (bilineality) therefore decreasing the power to detect linkage.

2.2. Methods

2.2.1. Ethical Approval

Ethical approval was sought for and gained from the Human Science Ethical Committee, Cardiff University.

2.2.2. Recruiting

2.2.2.1. Recruiting via Optometrists

Families for the Family Study of Myopia were ascertained by participating optometrists who identified high myopes willing to join the study. Probands were given an information sheet and a questionnaire, PQ1, (Appendix I) which asked them to indicate the name and address of their optometrist, the age they began to wear spectacles, any ocular surgery or treatment and any systemic conditions. Once a high myope had been contacted and their consent form and questionnaire returned, their optometrist was contacted again to obtain the spectacle prescription and ocular history of that subject. Other family members were invited to participate by sending them a "PQ2" questionnaire (Appendix I). Family members' spectacle prescriptions and ocular history details were obtained in the same way. Once a suitable family had been identified, mouthwashes were requested from each individual.

When optometrists mentioned on a returned PQ1 or PQ2 form were contacted to request the patient's details (forms *OP1* and *OP2*, Appendix I), they were also sent a recruitment pack and information on the study to pass on to any other highly myopic patient they might see, ensuring an accumulation of optometrists recruiting patients. They were also sent a Family Study of Myopia occluder, for use in their day-to-day clinical practice, with the contact details and recruitment criteria on it to serve as a constant reminder. Information packs were also distributed at local optical continuing education meetings.

An optometrist based in Carmarthen approached the Family Study of Myopia with information about his own family, who are mainly living in Ireland. Approximately 190 members of this family were contacted (PQIW, Appendix II) and asked to provide details of spectacle wear and, again, details of their optometrists, from whom we could obtain ocular history details. A "family get together" was arranged for this family to obtain further information, ocular measurements and DNA samples (chapter 3).

2.2.2.2. Recruiting by Databases

A large proportion of the highly myopic probands recruited were included in a database of patients who attended a group of optometric practices based in the North West of England and Scotland, Conlon & Sons (Opticians) Ltd. These individuals were contacted directly with a PQ1, consent form and covering letter explaining the research project (Appendix I). Their relatives were contacted with PQ2s as mentioned above. An alternative database containing those patients seen at the Cardiff University Eye Clinic was also utilised.

2.2.2.3. Recruiting by the Internet

The Family Study of Myopia has its own website containing information about the study and also an email address and contact details through which interested parties can request information or offer to participate.

2.2.2.4. Recruiting by Word of Mouth

The final way probands were recruited was simply by word of mouth. These probands included, for example, those who were optometry students within the university and high myopes known to one of the research study group.

2.2.3. Assortative Mating

To investigate assortative mating, all probands who were not recruited specifically for a family history of myopia were encouraged to provide contact details of their spouse. The spouse was also contacted and their spectacle prescription and ocular history details were collected from their own optometrist.

The sample included only those not specifically recruited for a family history of myopia to avoid any bias and provide an appropriate random sample representative of the general population.

2.3. Results

The recruiting process is summarised in figure 2.1.

2.3.1. Proband Recruitment

Between October 2002 and October 2005, a total of 30 families with affected sib pairs, 55 families containing at least one high myope and 98 trios were identified. A total of 1968 PQ1s were sent out and 250 were returned. The response rates ranged from 0 to 45% per month, averaging 12.7% (figures 2.2 and 2.3). Two different methods were tested for recruiting high myopes through the database. For most subjects in the database, a questionnaire, consent form, information sheet and freepost return envelope were sent and follow up questionnaires were sent to family members if consent was given. A sample of 200 subjects were sent a variation on this theme, comprising a shortened questionnaire, a card with details about the study and a tick box section to return if they would like to take part or just require further information. Those subjects who returned the card were then sent a full questionnaire, as previously, by their chosen method of communication (PQ1C, Appendix I). This approach ensures only interested parties would receive the information and increase the response rate from the questionnaires. From the 200 cards sent, 17 replies were received, (figure 2.3, highlighted region). The 8.5% response rate was not significantly different from the previous method. Furthermore, due to either a lack of response from these probands' families or from those probands themselves, only one suitable family was ascertained using this approach. Therefore the success rate was 5.9%. This compares to the original method (a mixture of sending out questionnaires to people in the database and recruiting by optometrist, where there were 101 PQ1s leading to a family useful for genetic analysis) in which a success rate of 5.1% was estimated. Due to the benefits in terms of time-investment, the original method was used.

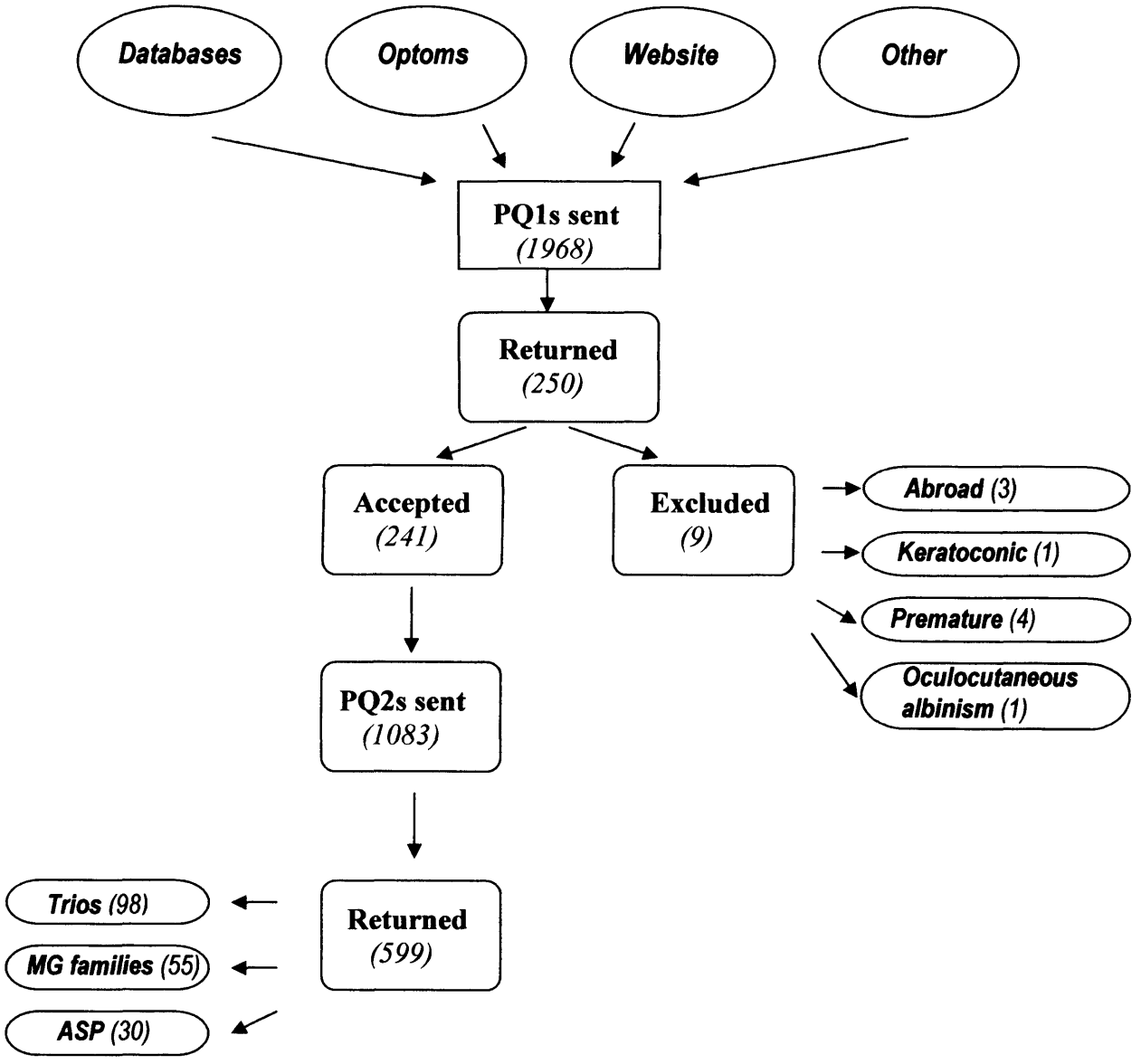


Figure 2.1. Flow chart summarising the recruiting process. A small proportion of the responding probands were excluded due to clinical factors.

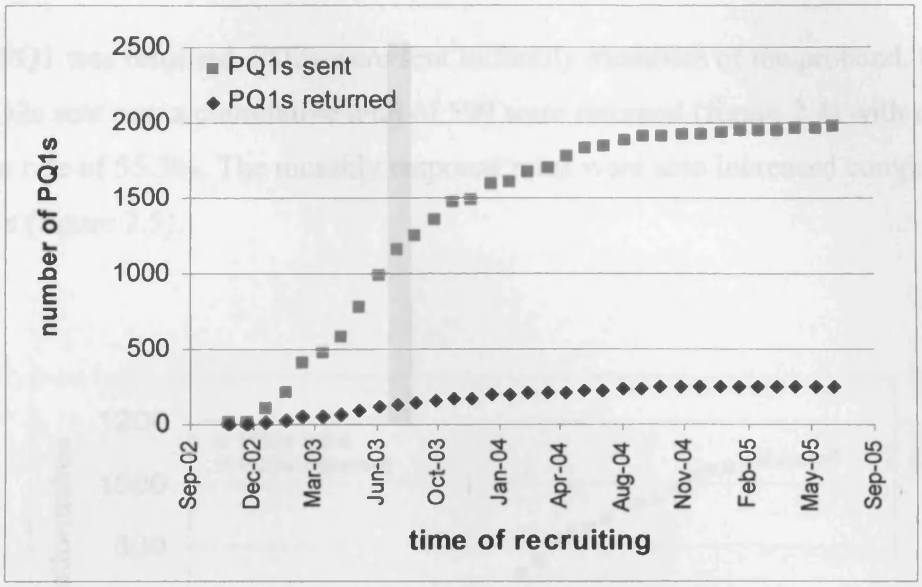


Figure 2.2. Number of first contact questionnaires (PQ1) sent out and returned between October 2002 and October 2005. The highlighted region represents the time period in which the alternative method of using a shortened questionnaire and card was used in an attempt to increase the response rate.

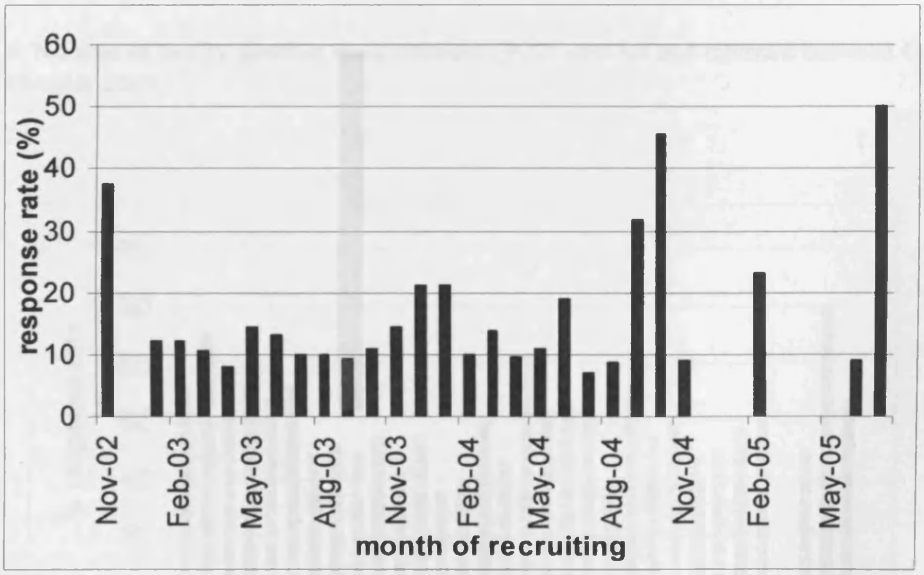


Figure 2.3. Response rate for PQ1s from October 2002 to October 2005. Highlighted region as described above.

2.3.2. Family Recruitment

Once a PQ1 was returned, PQ2s were sent to family members of the proband. Out of 1083 PQ2s sent out, a cumulative total of 599 were returned (figure 2.4) with a response rate of 55.3%. The monthly response rates were also increased compared to the PQ1s (figure 2.5).

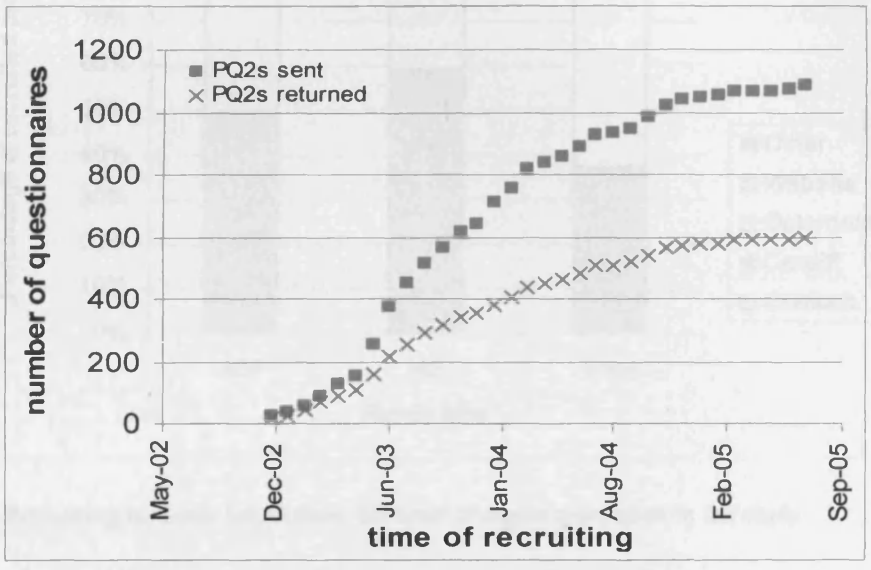


Figure 2.4. Number of family member questionnaires (PQ2) sent out and returned between October 2002 and October 2005.

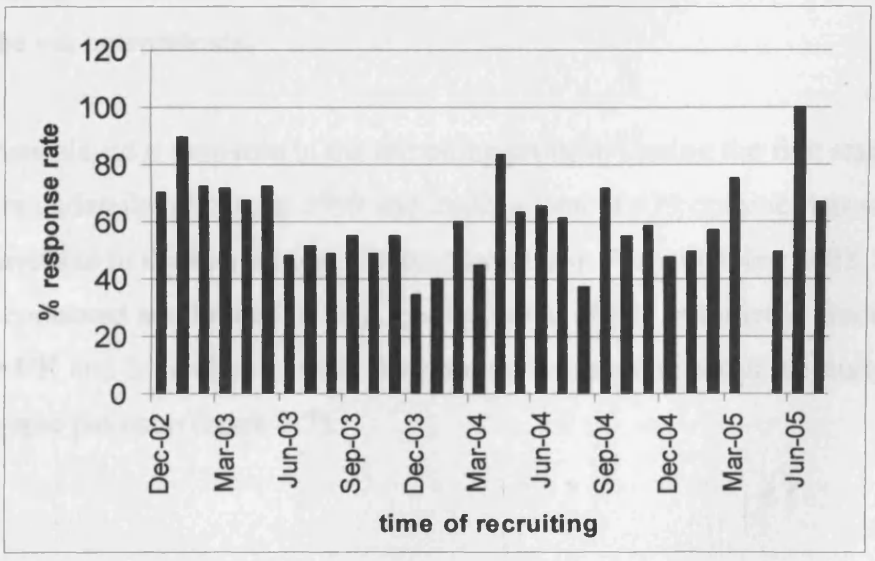


Figure 2.5. Response rate for PQ2s from October 2002 to October 2005.

2.3.3. Recruiting Methods

Probands were recruited by a variety of methods as shown in figure 2.6.

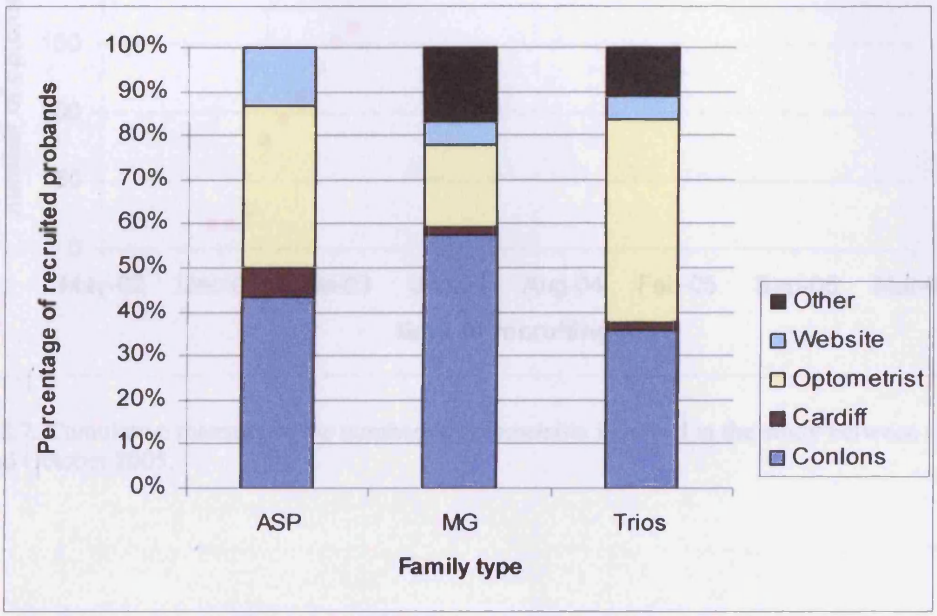


Figure 2.6. Recruiting methods for all three different pedigree types used in the study.

The largest proportion of affected sibling pairs and multigenerational families were recruited using the Conlons' database. The most successful method for recruiting trios seems to be via optometrists.

Optometrists played a vital role in the recruiting process. During the first stage of recruitment undertaken between 1999 and 2002, a total of 479 optometrists were actively involved in the recruitment (Farbrother, 2003). From October 2002, a further 323 were contacted and invited to help, giving a total of 802 optometric practices within the UK and Ireland potentially distributing information about the study to their highly myopic patients (figure 2.7).

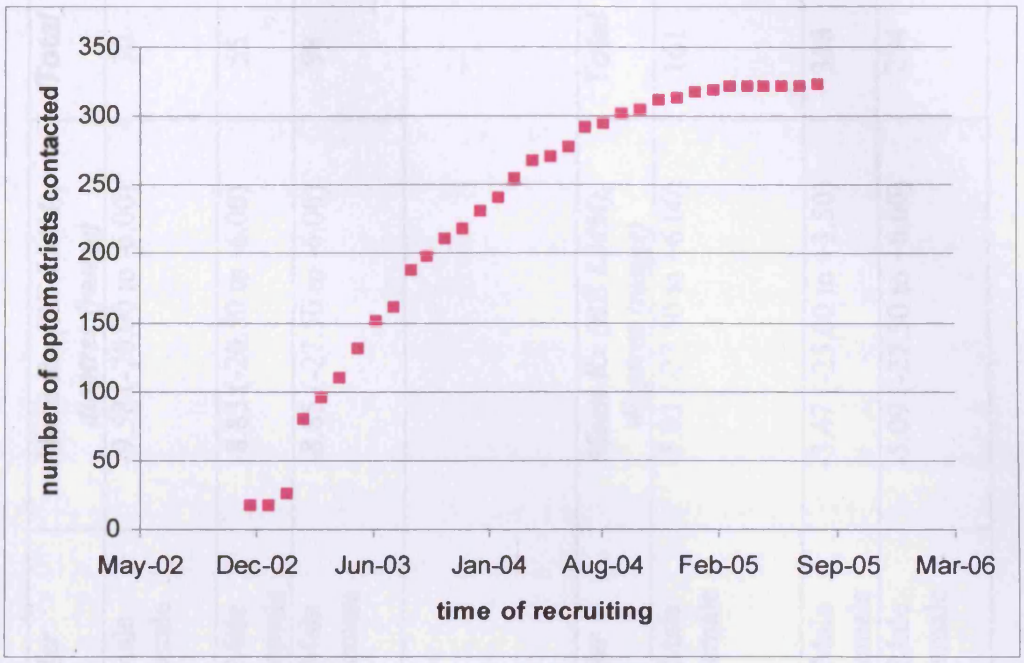


Figure 2.7. Cumulative measure of the number of optometrists involved in the study between October 2002 and October 2005.

2.3.4. Characteristics for probands and families.

Tables 2.1. and 2.2. show the characteristics of probands and family members of the three family types. Subjects were aged between 7 and 89, and were mostly of White European descent. More females than males participated. By investigating only those subjects who were recruited by the Conlon's database and comparing those who agreed to participate and those who did not, there appears to be supporting evidence of a gender bias towards females willing to participate (χ^2=11.49, p=7.01x10^{-4}). However, there is no significant difference in severity of myopia (Mann Whitney, p=0.01) nor with age (Mann Whitney, p=0.01).

By looking at the affected and unaffected members of the recruited families there was no obvious pattern of inheritance of myopia indicated.

Table 2.1. Characteristics of probands.

Type of family	*Mean age in yrs (range)*	*Ethnicity*	*Gender*	*Mean Rx (RE LMM), dioptres (range)*	*Total*
ASP	38 (14 to 71)	93.3% White European 3.3% Asian 3.3% Euro-Caribbean	30% Male 70% Female	-9.58 (-20.00 to -6.00)	30
MG	34 (14 to 67)	100% White European	23.6% Male 76.4% Female	-8.83 (-20.50 to -6.00)	55
Trios	29 (14 to 58)	98.0% White European 1% Asian 1% European-African	27.6% Male 72.4% Female	-8.85 (-27.50 to -6.00)	98

Table 2.2. Characteristics of all individuals recruited for the Family Study of Myopia

Type of family	*Mean age, yrs (range)*	*Ethnicity*	*Gender*	*Mean Rx (RE LMM), dioptres (range)*	*Total*
ASP	41 (7 to 80)	97.5% White European 1.2% Asian 0.65% Euro-Caribbean 0.65% Unknown	40.4% Male 59.6% Female	-5.03 (-27.50 to +6.00)	161
MG	49 (7 to 89)	98.4% White European 1.6% Unknown	41.8% Male 58.2% Female	-3.47 (-25.00 to +3.50)	318
Trios	42 (14 to 85)	98.0% White European 1% Asian 1% European-African	42.5% Male 57.5% Female	-5.09 (-27.50 to +6.00)	294

2.3.5. Trait distributions

Refractive error (RE LMM) distribution is not normally distributed. Considering the family members in the ASP and MG families, there is an apparent bimodal distribution (figure 2.8). This maximum around -6.00D is due to the ascertainment criteria. Removing the probands from this analysis reduces this peak, but not completely as the ascertainment criteria for the ASP families will increase the amount of high myopes (figure 2.9). The trait distribution of probands only (figure 2.10.) shows the majority of probands being within the lower end of the limits. Generally, refractive error is leptokurtosed at emmetropia and skewed towards myopia (Sorsby et al., 1957), this is also reflected in the distribution of the study subjects.

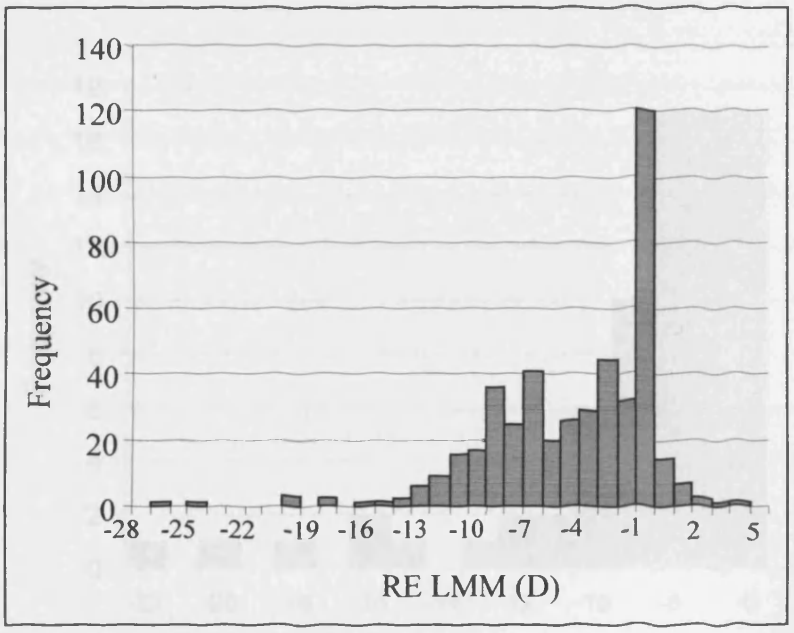


Figure 2.8. Trait distribution of right eye, least minus meridian (dioptres) for subjects within ASP and MG families.

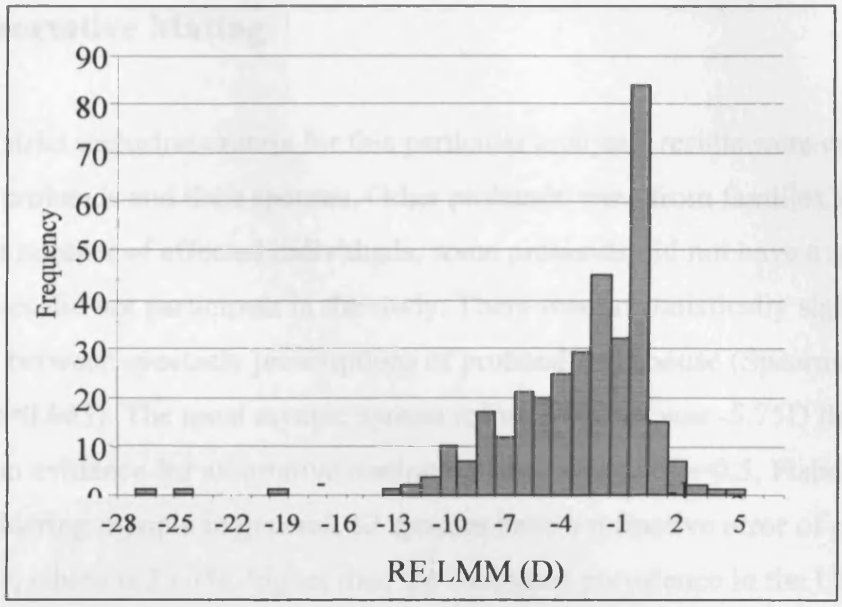


Figure 2.9. Trait distribution of right eye, least minus meridian (dioptres) for subjects within ASP and MG families, not including probands.

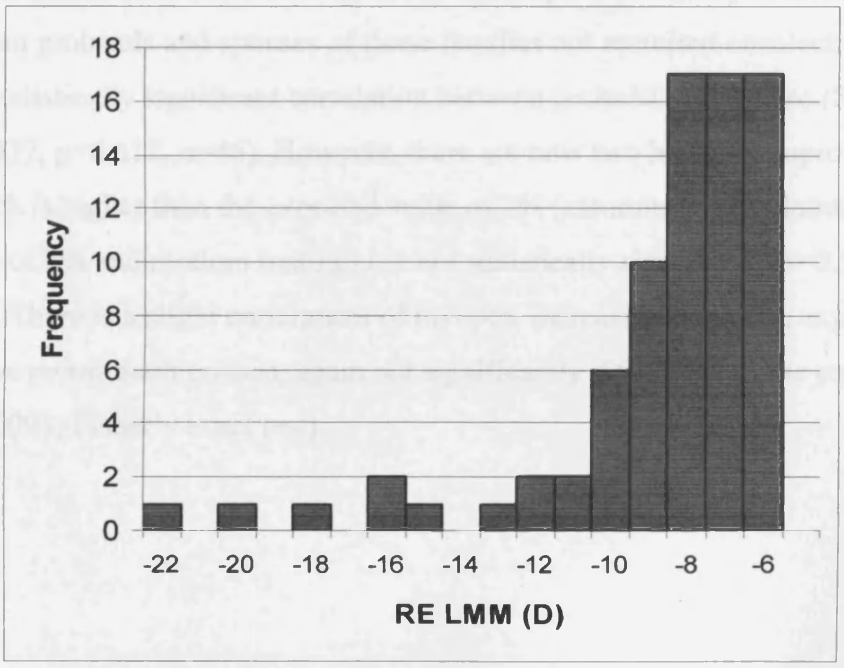


Figure 2.10. Trait distribution of right eye, least minus meridian (dioptres) for probands within ASP and MG families.

2.3.6. Assortative Mating

Due to the strict inclusion criteria for this particular analysis, results were collected for just 36 probands and their spouses. Other probands were from families known to have a high number of affected individuals, some probands did not have a spouse, and some spouses did not participate in the study. There was no statistically significant correlation between spectacle prescriptions of proband and spouse (Spearman's rank r=-0.075, p=0.663). The most myopic spouse refractive error was -5.75D therefore there was no evidence for assortative mating for high myopes (p=0.5, Fisher's exact test). Considering myopia in general, 12 spouses have a refractive error of greater than -0.50D, which is 33.3%, higher than the estimated prevalence in the UK population, but not significantly different (p=0.302, Fisher's exact test). This suggests that this population is not enriched for myopes due to the recruiting criteria.

If this investigation was extended to consider not only the randomly ascertained probands but probands and spouses of those families not recruited unselectively, there is still no statistically significant correlation between proband and spouse (Spearman's rank r=-0.077, p=0.612, n=46). However, there are now two highly myopic spouses, 4.3%, which is higher than the expected value of 2% (assuming a population prevalence of 2% and random mating) but not statistically significant (p=0.5, Fisher's exact test). There is a slight enrichment of myopes, increasing to 41.3% explained by the selective recruitment criteria, again not significantly difference to the expected value (p=0.093, Fisher's exact test).

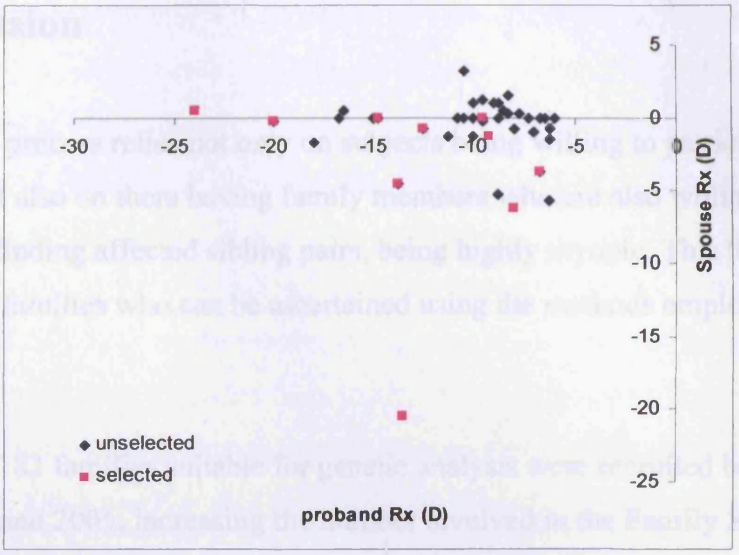


Figure 2.11. Proband spectacle prescriptions (Rx, RE LMM, dioptres) plotted against that of their spouses.

2.4. Discussion

The recruiting process relies not only on subjects being willing to participate themselves but also on them having family members who are also willing to help and, in the case of finding affected sibling pairs, being highly myopic. This therefore limits the number of families who can be ascertained using the methods employed in this study.

Nevertheless, 183 families suitable for genetic analysis were recruited between October 2002 and 2005, increasing the number involved in the Family Study of Myopia to over 240 (including all family types). The subjects for this project were mainly recruited via database-based recruiting and by optometrists. The subjects vary in age and had a gender split biased towards females. The majority of the families were of White European ethnicity which suggests the population will be highly homogenous, thereby increasing the power to detect linkage within the study. Including a mixture of ethnicities into the pedigrees to be analysed can increase the possible heterogeneity and therefore decrease the power to detect linkage (Ott, 1991). All of the trios were included into the association analysis (chapter 6) as there was only a small percentage of different ethnicities. If an effect was found it would be possible to repeat the analysis with sub-sections of different ethnicities or remove individuals from the analysis to create a homogenous population.

A study undertaken in Australia found that participation of probands seems to increase with the proband's perception of disease severity (Garoufalis et al., 2005). This was not found to be the case for these probands, only 25% of the sample representing extreme myopia (over -10.00D, RE LMM) compared to 56% in the Australian study, although these two studies did differ in study design and selection criteria. There did not appear to be any other significant pattern involving specific characteristics of families and their likelihood to participate although this analysis is limited as only specific information was obtained from those individuals who were only invited to take part and declined compared to those who agreed.

Optometrists appeared to play a vital role in the recruitment of probands and their families. They provide a valuable position in that they are qualified to recognise suitable candidates by including relevant questions into their detailed history taking along with having the specific inclusion and exclusion details in information sheets provided to them. This allows the numbers involved in the study to be massively increased compared to the situation in which one optometrist were to examine all patients individually with no decrease in the quality of the data collected.
Only nine probands were excluded from the study and only six of these were excluded due to clinical issues; the other three were excluded due to difficulties in obtaining phenotypes and DNA samples due to them living abroad. Out of those six probands, a premature birth was the main reason for exclusion. One person was keratoconic and one had oculocutaneous albinism which prevented them from being involved in this genetic study as they would increase the genetic heterogeneity of the sample and allow some possible confounding factors as to the basis of the myopia.

There is a wide range of refractive errors within the probands and it is possible that a high myope of -6.00D has a different clinical phenotype than one with a much higher refractive error of -22.00D. However, with the obtained clinical information from the optometrists any abnormalities, systemic or ocular, allow the exclusion of these individuals from the analysis. There is considerable current interest in sub-classifying complex disease phenotypes in order to improve linkage power which could be adapted to be used in the Family Study of Myopia (Hauser et al., 2004; Gordon et al., 2006).

There appears to be no evidence of assortative mating. However, the small number of spouses involved in the study suggests that the power to detect such an effect would have been low.

The following analyses did not make use of all of the families recruited within this time period but this recruitment process added to a large and unique database of pedigrees willing to participate with future genetic analysis projects investigating myopia within the UK and Irish population.

3. Heritability

3.1. Introduction

Whilst recruiting families to take part in the Family Study of Myopia, an optometrist based in Carmarthen approached the study to offer the participation of his own family, a large seven generational pedigree consisting of 294 contactable members within three generations (IW pedigree, Appendix III). Due to the size of the family and a reportedly high prevalence of myopia, they were considered an ideal cohort for linkage analysis. Initial efforts concentrated on calculating a heritability estimate in order to give an indication of the potential for mapping refractive error susceptibility loci responsible for the control and development of refractive error for this family.

3.1.1. Definition

Heritability is the proportion of observed phenotypic variance attributed to genetic variance (Camp and Cox, 2002). This broad sense heritability (H^2) therefore describes both additive and dominance effects on the phenotype and also any interactions. The broad sense heritability can be calculated using a twin study (section 3.1.2.). Narrow sense heritability, h^2, is more commonly used. The genotypic variance is partitioned into separate components of additive and dominance effects, the narrow sense heritability being the ratio of additive genetic variance to phenotypic variance in a specific population (equation 3.1.) (Falconer, 1996). It is a quantitative value representing the effect of genetic factors upon a trait and can be estimated from the degree of resemblance between relatives. Specifically, heritability is the observed correlation as a proportion of the correlation that would be found if the character were completely inherited (Falconer, 1996).

$h^2=V_A/V_P$. *where V_A is the additive genetic variance and V_P is the phenotypic variance*

Equation 3.1. Narrow sense heritability (h^2)

It is vital to be aware of the heritability of a trait before starting a genetic study. The estimate calculated may not only govern the choice of ascertainment scheme or study design but also reinforce the hypothesis of the likelihood of identifying a quantitative trait locus for that trait (Camp and Cox, 2002).

3.1.2. Twin Studies

An alternative method to investigating the degree of resemblance between relatives is a twin study, comparing the concordance of the trait in monozygotic twins to that of dizygotic twins, for example see figure 3.1. The increased correlation of phenotypic values from the monozygotic twins compared to the dizygotic twins suggests a highly genetic influence and therefore a high heritability value. Twins are ideal subjects for heritability studies as the genetic and, generally, the environmental factors of monozygotic twins are the same from conception onwards, whereas for dizygotic twins it is purely the genetic factors that will differ in most cases. However, this shared environment may inflate the heritability estimate; therefore the value must be regarded with these limitations in mind and will show the upper limit of the heritability. This is also true of sibling-sibling methods.

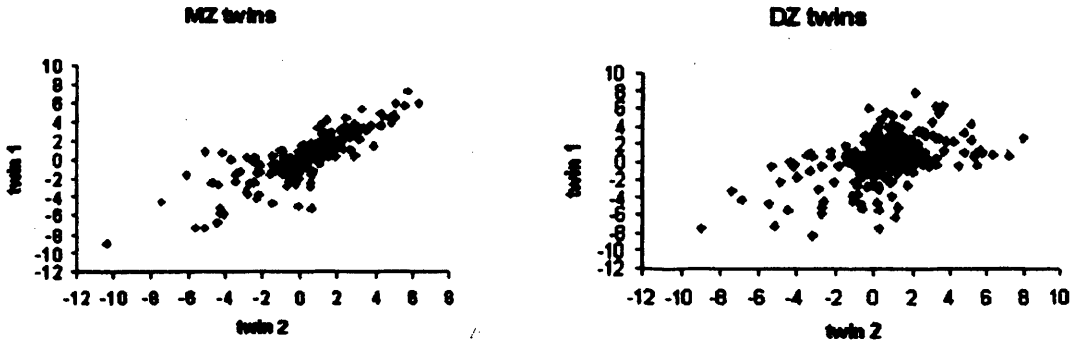


Figure 3.1. Taken from (Hammond et al., 2001) showing the values of spherical equivalent refractive error (D) of the left eye for twin 1 plotted against twin 2 for monozygotic (MZ) and dizygotic (DZ) twins.

For twin studies, when comparing the traits of monozygotic and dizygotic twins, the assumption of equal environmental effects is made and also that the genetic variance is equal in both types of twins. This may not always be the case due to factors including gene-environment interaction, competition between twins *in utero* and parental treatment (Falconer, 1996).

3.1.3. Assumptions

Heritability calculations rely on several assumptions. This estimate corresponds only to the population and conditions under investigation. It is assumed the population is random-mating. Assortative mating (chapter 2) as opposed to random mating, will inflate the heritability estimate by increasing the resemblance between spouses and therefore increasing the resemblance between family members. For midparent-offspring methods (see 3.1.4.1) it will also decrease the variance of the parents and so still provide a valid estimate.

It is assumed there is no epistasis, which is no interaction between two or more genes acting on the same trait. Also that there is no gene-environment interaction, i.e. the environmental factors have no influence on the gene effects. The final assumption is

that of no dominance. As narrow sense heritability is the phenotypic variance due to additive genetic effects, dominance must be assumed to be zero and so the offspring trait values should tend to the mean of the parental values.

The involvement of dominance and epistatic effects confound the parental genotype to phenotype relationship. The inclusion of these factors makes this relationship more difficult to explain, the relationship between parental and offspring phenotype is similarly affected. The calculation of narrow sense heritability using only additive genetic effects, as opposed to broad sense heritability which includes these dominance and epistasis effects, is therefore more useful in this prediction of offspring phenotype.

3.1.4. Methods of calculating heritability

3.1.4.1. Within Family Regression

Taking into account a definition involving the degree of resemblance between relatives, regression between relatives, either between offspring and parents or between siblings can provide heritability estimates (table 3.1).

Table 3.1. Calculation of heritability estimates (h^2) based on regression and correlation of the phenotypes of relatives (adapted from (Falconer, 1996)).

Relatives	Regression (b) or correlation (t)
Offspring and one parent	$b=\frac{1}{2}h^2$
Offspring and mid-parent	$b=h^2$ or $t=(\sqrt{0.5})h^2$
Half sibs	$t=\frac{1}{4}h^2$
Full sibs	$t\geq\frac{1}{2}h^2$

For example by plotting the mean phenotypic values of two parents against the phenotypic value of their offspring, the heritability is equal to the regression coefficient.

3.1.4.2. Variance Components Method

The SOLAR (Sequential Oligogenic Linkage Analysis Routines) (Almasy and Blangero, 1998) software package provides a heritability estimate by decomposing the variance of the phenotype into that caused by genetic effects and that by residual, presumably environmental factors. The genotypic variance can be further partitioned into additive and dominant gene effects, and again into the variance due to a major effect quantitative trait locus and that of combined smaller effect polygenic influences. It is the variance due to the additive effects that is considered for the heritability estimate. This involves taking into account all members of the pedigree, exploiting all informative relationships as well as those between parents and offspring, or between siblings.

3.1.4.3. Markov Chain Monte Carlo (Gibbs sampling) methods

A Markov chain is a sequence of random numbers, in which each following number is conditionally independent of the past sequence, given the current number. To construct a Markov chain that has the required distribution, sampling occurs from probability distributions using specific algorithms. This method is known as Markov Chain Monte Carlo method (MCMC). The probability distributions can be joined together forming one distribution by means of Gibbs sampling.

As an example, [taken from Sorensen and Gianola (Sorensen and Gianola, 2002)] (figure 3.2.), imagine there are three compartments numbered 1, 2 and 3 which are

interconnected and a mouse is placed into one of the compartments. The sequence of the compartments visited makes up the numbers in the Markov chain. The movements of the mouse are dependent only on its current position. The probability of the mouse moving into either of the other compartments, the transitional probability, can be calculated, which is conditional on it actually starting this move in the stated compartment. The distribution of the proportion of the number of times spent in each compartment is produced. After a large number of movements, or transitions, the distribution may stabilise and "converge", the evidence suggesting that the compartment that is visited the most times is the most likely state. For heritability estimates, each state or compartment is equivalent to a specific heritability estimate based on an assumed gene flow through the pedigree. The heritability estimate that is most likely will be revisited and therefore provide the modal value of the distribution when it has stabilised. Gibbs sampling samples from all of the fully conditional posterior distributions to produce a joint distribution.

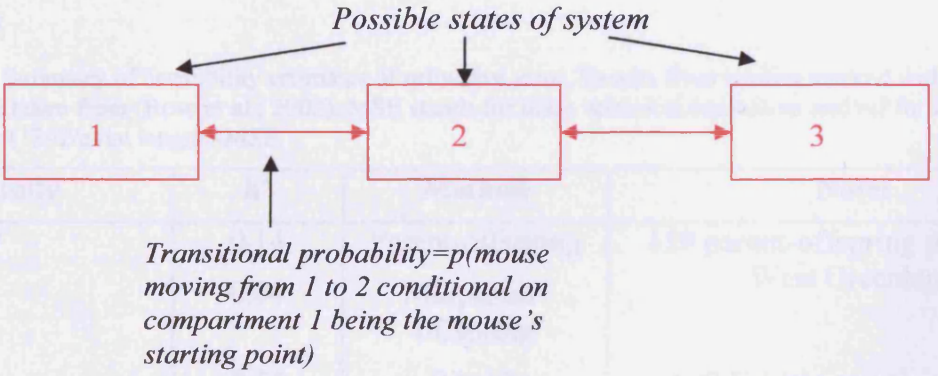


Figure 3.2. Pictorial representation of Markov Chain.

Two main disadvantages of this method are the time needed to run the analysis and the difficulty in knowing when the likely states have been sampled proportionally to their likelihood (E. Sobel, 2004 WTAC). Algorithms and transition rules can be set up using different software programmes to decrease the time needed for these

calculations but maintaining the correct sampling guidelines, i.e. proportional to their likelihood.

3.1.3. Heritability and Refractive Error

Studies which have investigated the heritability of refractive error vary from twin studies to family studies of indigenous populations. The estimates range from -0.03 to 0.98 (table 3.2.). The lowest values suggest little or no genetic influence on the trait whereas the higher values suggest a strong association with family history.

The higher values of heritability were calculated in twin studies. These will give an upper limit to the heritability estimates but also provide strong evidence of a major genetic component.

Table 3.2. Summary of heritability estimates of refractive error. Results from studies marked with an asterisk are taken from (Rose et al., 2002). MSE stands for mean spherical equivalent and ref for total refraction= (1392/axial length)-MSE.

Study	h^2	Method	Notes
Alsbirk*	0.14	Parent-offspring	159 parent-offspring pairs from West Greenland
	0.04	Midparent-offspring	
	0.50	Sib-sib	~600 Eskimo people aged 15+
(Angi et al., 1993)	0.11	Twin	19 MZ and 20 DZ twin pairs age 3-7 years
Ashton*	0.49	Midparent-offspring	Nuclear families of Japanese and European descent in Hawaii
	0.74	Sib-sib	
(Biino et al., 2005)	0.18 (MSE)	Midparent-offspring	201 nuclear families, age 4-90 years
	0.50 (ref)		
(Hammond et al.,	0.84-0.86	Twin	226 MZ and 280 DZ twin pairs

2001)			age 49-79 years
Hu*	0.61	Twin	49 MZ and 37 DZ twin pairs age 7-19 years
Johnson*	-0.03	Parent-offspring	Inuit and mixed Inuit-Caucasian population in Labrador
Kimura*	0.80	Twin	33 MZ and 16 DZ twin pairs aged 15-20 years
Lin & Chen*	0.25	Twin	90 MZ and 36 DZ twin pairs aged 7-23 years
(Lyhne et al., 2001)	0.89-0.94	Twin	53 MZ and 61 DZ twin pairs age 20-45 years
Nakajina et al*	0.83 0.73 0.16 0.42	Twin Twin Parent-offspring Parent-offspring	39 MZ and 10 DZ Japanese twin pairs age 12-17
Sorsby et al*	0.87 0.45 0.49 0.72	Twin Parent-Offspring Midparent-offspring Sib-sib	78 MZ and 40 DZ twins age 4-14 28 UK nuclear families
Teikari*	0.58	Twin	54 MZ and 55 DZ twin pairs age 30-31 years
(Wojciechowski et al., 2005)	0.61	Sib-sib	759 siblings in 241 families, mean age 73.4 years
(Young et al., 1969)	0.10 0.98	Parent-offspring Sib-sib	197 subjects in 41 Eskimo family units
(Young and Leary, 1972)	0.46	Midparent-offspring	71 Eskimo families

Despite the fact that heritability studies are difficult to compare due to the estimate being population specific, the previous studies suggest that a high degree of refractive error is determined by genetic factors, although not discounting effects from environmental influences. There is a large range of values of heritability estimates for refractive error and these will vary due to study design and sample population but the

main focus of this study is to regard the trend. As the study populations are different the estimates cannot be compared but an overall pattern may be shown.

3.1.4. Ocular Components and Heritability

The eye is made up of a number of components correlated to produce an accurate refracting system, see figure 1.2., including the vitreous chamber depth, radius of corneal curvature, anterior chamber depth and lens thickness. In an emmetropic eye all of the ocular components are correlated so producing a focal point coincident with the retina. A disruption to this correlation will cause the eye to be ametropic by focusing the light in front of or behind the retina (Sorsby et al., 1957).

These ocular biometric measurements can be calculated using standard clinical equipment, including the Zeiss IOLMaster (Carl Zeiss Meditec). This instrument works using partial coherence interferometry to measure the axial length of the eye. It incorporates a keratometer to enable measurement of the corneal curvature. Anterior chamber depth is estimated by measuring the distance between the reflected light from the optical sections of the lens and corneal surface using lateral slit illumination (Zeiss IOL Master manual website).

Table 3.3. shows heritability estimates calculated for ocular refractive components. Values range between 0.55 to 0.88 for corneal curvature and between 0.31 and 0.84 for axial length suggesting highly heritable traits. Anterior chamber depth has a larger and lower range between 0.05, suggesting genetic factors play a very minor role in the development of this refractive component in this population, to a higher estimate of 0.68.

Table 3.3. Heritability estimates of ocular refractive components. Studies marked with ** estimates taken from (Goss, Hampton and Wickham, 1988). (f) is female specific estimate and (m) is male specific. [a]optically derived axial length measurements and [b]is via ultrasound.

Study	Method	h^2 estimates		
		Corneal curvature	***Axial length***	***Anterior chamber depth***
Alsbirk**	Parent-offspring	0.64	0.76[b]	0.56
(Biino et al., 2005)	Midparent-offspring	0.57	0.31[a] (f) 0.60[a] (m)	0.47 (f) 0.44 (m)
Kimura**	Twin	0.55	0.68[a]	0.05
Nakajima**	Parent-offspring Twin	0.82 0.88	0.66[a] 0.72[b] 0.68[a]	0.30 0.26
(Sorsby, Sheridan and Leary, 1962; Sorsby, Leary and Fraser, 1966)	Parent-offspring Twin	0.75 0.78	0.62[a] 0.84[a]	0.52 0.68
(Young and Leary, 1972)	Parent-offspring	0.72	0.44[b]	0.52

3.2. Methods

3.2.1. Recruitment

A questionnaire (PQ1(IW), Appendix II) was sent to all of the family members in the large Irish-Welsh pedigree. The questionnaire asked subjects to indicate the name and address of their optometrist, the age they began to wear spectacles, any ocular surgery or treatment and any systemic medical conditions. Subjects were also asked if they were born prematurely. A copy of the family tree was also sent to each family member and they were asked to indicate if there were any other relatives missing (Appendix III). Any additional relatives were also sent a questionnaire and were included in the analysis if appropriate. Spectacle prescriptions were obtained from the subjects' optometrists, as previously.

The family were then invited to attend either one of two get-togethers in Kilkenny or Dublin, Ireland at which ocular biometry was undertaken along with autorefraction and DNA sample collection. Due to the lack of interest for the Dublin gathering, only the Kilkenny one went ahead.

3.2.2. Heritability Estimates

For each of the methods, the trait under investigation was refractive error of the right eye in the least minus meridian. (There was no statistically significant difference between the refractive error in the right and left eyes; Spearman's rank $r=0.957$, $p<0.001$) (table 3.5.). Least minus meridian is used to ensure the most conservative result and therefore maintaining a low type I error rate.

3.2.2.1. Within Family Regression

3.2.2.1.1. Midparent-offspring regression

The midparent trait value, i.e. the mean parental refractive error (RE LMM), was regressed against the trait value of the offspring. If there was more than one offspring within a nuclear family, the mean RE LMM value of the offspring was used. The regression coefficient was taken to be equal to the heritability. These calculations were performed both non-weighted and weighted for the number of children in the family. In the non-weighted situation, each family group (i.e. each midparent with the average of all offspring) provides an equal contribution to the calculation regardless of the number of offspring. Weighting provides a method to distribute the contribution of each of the families depending on the number of offspring; the larger families therefore contribute a higher proportion to the calculation.

3.2.2.1.2. Sibling-sibling regression

One sibling's refractive error was regressed against the other to obtain a regression coefficient equal to the heritability. If there are more than two siblings in a family, a variety of methods were used; following the approach of (Schmidt et al., 2005):

1. One sibling is chosen at random and the mean of the others is used as equivalent to the second sibling's trait value [*sib1 v mid sib (not weighted)*].

2. As 1. and then weighted using the weighted least squares method [*sib1 v mid sib (weighted)*].

3. As 1. with each sibling in turn being chosen as the first sibling [*multiple sib1 v mid sib (not weighted)*].

4. As 3. and weighted using the weighted least squares method [*multiple sib1 v mid sib (weighted)*].

5. A random sibling within each family was plotted against all of the other siblings separately, each point representing one pair of siblings [*multiple sib1 v sib2 (not weighted)*].

6. As 5. weighted using the weighted least squares method [*multiple sib1 v sib2 (weighted)*].

Although naturally weighted, cases 3 and 5 are described as non-weighted because the weighted least squares method was not applied. Cases 4 and 6 used the weighted least squares method to regress the data using the inverse number of siblings as the weighting variable.

The correlation coefficient between the two sets of values is equal to half of the heritability estimate.

3.2.2.2. Variance Components Method

The SOLAR (Sequential Oligogenic Linkage Analysis Routines) variance components linkage analysis software package was used to calculate heritability estimates by partitioning the phenotypic variance of the trait (refractive error, RE LMM) into genotypic and environmental components.

The pedigree structure, phenotypes and any covariates, for example, sex and age, were inputted into the programme. Each individual had an identifying number and by stating the parental identifiers the whole pedigree structure was constructed, thereby benefiting from using all familial relationships. This package also enables any effect of covariates on the phenotype to be accounted for.

The files were altered to use selected family members for further tailored analysis, removing females and males consecutively to produce gender specific results.

3.2.2.3. Markov Chain Monte Carlo (Gibbs sampling) methods

This analysis was performed at the Roslin Institute, Edinburgh, by R. Pong-Wong and C.S. Haley, using custom-written software.

3.3. Results

3.3.1. Recruitment

Two hundred and thirty family members were contacted and 148 consented to take part in the initial stage of the research project. Sixty seven subjects were sent a second questionnaire in an attempt to increase the response. Out of these 67, 20 more subjects consented to participation. Therefore there was an initial response rate of 64.3%, a follow up response rate of 29.9% and a total response rate of 73.0%.

Spectacle prescriptions were obtained for 85 individuals, and 57 subjects stated that they did not have an optometrist but also did not wear glasses. We were unable to obtain spectacle prescriptions for 16 subjects due to the lack of response from their optometrists. Five subjects did not write their optometrists' details on the form, one subject's optometrist did not have any records for that subject and four subjects did not give consent to contact their optometrist.

3.3.2. Trait Characteristics

Sixty six subjects were able to attend the Kilkenny gathering. Successful autorefraction results were obtained for 65 subjects. Axial length measurements were obtained for 49 right eyes and 42 left eyes. Keratometry measures were obtained for 56 right eyes and 58 left eyes (table A, Appendix III). DNA samples (mouthwashes) were obtained for all family members who attended the party and a further 47 subjects provided DNA samples by post. All of the subjects involved in the second stage of the project provided written consent (form CO(IW), Appendix II).

Autorefraction results were available for some subjects for whom a spectacle prescription was not available so a regression between spectacle prescriptions (RE LMM) and autorefraction results was performed. The resultant regression equation was used to convert any autorefraction results to the equivalent spectacle prescription where none were available (figure 3.3). This increased the number of subjects included in the analysis from 85 to 155.

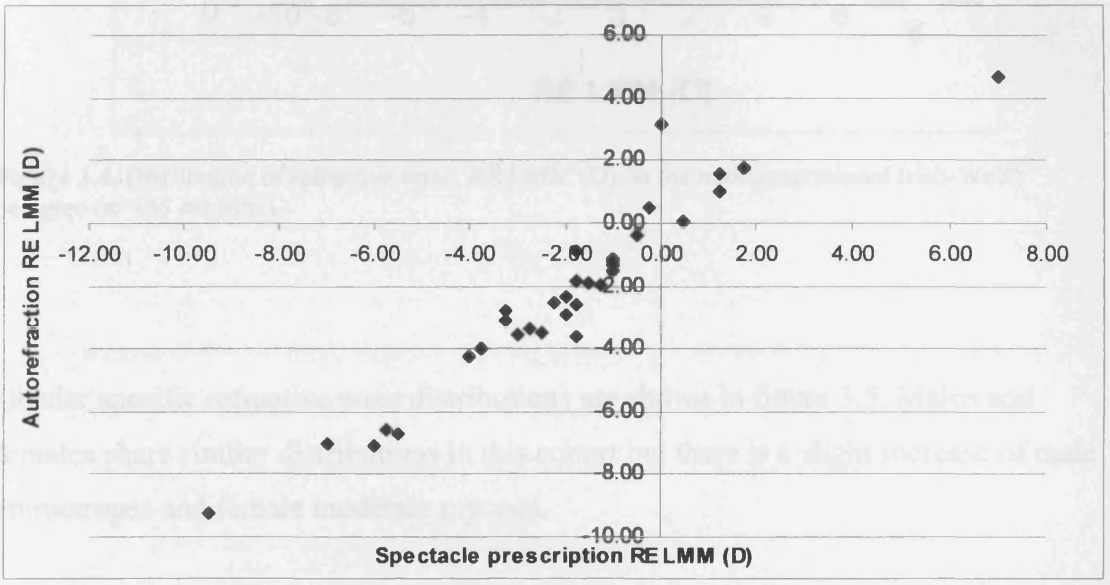


Figure 3.3. Regression of spectacle prescription against autorefraction (n=35 eyes from 35 subjects).

In order to assign a trait value to those subjects who did not have an autorefraction result and who also stated they wore no glasses and had never had an eye examination, the RE LMM of those people who had had an eye examination and/or an autorefraction was examined and the median value of this distribution was found to be zero (plano). Therefore those people with no phenotypic information available were assigned a refractive error trait value of zero (plano). The trait distribution was non-normal ($p<0.01$, Kolmogorov-Smirnov normality test), skewed towards myopia and leptokurtosed at emmetropia (figure 3.4.).

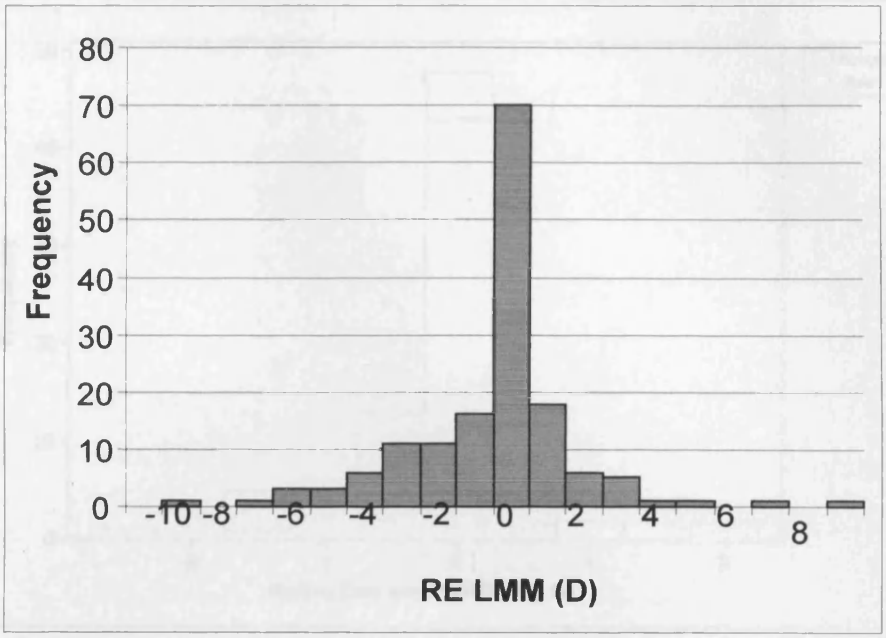


Figure 3.4. Distribution of refractive error, RE LMM (D), in the multigenerational Irish-Welsh pedigree (n=155 subjects).

Gender specific refractive error distributions are shown in figure 3.5. Males and females share similar distributions in this cohort but there is a slight increase of male emmetropes and female moderate myopes.

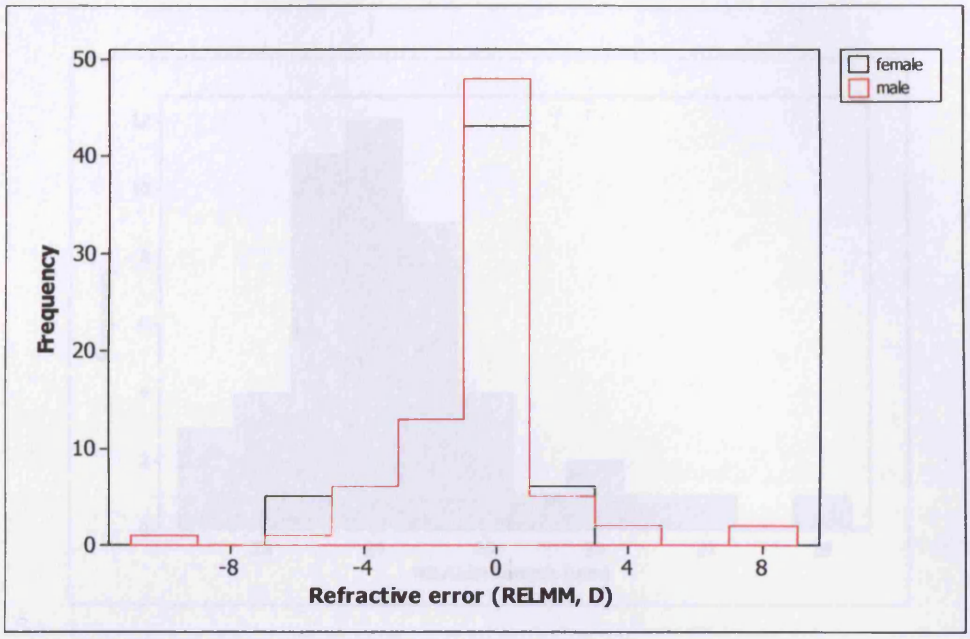


Figure 3.5. Female and male trait distributions of refractive error.

The characteristics of the biometry measurements are shown in table 3.4. and the frequency distribution curves are seen in figures 3.6. and 3.7.

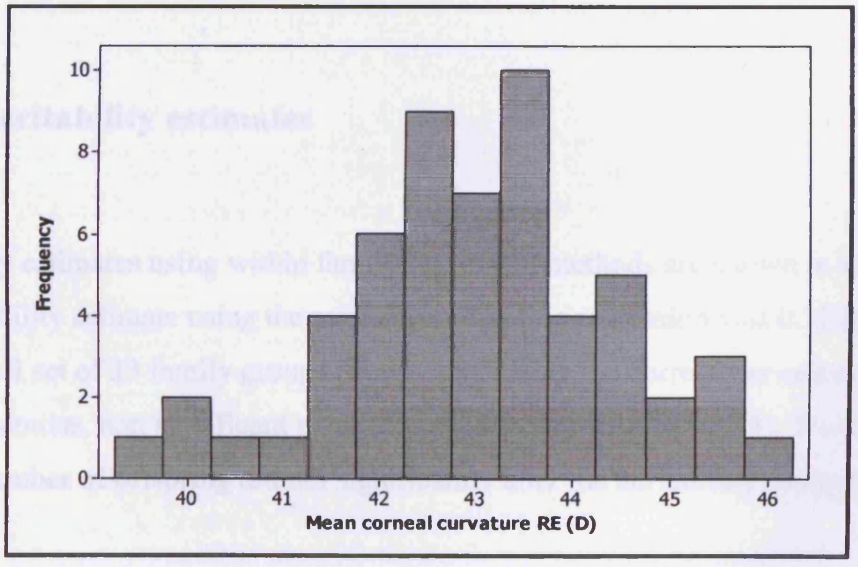


Figure 3.6. Frequency distribution of RE mean corneal curvature (D), in the multigenerational Irish-Welsh pedigree, (n=56 subjects).

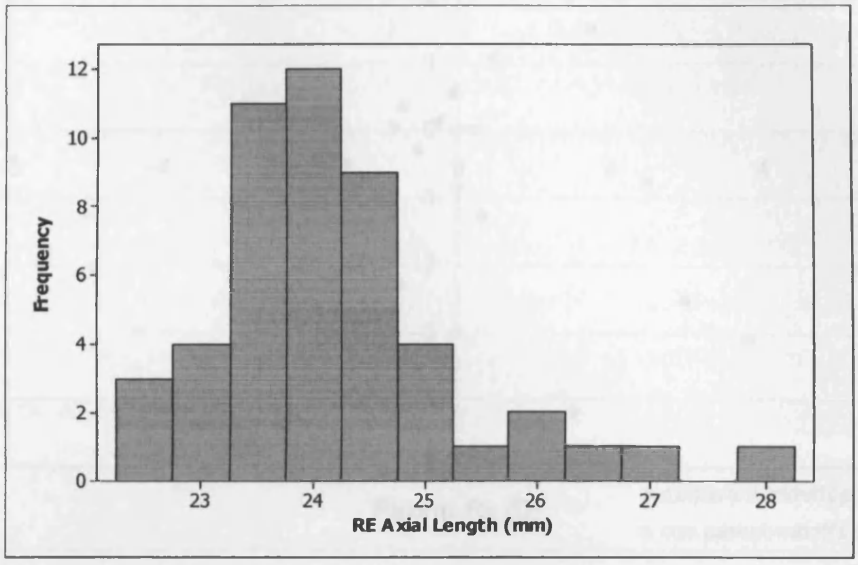


Figure 3.7. Frequency distribution of RE axial length (mm), in the multigenerational Irish-Welsh pedigree (n=49 subjects).

Right and left eye measurements are shown in table 3.5. The high correlation between eyes supports the use of only the right eye measurements.

3.3.3. Heritability estimates

Heritability estimates using within family regression methods are shown in table 3.6. The heritability estimate using the midparent-offspring regression was 0.12 (p=0.61) with the full set of 33 family groups (figure 3.8). Using the correlation calculation method a similar, non significant result was produced (r=0.19, p=0.55). Weighting the data for number of offspring did not significantly alter the heritability estimates.

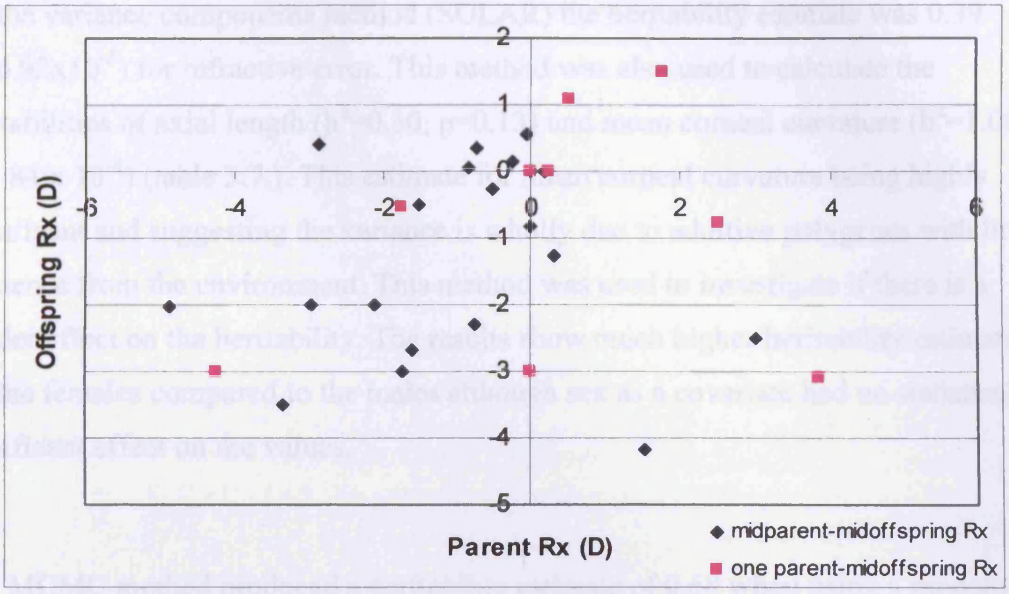


Figure 3.8. Midparent-offspring regression.

Sibling-sibling regression (figure 3.9.) gave heritability estimates of between 0.24 and 0.43 depending on which method was used (table 3.6). In general the correlation results gave higher heritability estimates, those statistically significant values being between 0.52 and 0.56.

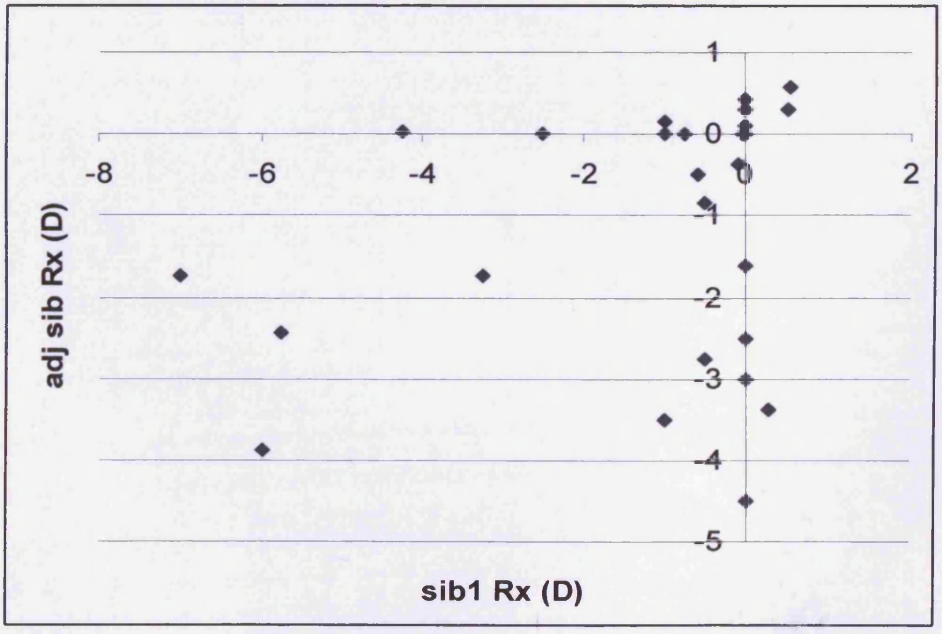


Figure 3.9. Sibling-sibling regression of refractive error (Sib1 v mid sib (not weighted)).

By the variance components method (SOLAR) the heritability estimate was 0.39 ($p=6.92x10^{-5}$) for refractive error. This method was also used to calculate the heritabilities of axial length ($h^2=0.30$, $p=0.13$) and mean corneal curvature ($h^2=1.00$, $p=1.84 \times 10^{-4}$) (table 3.7.). This estimate for mean corneal curvature being highly significant and suggesting the variance is wholly due to additive polygenes with little influence from the environment. This method was used to investigate if there is a gender effect on the heritability. The results show much higher heritability estimates for the females compared to the males although sex as a covariate had no statistically significant effect on the values.

The MCMC method produced a heritability estimate of 0.68 when using a random starting point for the Markov chain and 0.73 when using the mode of posterior densities for refractive error (table 3.8.); again suggesting a high proportion of variance is due to genetic factors. This method also included covariates of age and sex.

Table 3.4. Characteristics of biometry measurements in the multigenerational Irish-Welsh pedigree.

Biometry Measure	Mean		Median		Range		Non-Normally Distributed?	
	R.E.	L.E.	R.E.	L.E.	R.E.	L.E.	R.E.	L.E.
Autorefraction (D)	-0.96 (S.D.=2.73)	-1.03 (S.D.=2.75)	-0.33 (S.E.M.=0.33)	-0.25 (S.E.M.=0.34)	-9.75 to 4.78	-12.25 to 4.47	Yes ($p=0.03$)	Yes ($p<0.01$)
Axial Length (mm)	24.18 (S.D.=1.11)	24.03 (S.D.=1.15)	23.94 (S.E.M.=0.16)	24.02 (S.E.M.=0.18)	22.38 to 27.82	20.49 to 26.63	Yes ($p<0.01$)	No ($p>0.15$)
Mean Corneal Curvature (D)	43.00 (S.D.=1.38)	43.06 (S.D.=1.31)	42.94 (S.E.M.=0.19)	42.89 (S.E.M.=0.17)	39.27 to 45.86	39.59 to 45.61	No ($p>0.15$)	No ($p>0.15$)

Table 3.5. Right and left eye characteristics of traits.

Biometry Measure	N		RE v LE correlation	
	R.E.	L.E.	Pearson	Spearman's rank
Autorefraction (D)	67	67	$r=0.65$, $p<0.001$	$r=0.80$, $p<0.001$
Axial Length (mm)	49	42	$r=0.95$, $p<0.001$	$r=0.96$, $p<0.001$
Mean Corneal Curvature (D)	56	58	$r=0.97$, $p<0.001$	$r=0.97$, $p<0.001$

Table 3.6. Heritability estimates for refractive error using within-family regression and correlation methods.

Within families: Method	N	Regression				Correlation			
		h^2	S.E.	95% C.I.	P value	h^2	S.E.	95% C.I.	P value
Midparent-offspring ***(not weighted)***	22	0.10	0.19	-0.30 to 0.50	0.61	0.24	1.74	-3.17 to 3.65	0.61
Midparent-offspring ***(weighted)***	22	0.11	0.32	-0.56 to 0.78	0.21	0.26	1.02	-1.74 to 2.26	0.56
Sib1 v mid sib ***(not weighted)***	30	0.22	0.13	-0.05 to 0.49	0.10	0.60	1.46	-2.26 to 3.46	0.10
Sib1 v mid sib ***(weighted)***	30	0.27*	0.11	0.04 to 0.50	0.02	0.84*	2.16	-3.39 to 5.07	0.02
Multiple Sib1 v mid sib ***(not weighted)***	94	0.28*	0.07	0.14 to 0.42	0.00	0.74*	1.44	-2.08 to 3.56	0.00
Multiple Sib1 v mid sib ***(weighted)***	94	0.30*	0.07	0.14 to 0.42	0.00	0.64*	0.82	-0.97 to 2.25	0.00
Multiple sib1 v sib2 ***(not weighted)***	146	0.29*	0.08	0.14 to 0.46	0.00	0.58*	2.19	-3.83 to 4.99	0.00
Multiple sib1 v sib2 ***(weighted)***	146	0.26*	0.08	0.10 to 0.42	0.00	0.48*	0.85	-2.01 to 2.97	0.00

**significant at* $p \leq 0.05$

Table 3.7a. Heritability estimates for refractive components using SOLAR for all subjects

Ocular Component	n	All subjects			
		h^2	SE	95% C.I.	p
Refractive Error (D)	153	0.39*	0.14	0.12 to 0.66	6.92×10^{-5}
Transformed Refractive Error (D)	153	0.39*	0.13	0.14 to 0.65	9.90×10^{-6}
Axial Length (mm)	49	0.30	0.33	-0.35 to 0.95	0.13
Mean Corneal Curvature (D)	56	1.00*	0.22	0.57 to 1.43	1.84×10^{-4}

**significant at* $p \leq 0.05$

Table 3.7b. Heritability estimates for refractive components using SOLAR for male and female subjects separately

Ocular Component	Male					Female				
	n	h^2	SE	95% C.I.	p	n	h^2	SE	95% C.I.	p
Refractive Error (D)	80	0.37*	0.21	-0.04 to 0.78	0.03	73	0.61*	0.22	0.18 to 1.04	9.00×10^{-4}
Transformed Refractive Error (D)	80	0.39*	0.21	-0.02 to 0.80	0.02	73	0.72*	0.21	0.31 to 1.13	6.20×10^{-4}
Axial Length (mm)	27	0.08	0.32	-0.55 to 0.71	0.40	22	0.73	0.60	-0.45 to 1.91	0.16
Mean Corneal Curvature (D)	29	1.00*	0.00	1.00 to 1.00	3×10^{-3}	27	1.00*	0.00	1.00 to 1.00	0.02

Table 3.8. Heritability estimates for refractive error using MCMC methods.

Method	N	h^2	Standard error	95% C.I.
MCMC	89 individuals	0.68	0.17	0.35 to 1.01
MCMC (mode of posterior densities)	89 individuals	0.73	0.17	0.35 to 1.01

3.4. Discussion

For refractive error a range of heritability estimates were obtained, ranging from 0.12 and not significantly different from zero, to 0.73 and highly significant, within this pedigree. The heritability estimates for the refractive components of the eye show only refractive error and mean corneal curvature to be significant. Plotting refractive error against mean corneal curvature shows a correlation of -0.35 (Pearson correlation coefficient, p=0.009) when considering all of the subjects together (figure 3.10.). However the wide spread of the data suggests there may be some interaction between the other components of the eye when determining refractive error, therefore the possibility of mapping a QTL controlling refractive error based on the control of the corneal curvature may not be as likely as the heritability value suggests.

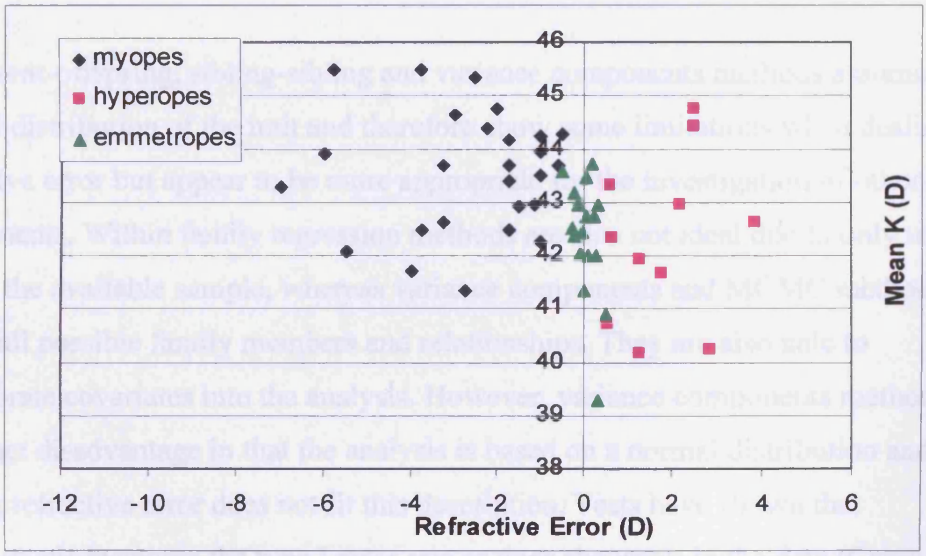


Figure 3.10. Correlation between refractive error and mean corneal curvature.

Axial length, despite its strong correlation with refractive error has an estimate which is statistically insignificant. However, for this measurement the sample size could have significantly affected the results (n=49 for axial length and n=153 for refractive

error); those included individuals may not be as closely related to optimise the heritability calculations and allow less assumptions to be made regarding the missing data. Animal studies have also suggested that axial length change (specifically vitreous chamber depth) is visually driven; changes in the environment producing a change in the rate of vitreous chamber elongation, and therefore the trait is less likely to be heritable. Corneal curvature, on the other hand, alters little with a change in environmental factors and is therefore more likely to be genetically based and so a highly heritable trait.

The results produced here have a general trend of having higher heritability estimates within the female section of the pedigree for refractive error compared to males and the same estimate for mean corneal curvature. Considering sex as a covariate the effects are not significant: however there is a slight increase in myopic females and a reduction in emmetropic females compared to males and therefore the increased heritability may be mostly driven by this difference.

Midparent-offspring, sibling-sibling and variance components methods assume a normal distribution of the trait and therefore show some limitations when dealing with refractive error but appear to be more appropriate for the investigation of other ocular components. Within family regression methods are also not ideal due to only using part of the available sample, whereas variance components and MCMC methods utilise all possible family members and relationships. They are also able to incorporate covariates into the analysis. However, variance components methods have a distinct disadvantage in that the analysis is based on a normal distribution and, as shown; refractive error does not fit this description. Tests have shown that leptokurtosis increases the type I error rate as does skewness in the data (Camp and Cox, 2002). Mean corneal curvature and axial length estimates using this method can therefore be regarded as accurate but it is more of a suggestive value for refractive error.

The estimates measured for the Irish-Welsh family are unique as heritability is a measure of the proportion of variation of phenotype due to genetic factors within the

specified population, therefore other studies using different or isolated populations are not comparable, but may be regarded as suggestive of a trend. The trends set by heritability studies of alternative populations of ocular refractive components being highly heritable are followed by this pedigree when considering refractive error and corneal curvature.

4. Evaluation of Quality and Quantity of DNA Extracted from Mouthwashes.

4.1. Introduction

4.1.1. Mouthwashes.

The standard procedure for collecting DNA for genetic and epidemiological studies involves obtaining blood samples from individuals. Each millilitre of blood will provide a yield of around 30μg DNA. However, from a 10ml mouthwash, a range of yields of up to ~200μg is possible. Studies have also shown that using treated cards (Harty et al., 2000) and cytobrush techniques (Garcia-Closas et al., 2001) for sample collection also work well and have similar advantages to the mouthwash samples as compared to blood but, on average, provide lower yields of human DNA. For the DNA collected by treated cards, a significant decline in yield occurred after storage at room temperature and –70°C after nine months (Harty et al., 2000).

Advantages of using mouthwashes as a sample collection technique over blood include increasing the likelihood of subject participation (by eliminating both the need for medical supervision and reducing the risk of infection from hepatitis and HIV) and decreasing the inconvenience to the subject. Also, the cost is less and the collection is made easier due to the ability to post the samples (Lench et al., 1988).

To obtain maximum DNA yield and integrity samples should be collected before tooth brushing; those collected after show a 40% reduction in yield (Feigelson et al., 2001). They should also be extracted within a week of collection to avoid degradation (Harty et al., 2000; Heath et al., 2001). For a genome scan, 300-400 polymerase chain reactions are required per subject, 40ng of DNA being required for each. To be certain of sufficient DNA, allowing for reactions that may need to be repeated, at least 14μg DNA yield is required from the two mouthwashes provided by each individual. As genetic analysis of large numbers of individuals by molecular methods is often

restricted by the availability of DNA (Lench et al., 1988), by using mouthwashes as a source, more samples can be requested at little inconvenience to the subject. Alternatively, a method of amplifying the DNA can be attempted. This is discussed further in section 4.1.2.

4.1.2. Whole Genome Amplification (WGA)

The quantity of high quality genomic DNA is often a major factor governing the breadth and potential of a genetic study. Low DNA yields lead to a reduction in the amount of genotyping possible, limiting the available information and therefore resulting in a loss of statistical power. One solution to this limitation is the application of whole genome amplification (WGA). Previous methods of increasing DNA yield are PCR based and have been found to provide amplification bias and inadequate genome coverage (Dean et al., 2001). Ideally, the WGA method should produce equal amplification of the total genome with no amplification error or non specific product amplification. Whole genome amplification by multiple displacement amplification (MDA) has been shown to fulfil these criteria by providing a highly uniform representation across the whole genome with minimal amplification bias (Dean et al., 2002; Paez et al., 2004). This method works by non specific primers binding to regions of complementary DNA within the genome, and then providing strand displacement (unwinding of the DNA double helix) and replication of the template by φ29 DNA polymerase. This φ29 DNA polymerase is a highly specific enzyme with both strand displacement activity and proofreading capabilities i.e. the ability to check the sequence of the new DNA strand and remove any mistakenly added bases. Further benefits of this method include potential amplification from as little as 1-10 copies of human genomic DNA (gDNA) and no need for adjustment of final concentration due to a saturation of the reaction yielding the same amount of final DNA independent of the amount of initial template.

The multiple displacement amplification method (MDA) involves firstly alkaline-denaturing the double stranded DNA template. Addition of a neutralising buffer reduces the pH again but due to the length and complexity of the newly single

stranded template, re-annealing does not occur. A hexamer primer comprised of six randomly ordered bases binds to the template at regions of complementarity. This primer remains intact in spite of the φ29 polymerase's proofreading (3' to 5' exonuclease activity) due to the inclusion of two 3' phosphorothioate bonds. Pyrophosphatase is added to the reaction to break down diphosphates released by the dNTPs when bases are added to create new DNA strands. An accumulation of diphosphates can inhibit the activity of the φ29 DNA polymerase. The MDA reaction takes place over 16 hours of incubation.

Using single nucleotide polymorphisms (SNPs) to test these methods for accurate and concordant genotyping, a concordance rate of over 99% was found (Barker et al., 2004; Paez et al., 2004), the latter investigation finding an increase in this concordance rate with the addition of a DNA alkaline denaturing step prior to amplification. Holbrook and colleagues (Holbrook et al., 2005) found that there was 100% concordance between genotyped microsatellite markers of pre- and post-WGA samples, suggesting accurate replication and therefore making MDA WGA a solution to increasing supplies of depleted DNA samples.

Unfortunately, MDA does not provide a solution to recovering already degraded DNA samples, the process producing a smaller yield and a slower rate of MDA reaction in these cases (Holbrook et al., 2005). The degraded DNA will not provide a high quality template for MDA and other techniques including PCR due to the low molecular weight and potential early sequence termination of the specified section. Amplification of this template will only produce larger volumes of the shorter and degraded sequences in an inefficient reaction and will not restore the quality of the original sample.

Dickson and colleagues (Dickson et al., 2005) evaluated the suitability of using WGA on samples designed for microsatellite marker genotyping and found only a very slight increase in error rate in the MDA samples of approximately 0.6% compared to the human genomic DNA samples along with some variation in allele peak heights, indicating some amplification bias. Both of these problems could be improved by pooling multiple samples. However, there was some inter-marker variability, with

some microsatellites providing the majority of the genotyping errors within the MDA group of samples.

An investigation of genotyping DNA from different tissue samples showed equal and complete concordance between blood and buccal cells using both SNPs and short tandem repeats (Hosono et al., 2003).

Genomic representation is also a critical issue. Reports have shown that coverage is comparable to that of non amplified DNA (Dean et al., 2002; Hosono et al., 2003; Paez et al., 2004). The study by Paez and colleagues stated there were several chromosomal sections that were not represented. There were also sections of chromosome that were consistently underrepresented in the amplified samples even though the SNPs were concordant. This could become problematic for some genetic analyses techniques. However, the evidence from these studies suggests that, in general, this method is an extremely useful solution to the problems of sample depletion for genetic studies.

4.2. Methods

4.2.1 Sample Collection

Two sterile tubes each containing 15ml sterile 0.9% saline solution were posted to each participating subject. The subject poured the solution into his/her mouth and swished it around for 20 seconds. The subject was asked to spit the solution back into the tube. This was repeated for the second solution and they were asked to post the samples back as soon as possible after they had taken them in a pre-paid envelope. The subjects were asked to provide the samples first thing in the morning before eating, drinking or tooth brushing (see mouthwash instructions –Appendix I).

4.2.2. DNA Extraction from Mouthwashes.

On receiving the samples, they were refrigerated as soon as possible for at least 30 minutes. The samples were centrifuged at 3500rpm for 5-7 minutes to pellet the buccal cells (this was repeated if the pellet was not fully formed) and the supernatant was discarded. The pellet was resuspended by adding 380µl proteinase K buffer (10mM tris-HCl, pH8.0, 1mM EDTA, 0.5% SDS) and pipetted up and down until fully resuspended. The mixture was transferred to a 1.5ml screw top tube and frozen at -20°C until ready to process further. The samples were removed from the freezer, thawed at 37°C, vortex-mixed and briefly centrifuged. Twenty microlitres of proteinase K (10mg/ml) was added to each tube and incubated at 37°C for two hours in a waterbath with continuous shaking (100rpm). The samples were removed from the waterbath and centrifuged at 14000rpm for three minutes to pellet insoluble material and the supernatant was transferred to a 1.5ml "silicon grease" Eppendorf tube. These tubes were prepared by adding approximately 100µl silicon grease (high vacuum grease; Dow Corning®, Midland, U.S.A.) to the underside of the hinge of an Eppendorf tube, then centrifuging the tube for four seconds at 3000rpm. Phenol/chloroform (phenol:chloroform:isoamyl alcohol-25:24:1), 470µl, was added and vortexed vigorously for 30 seconds and then the tube was centrifuged at 14000rpm for two minutes. If debris remained in the supernatant, this layer was

transferred to a second silicon grease Eppendorf tube and the phenol/chloroform extraction was repeated. When no debris remained in the supernatant, it was transferred to a sterile 1.5ml screw top vial and 19µl 5M NaCl was added. The mixture was vortex-mixed and briefly centrifuged. One millilitre of 100% ethanol was added, mixed and left to precipitate at -20°C for 30 minutes or more. The samples were removed from the freezer, inverted a few times to mix and then centrifuged at 14000rpm for 10 minutes. The supernatant was discarded, 1ml of ice cold 70% ethanol was added and the solution was then centrifuged at 14000rpm for two minutes. The majority of the supernatant was removed, then a narrow bore pipette tip was used to remove the last traces of ethanol. The tube was air dried in an inverted position for five minutes. The pellet was resuspended in 51µl TE (10mM Tris, 1mM EDTA) and incubated for 15 minutes at 37°C with periodic gentle vortexing.

4.2.3. Quantification of DNA Yield and Quality by Spectrophotometry.

A 1:100 dilution was prepared by adding 1µl sample to 99µl TNE (10mM Tris, 100mM NaCl, 1mM EDTA, pH 8.0). The absorption at wavelengths of 260, 280 and 320nm was measured (A_{260}, A_{280} and A_{320} values) by spectrophotometry (GeneQuant™, GE Healthcare Life Sciences, Amersham, UK) and the concentration and yields were calculated using equations 4.1. and 4.2.

[1] $$\text{DNA concentration } (\mu g/\mu l) = \frac{A_{260} \times 50 \times \text{dilution factor}}{1000}$$

[2] $$\text{DNA Yield } (\mu g) = \text{DNA concentration } (\mu g/\mu l) \times \text{volume } (\mu l)$$

Equations 4.1. and **4.2.** DNA Concentration and Yield Calculation.

4.2.4. Investigation of origin of DNA

Spectrophotometry provides a quantification method of DNA samples. However, this value may be inaccurate due to the inability of this method to distinguish between human genomic DNA, RNA and DNA from sample contaminants such as bacteria and food (Garcia-Closas et al., 2001). Therefore to calculate the actual human genomic DNA content a human DNA assay was carried out.

4.2.4.1. SYBR® Green I human genomic DNA assay

DNA samples under test were diluted to a concentration of 6.67ng/µl, according to their spectrophotometric absorbence. Six microlitres of this DNA template solution was added to 6µl master mix solution to give final concentrations of 1 x Hotstar PCR buffer, 0.2mM dNTP mix, 0.3µM primer mix (D1S534, D4S1625 or D7S3506, upstream and downstream) and 0.033U/µl Hotstar Taq. The same was repeated for human placental genomic DNA standards, purchased from Sigma, of concentrations of 6.67ng/µl, 5ng/µl, 3.33ng/µl and 1.67ng/µl. The PCR was cycled as follows:

15 minutes @ 95°C
20 seconds @ 94°C
30 seconds @ 63°C } 30 cycles
45 seconds @ 72°C
5 minutes @ 72°C

Three microlitres 5x Ficoll EDTA loading dye (15% Ficoll 400, 10mM EDTA, 0.5% xylene cyanol FF) was added to each PCR sample and 10µl of this sample was run on a 3% agarose gel. The samples were run at 70V for approximately 30 minutes. The gel was stained with SYBR® Green I for 20 minutes and then destained for 15 minutes. A photograph was taken and then analysed using ImageMaster 1D Prime program (ImageMaster 1D Prime, GE Healthcare Life Sciences, Amersham, UK). A standard curve was produced from the human DNA standards and calculations for the other samples were taken from this curve.

Final human genomic DNA concentrations were then calculated for the subjects' mouthwashes. Samples containing less than 15µg human gDNA underwent whole genome amplification and the human gDNA assay was then repeated until at least 15µg human gDNA was present.

4.2.5. Whole Genome Amplification

The MDA reaction saturates at a specific yield and this occurs independently of the initial template. Therefore one microlitre of DNA from the first sample from each subject was added to 4µl TE (10mM Tris, 1mM EDTA) irrespective of its initial DNA concentration. "Solution A" was prepared by adding 900µl H_20 to 80µl 5M KOH and 20µl 0.5M EDTA, pH8.0. Five microlitres of a 1:8 dilution of solution A were added to each sample, mixed by pipetting and incubated at room temperature for three minutes. Ten microlitres of 20mM HCl were then added. Eighty microlitres of mastermix were added giving final concentrations of 1 x MDA buffer (37mM Tris-HCl (pH 7.5), 50mM KCl, 10mM $MgCl_2$, 5mM $(NH_4)_2SO_4$), 1mM dNTPs, 50µM WGA random hexamer primer (with 2 phosphorothioate linkages at the 3' end), 0.1U pyrophosphatase and 80U φ29 DNA polymerase. The plate was heat sealed and placed in a thermal cycler for 16 hours at 30°C followed by five minutes at 65°C. The final yield was quantified by spectrophotometry as previously and then human genomic DNA content was calculated by performing the SYBR® green human genomic DNA assay. For those subjects with no PCR band on the gel, i.e. a minimal human gDNA yield, WGA was repeated.

4.3. Results.

4.3.1. Spectrophotometry

A total of 113 subjects from the Irish-Welsh family provided mouthwashes. Figures 4.1 and 4.2 show the distribution of yields of DNA calculated by spectrophotometry from all mouthwashes and from all subjects (sum of both mouthwashes for each subject) respectively.

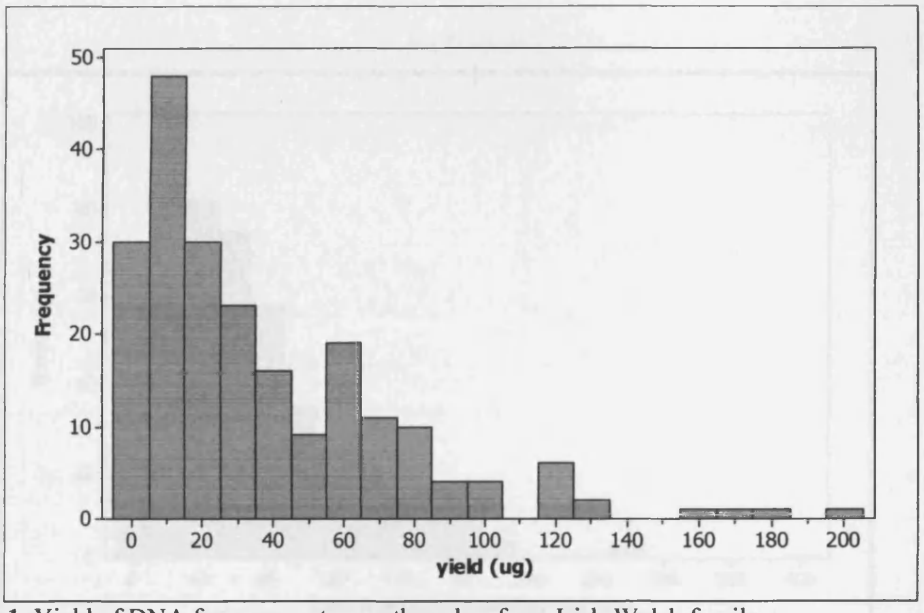


Figure 4.1. Yield of DNA from separate mouthwashes from Irish-Welsh family.

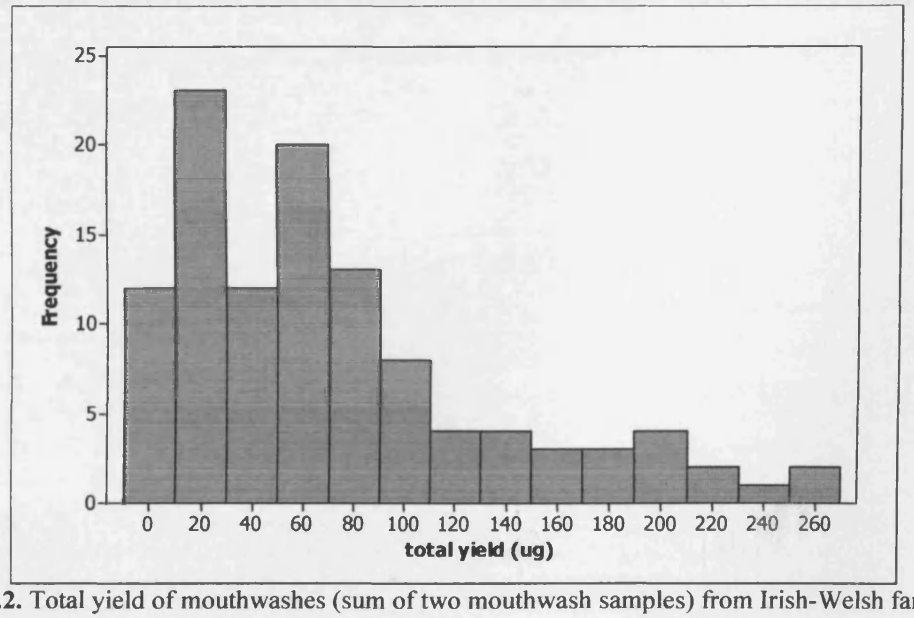


Figure 4.2. Total yield of mouthwashes (sum of two mouthwash samples) from Irish-Welsh family.

Each mouthwash gives an average yield of 40.56μg, the median value being 29.25μg (SEM=2.80) and the range from 0 to 184.25μg. The total yields from both mouthwashes have an average of 78.76μg, a median of 65.25μg (SEM=6.57) and a range from 3.43 to 255.25μg.

Other members of the Family Study of Myopia also provided mouthwash samples for alternative genetic analyses including an association study (chapter 6) and for future linkage analysis studies. Of the families recruited, approximately 243 subjects provided mouthwash samples. The distributions are shown in figures 4.3 and 4.4.

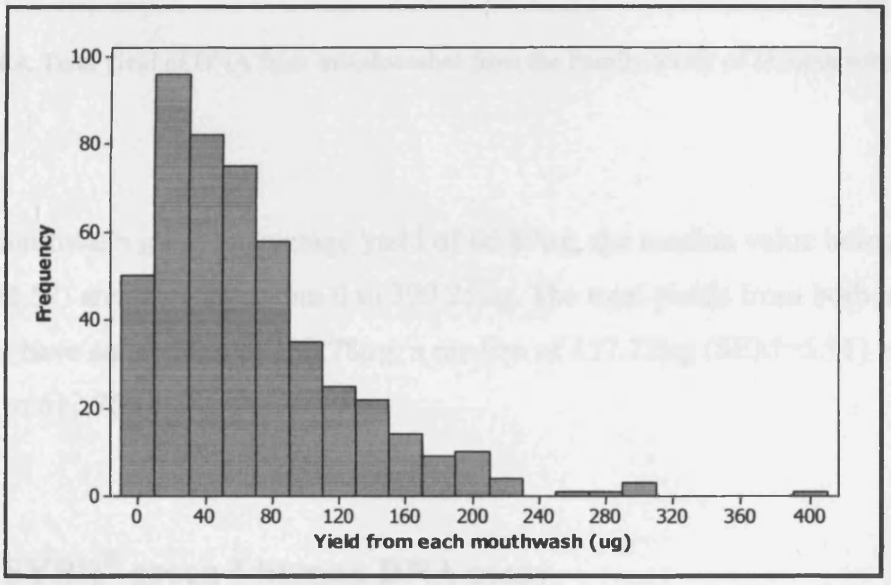


Figure 4.3. Yield of DNA from separate mouthwashes from the Family Study of Myopia subjects.

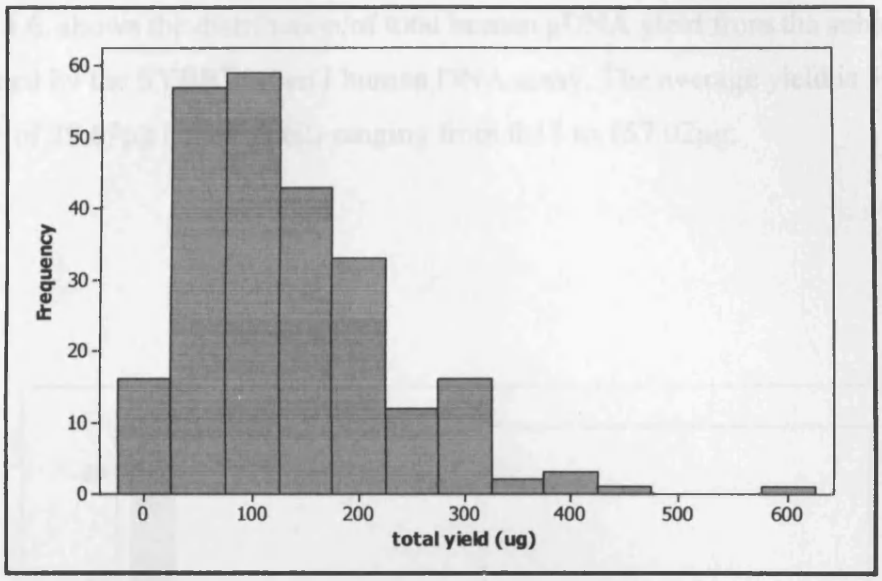


Figure 4.4. Total yield of DNA from mouthwashes from the Family Study of Myopia subjects.

Each mouthwash gives an average yield of 66.89µg, the median value being 52.88µg (SEM=2.57) and the range from 0 to 399.25µg. The total yields from both mouthwash samples have an average of 133.78µg, a median of 117.75µg (SEM=5.91) and a range from 1 to 612.75µg.

4.3.2. SYBR® green I human DNA assay

Figure 4.5. shows an example of the agarose gel image produced for the human genomic DNA assay.

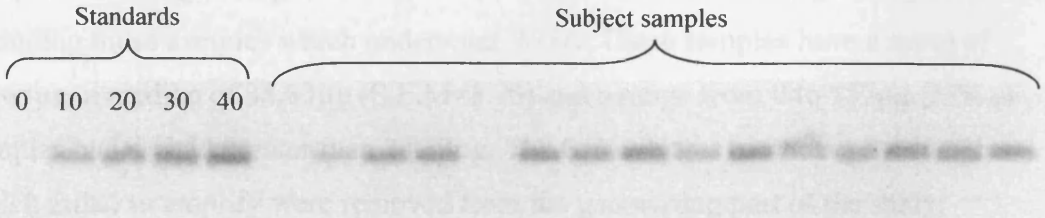


Figure 4.5. Agarose gel image. Bands shown are human specific PCR products stained with SYBR® green I. The bands' pixel density is compared to that of the standards to quantify the DNA yield. The figures for the standards refer to the amount of human genomic DNA (ng) used as a template.

Figure 4.6. shows the distribution of total human gDNA yield from the subjects calculated by the SYBR® green I human DNA assay. The average yield is 36.97µg, median of 29.67µg (SEM=3.62) ranging from 0.11 to 157.02µg.

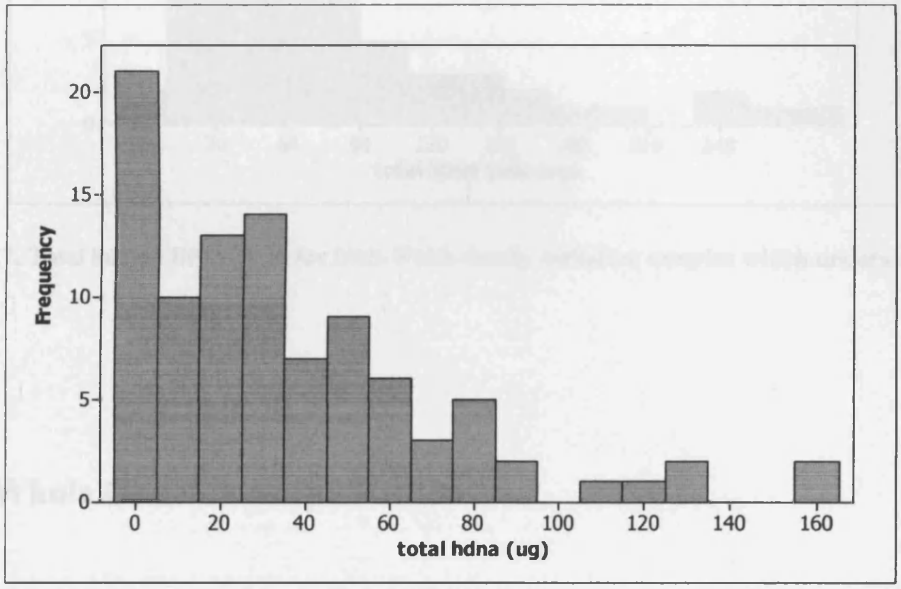


Figure 4.6. Total yields of human gDNA from SYBR® green I human gDNA assay.

As stated previously, it is necessary to have a total yield of approximately 15µg or more human DNA. Thirty one individuals in the Irish-Welsh family had less than this value and so underwent whole genome amplification. Out of 31 subjects, only two subjects failed to produce a WGA human gDNA yield, 29 subjects amplified to give adequate human gDNA yields. Figure 4.7. shows the final human gDNA yields including those samples which underwent WGA. These samples have a mean of 50.67µg, a median of 38.83µg (S.E.M=3.70) and a range from 0 to 157µg. 75% of samples had a yield greater than 24.33µg. The two subjects providing the samples which failed to amplify were removed from the genotyping part of the study.

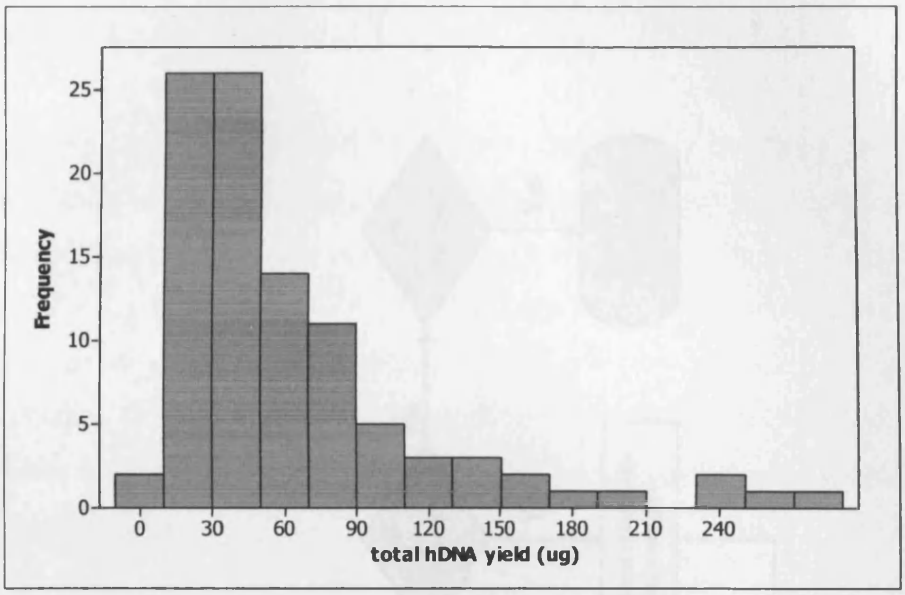


Figure 4.7. Total human DNA yield for Irish-Welsh family including samples which underwent WGA.

4.3.3. Whole Genome Amplification

See figure 4.8. The samples requiring whole genome amplification were identified and the first 16 samples were amplified. Only 6.25% of the samples amplified successfully. The next step was to try the alternate mouthwash from those subjects whose MDA WGA did not provide a human gDNA yield of over 15µg. The alternate mouthwash from these 15 subjects was taken along with both mouthwash samples from a further eleven subjects for amplification. From these first two attempts, 20 samples from 15 subjects amplified to give a human gDNA yield of over 15µg giving a sample success rate of 35.1% and a subject success rate of 55.6%. In an attempt to increase this success rate the volume of initial template was increased from 1µl to 5µl. From 29 samples, 51.9% were successfully amplified. The addition of an extra phenol-chloroform extraction step further increased the number of samples with an amplified human gDNA yield of over 15µg by 10.2% of subjects. Therefore in total, the DNA for 28 subjects was amplified using this technique to be used for genotyping and only 3% of subjects were excluded due to non-amplification.

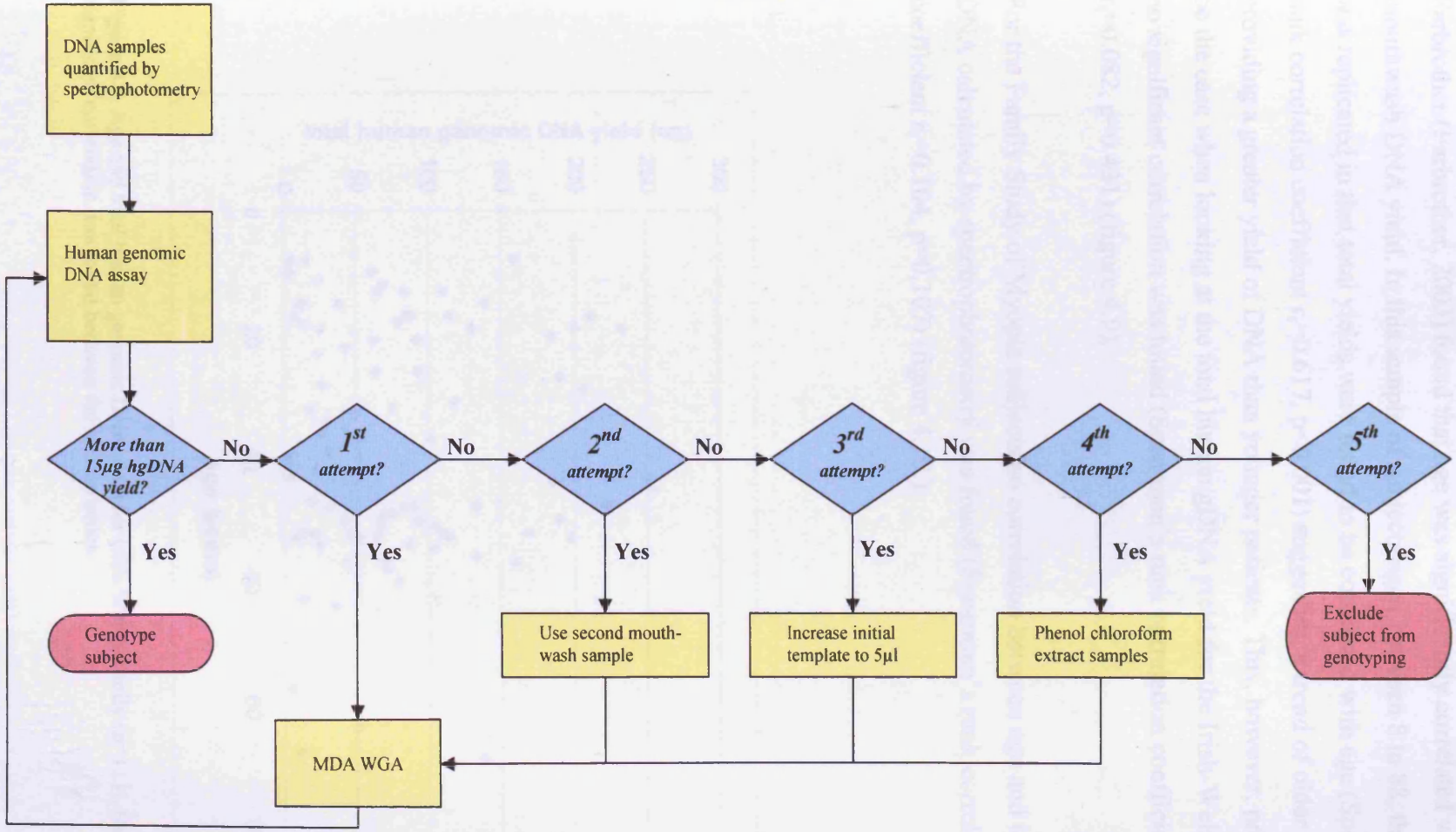


Figure 4.8. Flowchart of the sample preparation pre-genotyping process using a human genome DNA assay and multiple displacement amplification for the whole genome.

4.3.4. Age and total human DNA yield

Farbrother (Farbrother, 2003) found that age was significantly correlated with mouthwash DNA yield. In this sample of subjects aged between 8 to 88, this finding was replicated in that total yields were found to be correlated with age (Spearman's rank correlation coefficient r_s=0.617, p<0.001) suggesting a trend of older patients providing a greater yield of DNA than younger patients. This, however, proved not to be the case when looking at the total human gDNA yield for the Irish-Welsh family, no significant correlation was found (Spearman's rank correlation coefficient r_s=0.082, p=0.431) (figure 4.9).

For the Family Study of Myopia subjects no correlation between age and total yield of DNA calculated by spectrophotometry was found (Spearman's rank correlation coefficient r_s=0.104, p=0.107) (figure 4.10.).

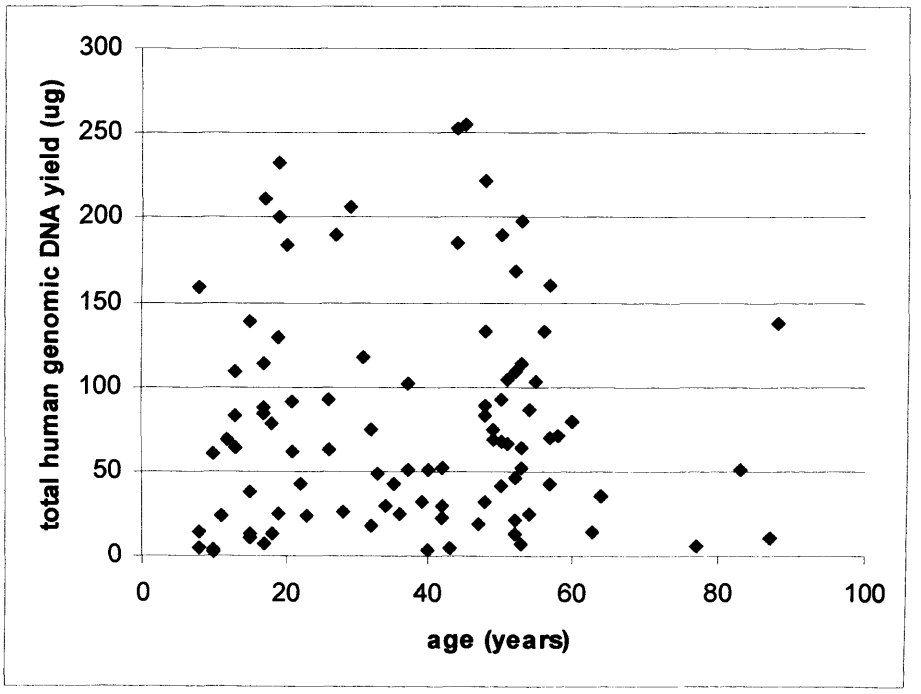


Figure 4.9. Age and total human genomic DNA yield for Irish-Welsh family (n=113). No statistically significant correlation was found between the two variables.

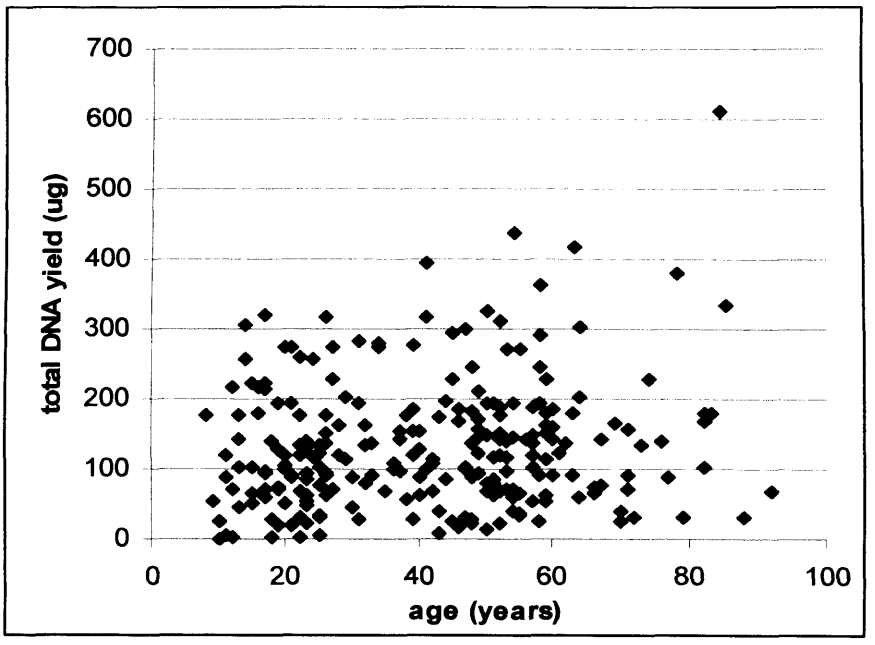


Figure 4.10. Age and total DNA yield for FSM subjects (n=243). No statistically significant correlation was found between the two variables.

4.4. Discussion

For those DNA samples with an initial adequate yield (i.e. a total yield of more than 15µg) the total yield of DNA estimated by spectrophotometry and by the human DNA assay, showed a statistically significant shift in yield distribution (Mann Whitney, $p<0.001$ for spectrophotometry>human gDNA yield) (figure 4.11.).

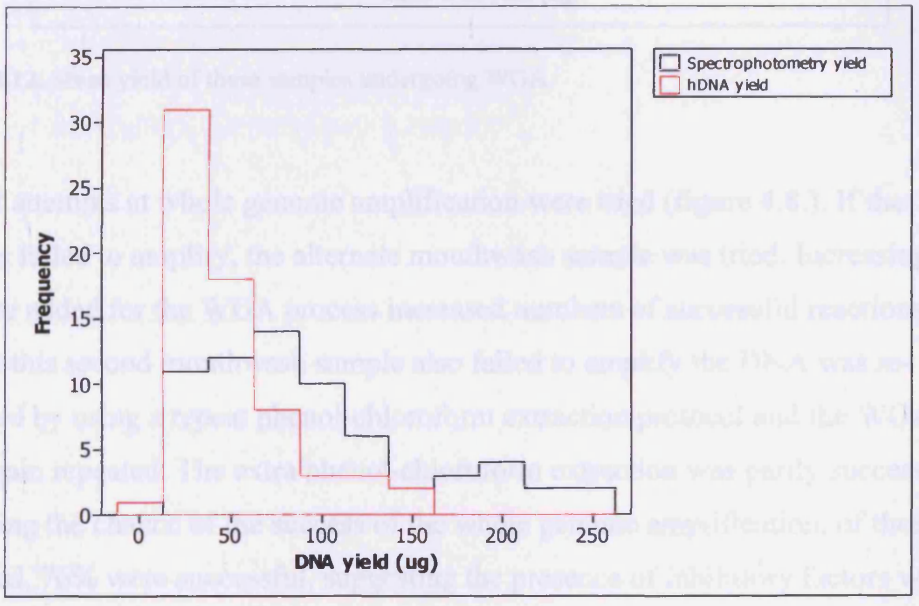


Figure 4.11. Yield distributions for spectrophotometry and human gDNA yields.

Mouthwashes appear to be a very useful method of DNA collection, providing up to a maximum of 157µg human gDNA, with 95% of subjects providing a total yield of 28µg or more. For those samples providing a low value, whole genome amplification can be performed increasing the yield to up to 243µg (mean=70.1µg, median=50.3µg, stdev=62.4, S.E.M.=10.9) from as little as 20ng of initial template.

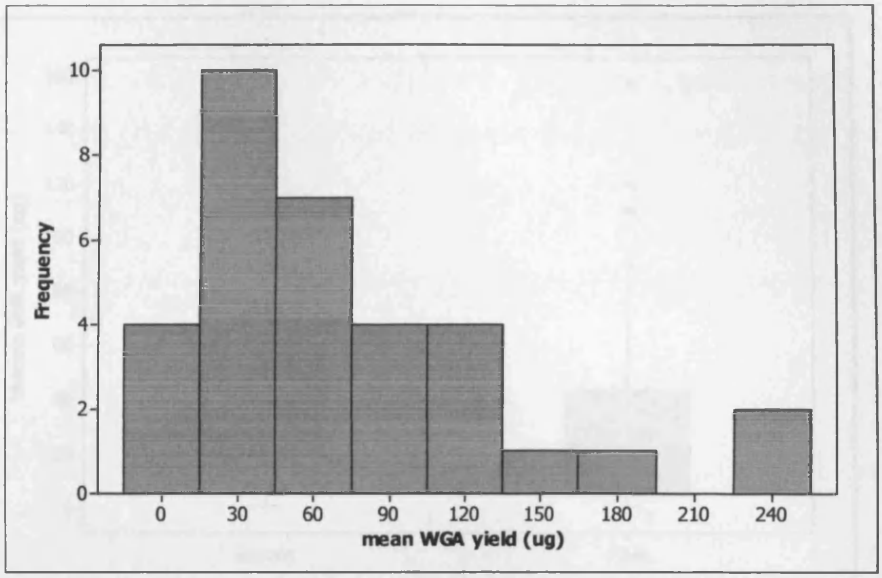


Figure 4.12. Mean yield of those samples undergoing WGA

Several attempts at whole genome amplification were tried (figure 4.8.). If the samples failed to amplify, the alternate mouthwash sample was tried. Increasing the template added for the WGA process increased numbers of successful reactions by 45%. If this second mouthwash sample also failed to amplify the DNA was re-extracted by using a repeat phenol-chloroform extraction protocol and the WGA was once again repeated. The extra phenol-chloroform extraction was partly successful in increasing the chance of the success of the whole genome amplification, of those re-extracted, 76% were successful, suggesting the presence of inhibitory factors within the initial template, possibly due to protein contamination within the mouthwash prior to extraction. Contamination of the mouthwash with non-human DNA, for example from food or bacteria, may partly explain the difference between spectrophotometry readings and the human DNA assay values, the alien DNA absorbing the light along with the human DNA.

Age was found not to affect the human gDNA yield in this sample population. Separating the Irish-Welsh mouthwash samples into male and female and comparing these yields shows a statistically significant difference between genders such that females produce a larger human DNA yield than males ($p=0.0075$, Mann Whitney) (figures 4.13 and 4.14.).

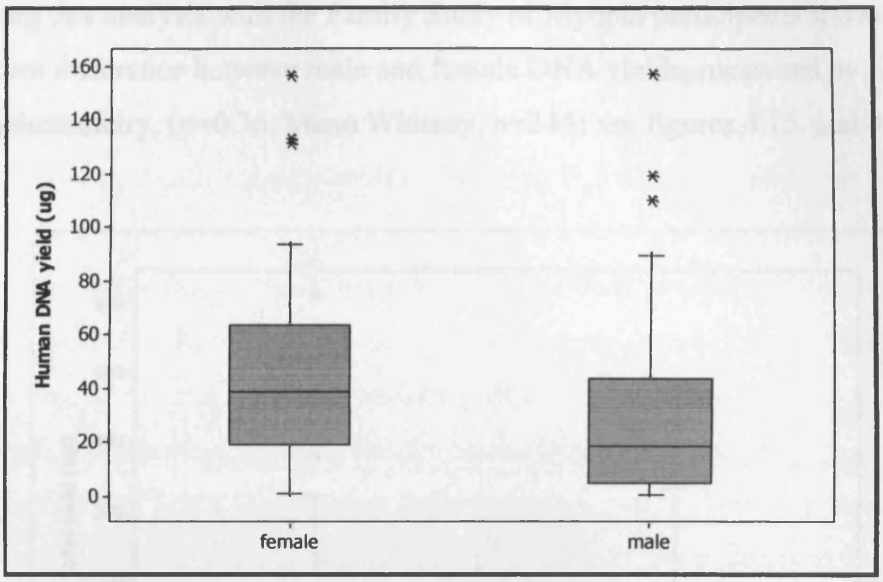


Figure 4.13. Human DNA yield in females and males

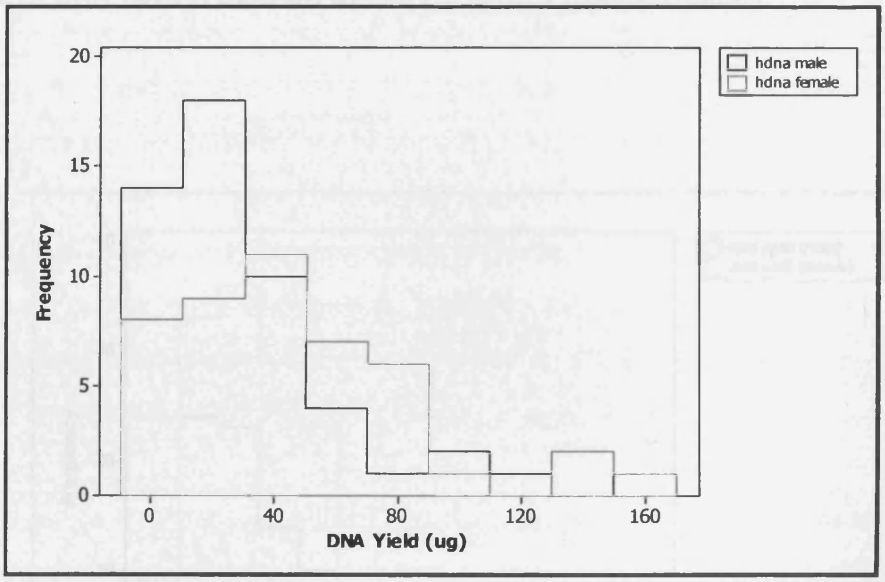


Figure 4.14. Human DNA yield in females (red outline) and males (black outline)

Repeating this analysis with the Family Study of Myopia participants showed no significant difference between male and female DNA yields, measured by spectrophotometry, (p=0.36, Mann Whitney, n=244) see figures 4.15. and 4.16.

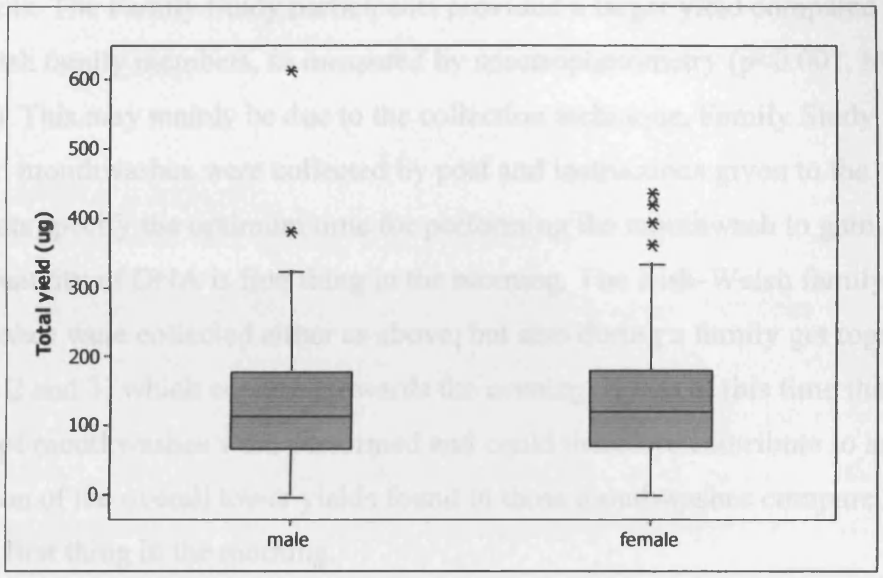


Figure 4.15. DNA yields in males and females measured by spectrophotometry.

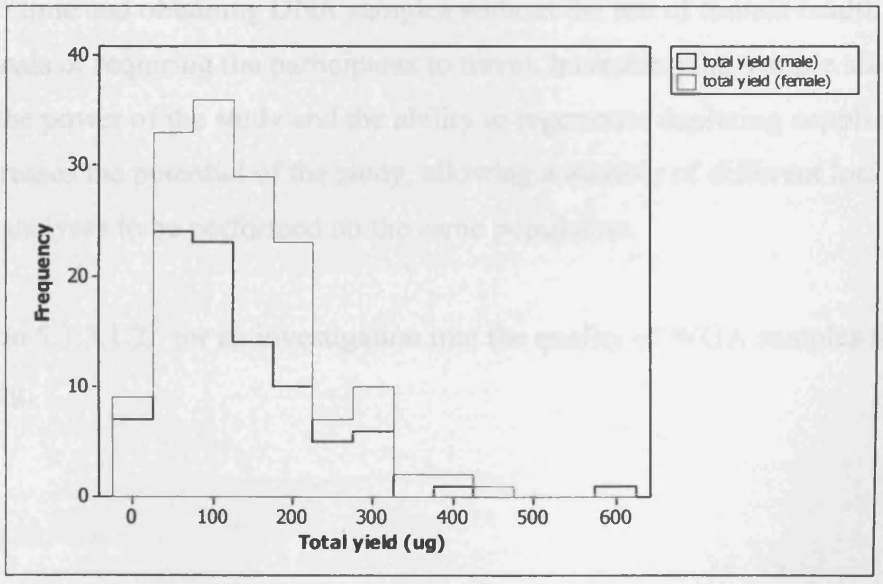


Figure 4.16. Distribution of DNA yields in males and females measured by spectrophotometry.

In summary, mouthwashes provide adequate quality DNA for the purposes of genetic analyses. The yield obtainable is highly variable between subjects and is affected by gender in the Irish-Welsh family, females providing a greater average yield than males. However, this is not supported by the other Family Study of Myopia participants. The Family Study participants provided a larger yield compared to the Irish-Welsh family members, as measured by spectrophotometry ($p<0.001$, Mann Whitney). This may mainly be due to the collection technique. Family Study members' mouthwashes were collected by post and instructions given to the participants specify the optimum time for performing the mouthwash to gain the highest quantity of DNA is first thing in the morning. The Irish-Welsh family's mouthwashes were collected either as above, but also during a family get together (chapters 2 and 3) which occurred towards the evening. It was at this time that the majority of mouthwashes were performed and could therefore contribute to an explanation of the overall lower yields found in these mouthwashes compared to those collected first thing in the morning.

If an inadequate quantity is obtained, whole genome amplification is suitable for increasing the yield of human DNA. Using mouthwashes as a source of DNA has the potential to ease the methodology of recruiting large numbers of subjects within short periods of time and obtaining DNA samples without the use of trained health professionals or requiring the participants to travel. Increasing the sample size will increase the power of the study and the ability to regenerate depleting supplies of DNA increases the potential of the study, allowing a number of different loci or different analyses to be performed on the same population.

See section 5.3.3.1.2. for an investigation into the quality of WGA samples for genotyping.

5. Linkage Analysis for Myopia in the Multi-Generational Irish-Welsh Pedigree.

5.1. Introduction

5.1.1. Linkage Analysis

Linkage is present when alleles at two loci are inherited more often than would have been expected by chance, due to the two loci lying close to each other on the same chromosome. Linkage analysis is the investigation of the cosegregation between a putative disease locus, i.e. the locus controlling the trait phenotype, and a marker locus. Specifically, observing the inheritance of the trait with regard to the marker genotypes. If a particular marker allele is transmitted along with the phenotype it may be that this allele and the disease pre-disposing allele are linked (chapter 1).

There are various different methods of linkage analysis, the appropriate method dependent on the assumed mode of inheritance of the trait, the penetrance and the disease gene frequency. One exceptionally large pedigree was recruited into the Family Study of Myopia cohort (Irish-Welsh pedigree). The large size and the higher than average myopia prevalence of this family prompted an investigation separate to that of the Family Study of Myopia's general strategy of examining a large number of small families within a combined cohort and so was investigated in isolation. Considering the pedigree and phenotype of interest, a variance components approach was deemed applicable.

5.1.1.1. Power and SLINK

For a genetic study, it is important to know if the recruited cohort is large enough to detect linkage, if it exists. In order to obtain an estimate of the potential power of this pedigree, a computer simulation program, SLINK, was used (Ott, 1989; Weeks et al.,

1990). SLINK simulates a genotype for one pedigree member conditional on the phenotype. Each following pedigree member is also assigned a genotype based not only on the phenotype but also considering the genotype assigned to the previous individual until all members of the pedigree have a genotype. This requires the input of pedigree structure information and disease parameters in standard pedigree and parameter files. The number of required replicates, a random seed number and the proportion of unlinked families (if allowing for heterogeneity) are also required to create a third input file. Other information required by SLINK regards the availability of phenotype and genotype data. A code is added to the end of the pedigree file to state whether the genotypes are available (i.e. if DNA is obtainable) and whether the phenotypic information is available or if it has to be simulated.

5.1.1.2. Variance Components Linkage Analysis

This type of linkage analysis is mostly suited to extended pedigrees and quantitative traits. The main benefit over other analysis methods is that no assumptions are made regarding the genetic model, i.e. it is a non parametric approach. It is most powerful when the trait distribution follows that of multivariate normality; a deviation from this distribution increases the type I error rate. However, when the distribution is highly leptokurtosed, it is the type I error that is inflated as opposed to the type II error (Majumder and Ghosh, 2005), the potential areas of linkage are therefore not necessarily excluded. The increase in power of variance components analysis occurs because all of the available information is used, including that from relatives who may not be affected by the disease or who have a comparatively low trait value. The pedigree is also seen as a whole entity and the analysis is not limited to analysing separate relative pairs (Almasy and Blangero, 1998; Williams and Blangero, 1999) thereby increasing the amount of information extracted from the pedigree.

Most variance component analyses make use of allele-sharing theory (chapter 1). Allele sharing methods are based on the theory that, for a quantitative trait, the more similar the phenotypes of a sibling pair, the higher the number of alleles shared identical by descent (IBD) at that locus.

5.1.1.2.1. SOLAR

Sequential Oligonucleotide Linkage Analysis Routines (SOLAR) (Almasy and Blangero, 1998) is a software package which applies a variance components method of analysis to investigate allele sharing within family members and is not restricted to relative pairs. As previously described (chapter 3) the variance of the trait under investigation can be partitioned into additive genetic effects and environmental effects, thereby calculating a heritability estimate for that pedigree. The hypothesis is stated that the variance of the trait is due to a major QTL acting on a polygenic background. This hypothesis is compared to the null hypothesis when only the polygenic background effects are modelled. The difference in the log likelihoods of these situations equates to a conventional LOD score (Almasy and Blangero, 1998). As the variance components method calculates the likelihood of the data at a certain variance estimate, the likelihood value (as well as the variance) must be positive values. The LOD score is equivalent to the difference in these values and therefore must also be a positive value. This contrasts to conventional LOD score analysis where negative values occur where linkage is not present. In variance components analysis the exclusion of linkage is indicated if the LOD score is around 0, rather than at a LOD score of -2 or below as in conventional linkage analysis.

Following on from this polygenic analysis, identity by descent matrices are calculated based on specifying the expected genetic covariances between relatives as a function of the IBD relationships at a QTL (Almasy and Blangero, 1998).

The IBD probability matrices are produced in two different ways, the choice of method is dependent on the complexity of the pedigree and data:

1. Based on a Curtis and Sham algorithm (Curtis and Sham, 1994) this method produces IBD matrices based on the probabilities that two relatives share the alleles considering the observed genotypic data.
2. A Monte Carlo method. Uses Monte Carlo algorithms to input marker genotypes for those with missing data conditional on the observed data and a maximum likelihood estimation method produces the IBD matrices.

Twopoint and multipoint linkage analyses can then be performed based on these IBD calculations. These also provide an estimate of the effect size, specifically the relative variance of the trait attributable to the underlying QTL (Williams and Blangero, 1999) and the residual variance attributable to other additive genetic effects.

5.1.2. Genotyping

This process involves identifying the genotype at each locus under investigation using the polymerase chain reaction with fluorescently labelled primers and differentiating between the size of the products and therefore between alleles (chapter 1, section 1.2.1.4.).

5.1.2.4. Genotyping Errors

It is important to identify errors in genotyping as any incorrectly ascribed genotype can alter the apparent transmission pattern within a pedigree and therefore affect the results obtained from genetic analysis. Errors are likely to decrease the power of both linkage and association studies and show an increase in recombination where none exists (Badzioch, Thomas and Jarvik, 2003).

5.1.2.4.1. Description and Sources of Error

In most genetic studies there will be a low genotype error rate. Generally it is around 0.5 to 1% but can be higher if the DNA template used is of low quantity or quality as this reduces the reliability of PCR amplification (Pompanon et al., 2005). Error rates are useful to calculate as they can produce a measure of how reliable the genotypes are and therefore the data as a whole. It can also identify particular markers or alleles which are more prone to errors and therefore the decision of whether the benefits of including that specific locus into the analysis outweigh the disadvantages of including potentially unreliable genotypes can be made.

Mendel's second law states that one allele from each parent is transmitted to an offspring. Any inconsistency from this law can be identified as an error in genotyping. However, even if a pedigree "Mendelises" well, it could be that there are genotyping errors not identified. It follows that by genotyping a highly informative marker, any errors will be highlighted easily but those present with a marker with low heterozygosity may not be as easy to identify. The only way to eliminate the possibility of including undetected errors is to perform further laboratory based work with the identified samples. This may include duplicating samples, either using replicates or by strategic reamplification at chosen loci that may be more susceptible to errors, or by comparing genotypes with other matched tissues. Alternatively the errors may be reduced by pre-testing the quality of the DNA (Hoffman and Amos, 2005). There is also the possibility, in large studies, that a mutation may have occurred between generations and that the detected error is not a true error (Pompanon et al., 2005).

Errors can be caused by a number of sources (Pompanon et al., 2005):

1. Mutation of the sequence either in the amplified product or in the template itself, leading to variations in allele sizes and, in the case of insertion or deletion of one base, difficulty with allele calling.
2. Preferential amplification of one allele, either the shorter allele or that one with a lower GC content.
3. Contamination or cross contamination of the sample.
4. Mislabelling of the sample.
5. Addition of an extra base at the end of the new product (section 1.2.1.4.2.).
6. Mislabelling of alleles due to a poor signal. Specifically, microsatellite markers can produce stutter bands by slippage of *Taq* DNA polymerase during PCR, especially in highly polymorphic markers (Hoffman and Amos, 2005).
7. Mispriming. The coding of alleles that are not true alleles but amplification products of the PCR.

5.1.2.4.2. PEDCHECK

PEDCHECK (O'Connell and Weeks, 1998) is a software package which identifies inconsistencies in the marker data according to Mendelian laws of inheritance. It requires the input of the pedigree structure and genotypic data which will then be run through a series of checking algorithms as described.

Firstly, the nuclear family algorithm will identify any inconsistencies between parents and offspring, known as Level 1 errors. Specifically it will identify situations in that:

1. The alleles of a child and parent are incompatible
2. The alleles of a child are compatible with each parent separately but not with both parents together.
3. There are more than the appropriate number alleles in a sibship considering the number of homozygous children.
4. For an X-linked locus a male is not coded as homozygous.
5. An individual has only one allele.

The next algorithm is not restricted to nuclear families but also extends the search for errors to more complex relationships (Level 2 errors). This genotype-elimination algorithm identifies an error and then eliminates the genotype at that marker for each member of the family in turn and outputs the possible genotypes the rest of the pedigree may have. This is compared to the observed genotypes to identify the individual with the error. This can sometimes prove to be difficult as there may be more than one individual with a genotyping error or the error comes not from a nuclear family but from an adjoining relationship. If this is the case there are two further algorithms which can be used (Level 3 and 4 error checking algorithms).

Critical genotypes are those which eliminate inconsistency if removed from the data, and are found using the level three error detection and the critical genotype algorithm. The subject with the critical genotype is likely to be the individual with the error, however if there is more than one and there is no way to identify the erroneous genotype, an odds ratio statistic is used (Level 4 error). This nominates alternative genotypes for those subjects with critical genotypes. Likelihoods for the pedigrees are

calculated using each of the genotypes and a ratio of the highest likelihood to each of the others is produced. The best supported genotypes will have a likelihood ratio close to 1.

5.1.3. Exclusion loci

A number of systemic diseases are associated with myopia (section 1.1.3.3.). In order to restrict the potential genetic basis of the trait under investigation, it is necessary to exclude linkage to these disease loci. Specifically, mutations in the gene FBN1, which encodes fibrillin, a protein found in the connective tissues, are known to cause Marfan's syndrome (OMIM #154700). Also mutations in the COL2A1 and COL11A1 genes, which code for collagen types II and XI, give rise to Stickler's syndrome type I (OMIM #120140) or type II (OMIM #604841) respectively. Including the genotyping of markers at these chromosomal loci allows an extra filter to exclude these syndromes as a causative factor and is the method used in other myopia studies (Young et al., 1998b; Farbrother et al., 2004b). Larger scale studies may require alternative methods using SNPs to detect individuals with mutations at these loci. Popular high-throughput methods for mutation detection include a method using fluorescence resonance energy transfer detection and DNA microarray genotyping based method (Shi, 2001).

5.1.4. Candidate Genes

These are genes which are presumed likely to control the disease and are suggested due to their position or biological function. Polymorphisms often occur within a gene and the probability that they affect the phenotype is often dependent on whether they will alter the function of a protein by causing an amino acid sequence change or a premature stop to the protein synthesis, these changes being the most likely to cause disease associations (Tabor et al., 2002). Amino acid substitutions are the most frequent type of mutation, at around 59%. The importance of that amino acid in

relation to the function of the protein is the relevant issue when considering the clinical outcome (Botstein and Risch, 2003).

5.1.4.1. Myopia candidate genes

Despite the numerous loci identified by linkage to high myopia (see section 1.3.), there have not been any causal genes specifically identified.

At the MYP1 locus, cone pigment genes were sequenced and compared to the sequences of unaffected individuals, but no mutations were found in affected individuals suggesting that colour vision defects were not related to the underlying disease (Young et al., 2004a).

MYP2 is the most intensively investigated locus with respect to candidate gene sequencing. Nine positional candidate genes were identified for this locus, including two structural genes and six transcription factors. However, none of the polymorphisms found in the MYP2-linked pedigrees were found to segregate with the disease (Scavello et al., 2005; Zhou and Young, 2005). One of these candidate genes, TGIF-β, a transforming growth factor related protein, was excluded as a candidate gene (Scavello et al., 2004) despite being previously found to be associated with a susceptibility to high myopia in a Hong Kong Chinese population (Lam et al., 2003). These differences may be due to the differences in ethnic background, as suggested by Scavello and colleagues as the wild type DNA sequences of the two cohorts were not concordant (Scavello et al., 2004).

The MYP3 locus incorporates a number of possible candidate genes. The first to be identified was phenylalanine hydroxylase (PAH) which is involved with metabolism. This was used as a marker in linkage analysis and resulted in a high LOD score although its physiological role makes it an unlikely candidate gene for myopia (Young et al., 1998a). The small leucine-rich proteoglycans (SLRPs): decorin, lumican and fibromodulin genes also reside within this locus and are expressed in ocular tissues. They have a structural function in that they interact with collagen

fibrils. Lumican and fibromodulin were both excluded by sequencing affected individuals: no sequence mutations appeared to segregate with the disease (Paluru et al., 2004).

The locus on chromosome 7, MYP4, does not contain any strong potential candidate genes based on their function alone (Naiglin et al., 2002). However, MYP5 contains four genes which have been evaluated, two of which are involved in connective tissue disorders (COL1A1 and chondroadherin) and two which have been found to be involved in eye development (TBX2 and TBX3). Again, none of the polymorphisms found segregated with the high myopia (Paluru et al., 2003; Young et al., 2004b). The region around PAX6 was found to be linked to myopia (MYP7) but no association was found with a selection of haplotype-tagging single nucleotide polymorphisms (tSNPs) across the region (Hammond et al., 2004). Tagging-SNPs are highly informative polymorphisms selected from a group along a region of chromosome in which there is elevated linkage disequilibrium (Weale et al., 2003). Due to this high level of linkage disequilibrium, where SNPs on the same chromosome are inherited together more often than by chance, typing all of the SNPs will not increase the information compared to genotyping only those selected (tSNPs). By genotyping only the tSNPs within this region, the effect is that of typing a haplotype. Rare recombination events will enable further discrimination between non causal and causal SNPs once an association is found. PAX6 (MIM #607108) lies within the linked region on chromosome 11 and is known to be involved in oculogenesis. Due to the known functionality of the gene and the high linkage signal, this is potentially a strong candidate gene within this region. It is possible that the causal variant may simply have been missed by the tSNPs used, or it could be that the causal variants lie outside of the PAX6 gene itself (Hammond et al., 2004).

In summary, many positional candidate genes have been excluded but no causative genes have yet been found.

5.2. Methods

5.2.1. Power calculations and SLINK

SLINK requires the trait to be dichotomised and is also unable to handle large extended pedigrees. Thus, only a small section of the family could be used for the power calculation to produce an estimate of the minimum power of the total pedigree.

A sub section of 41 family members from Irish-Welsh pedigree was used for this power calculation (figure 5.2.). Subjects were assigned as affected if the refractive error was −5.00D or more in the least minus meridian of the right eye, unaffected if less and unknown if no phenotypic data was available, as used for the analysis detecting MYP5. Using a threshold of −6.00D would further decrease the power of this simulation as the number of affected individuals in the available pedigree section would be decreased. Dominant inheritance model parameters used were: disease gene frequency of 0.0133, 100% penetrance, no phenocopies and eight alleles all occurring with equal frequency at a co-dominant marker. Mutation rate was 0, with no sex difference. This model was used by Young and colleagues for the detection of MYP2, 3, and 5 (Young et al., 1998a; Young et al., 1998b; Paluru et al., 2003). No heterogeneity was allowed for. Power was calculated by simulation with 250 replicates.

This was repeated using the parameters set out for a linkage analysis study for common myopia (Stambolian et al., 2004). Affectation status was classified as a mean spherical equivalent (spherical component of refractive error plus half of the cylindrical component) of −1.00D and over as affected with an age based adjustment for individuals under 20 years old (figure 5.5). Twelve different models were used using penetrances of 0.584, 0.8 and 0.9. Phenocopy rates of 0, 0.05, 0.1 and 0.15 were also used. Gene frequencies and mode of inheritance were the same as described above.

5.2.2. Genotyping

Primers were designed and optimised to their PCR parameters by performing PCR as described below and running the products on an agarose gel with a size standard (figure 5.1.) on a selection of unrelated samples known to amplify well. Those with weak product bands or double bands were re-optimised or re-designed.

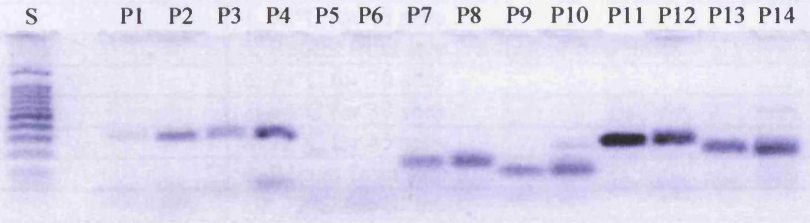


Figure 5.1. Example of a gel for testing primers prior to genotyping. Pn shows each separate primer pair and S is the size standard. Primer P10 will be re-optimised due to a shown double band, P5 and P6 will also be re-optimised due to weak product bands.

Forty nanograms of DNA were added to 8µl master mix solution to give final concentrations of 1 x PCR buffer (Promega), 200µM dNTPs, 1.5mM $MgCl_2$, 1µM fluorescently labelled upstream primer, 1µM downstream primer and 0.3 units *Taq* DNA polymerase in each well. PCR was set for programs as shown in table 5.1. and cycling was performed in a MJ Research DNA Engine DYAD. Primers used are shown in table 5.2. Forward primers were fluorescently labelled with FAM, HEX or NED as indicated.

Table 5.1. PCR programmes where *n* is the number of cycles.

NAME OF PROGRAM	PARAMETERS
58TD*n*	1. 94°C for 3 mins
	2. 94°C for 20 secs
	3. 62°C for 30 secs, reduce by 1°C per cycle
	4. 72°C for 45 secs
	5. Repeat from step 2 a further two times
	6. 94°C for 20 secs
	7. 58°C for 30 secs
	8. 72°C for 45 secs
	9. Repeat from step 6, *n*–1 times
	10. 72°C for 10 mins
62TD*n*	1. 94°C for 3 mins

	2. 94°C for 20 secs
	3. 66°C for 30 secs, reduce by 1°C
	4. 72°C for 45 secs
	5. Repeat from step 2 a further two times
	6. 94°C for 20 secs
	7. 62°C for 30 secs
	8. 72°C for 45 secs
	9. Repeat from step 6, $n-1$ times
	10. 72°C for 10 mins
64TD*n*	1. 94°C for 3 mins
	2. 94°C for 20 secs
	3. 68°C for 30 secs, reduce by 1°C
	4. 72°C for 45 secs
	5. Repeat from step 2 a further two times
	6. 94°C for 20 secs
	7. 64°C for 30 secs
	8. 72°C for 45 secs
	9. Repeat from step 6, $n-1$ times
	10. 72°C for 10 mins
60*n*	1. 94°C for 3 mins
	2. 94°C for 20 secs
	3. 60°C for 30 secs
	4. 72°C for 45 secs
	5. Repeat from step 2, $n-1$ times
	6. 72°C for 10 mins

5.2.2.1. Preparation for genotyping analysis

PCR products were pooled into their respective "gel sets" (see table 5.2.). Three microlitres of this diluted sample was added to 0.5µl size standard (Genescan 500 ROX; Applied Biosystems) and 8µl HiDi formamide reagent (Applied Biosystems). The samples were loaded into the ABI Prism® 310 Genetic Analyzer and run on program D. Genotyper software (ABI Prism®) was used to call the alleles.

Table 5.2. Marker, primer and PCR product descriptions.

Region	Marker Name	Position (cM)	No. Alleles	Heterozygosity (Mar)	Forward Sequence	Reverse Sequence	Range	Gel Set	PCR program
MYP1	DXS8106	162.62	11	0.85	GCCTTCTCAAAGTCAACAGGTA	GTTTCTTGGAATTCAGAATGGCTGCTAT	156-176	8	58TD30
	DXS8028	169.72	10	0.90	GATGACACTCGGACTGCTCTT	GTTTCTTCCTCACAAAGCTTTGGGATT	320-346	6	58TD30
	DXS998	173.40	4	0.78	GCAAATTGATGGGGAACTT	GTTTCTTCCAGGACTACTTAAAACTGTCCTGT	340-346	3	58TD30
	DXS8069	177.25	6	0.80	CCACTCCAAATTGTTGGGTAT	GTTTCTTGGGAAAGCACAAGTTTCAGAA	162-172	4	58TD30
	DXS8103	181.08	8	0.88	GCTTCATGGAGCTAAAGGGAA	GTTTCTTCCAAGGTGGGAGGATCACTT	190-206	4	58TD30
	DXS8087	184.34	5	0.82	GGCTGCCCTTTTGTTCAA	GTTTCTTCTGGCAAGGACGCATACAT	322-332	4	58TD30
	DXS1073	184.34	5	0.89	CGCCTTGCTAGAGAGACATGTA	GTTTCTTCCTGAAGATGACTCACTCACTT	246-262	4	58TD30
MYP2	D18S59	0	8	0.82	GCACTCTCTAAATCTTGCCTT	GGAACTGGTACTCTGAAAGAGA	157-173	10	58TD30
	D18S476	2.84	7	0.76	CCAGTATGATGGTGAAATCCT	CCTTAGTTCCTCTCTACAGAAGCA	283-295	11	58TD30
	D18S481	6.94	9	0.75	GTTTGCTGAGGGTCAGCAT	GCCTCAGCCTTCCAAAGT	189-211	15	64TD32
	D18S63	8.30	10	0.79	GGAAACATGTTCACCAAATCA	CTATGTCACCATCTGGCTGT	260-284	12	58TD30
	D18S1132	11.21	6	0.69	GTTTCTTTGCTGCCCTACCAGACCAA	AGCCACAACGACCAGCCTTT	120-130	10	58TD30
	D18S452	18.70	10	0.84	GCAAACCTATCCTGAAGTTCA	CAGATGAACCTGGAATGGT	102-131	11	58TD30
	D18S1163	24.08	7	0.59	GTTTCTTCCCACCCTTATGACTTCATTTAACC	CGCCACACACTCTCACACAC	201-217	10	60TD32
	D18S464	31.17	6	0.63	CCAGACTTTGTGCCATTTCT	CCTGAATCTCTTGTGGTTTGT	276-286	15	58TD28
	D18S1158	38.92	3	0.66	GCATCTATGCAGTGCCAAAT	GTTTCTTCACCAGGATGCCAGCC	103-107	10	58TD30
MYP3	D12S80	83.19	7	0.77	CCAGCCTGGAATGATATGTA	GAATGTCAATGGACCAGATG	205-219	11	58TD30
	D12S326	86.40	8	0.80	CCACTAAGATGCAGCTCAGTA	GCTAGTGTTGAATGGCTTCC	292-316	13	58TD30
	D12S64	89.42	8	0.59	CCAGATTTTGACCACTTCTCT	CGCAAACATGGAGAGAAC	132-146	13	58TD30
	D12S316	94.49	6	0.56	CGTGGATGATGCAATAGTCAGTA	GTTTCTTCCTCAACTACCTCCCTTTCACTT	172-184	13	58TD30
	D12S348	100.92	6	0.68	CAGATCTGCTCCTGGAATCAA	GTTTCTTGCAGGAAAGCTGGTTAGGAGAT	297-307	16	58TD30
	D12S332	105.18	7	0.64	GAAACTATTGGGCTGCTGA	AGTCGTAATTGGGAACAAAA	325-343	10	58TD30
	D12S1607	107.86	6	0.75	AGCTGAGATCATGCCACTG	TTGGTGAGCCCTGAAGAA	134-152	15	6034
	D12S78	111.87	13	0.91	CCTCTGTTGCAACTATCTTGAA	GCTTTGCAGCACCATGTA	192-216	12	58TD30
	D12S1605	116.66	5	0.78	CAGGGATCTCACAATCTGAAA	CTGAGCCTTGGAGGGTATG	193-201	16	58TD30
	D12S354	123.77	6	0.73	GGTGGTTCTGGGTCAGAT	GGTTTCCTAATTTCAAGTCAA	183-197	13	58TD28
MYP4	D7S798	168.98	7	0.79	GCTTCAAAGAGCTGCTCAA	GTTTCTTCCGATTACTATAGTCAAGCGAA	246-260	5	58TD30
	D7S2546	173.03	5	0.76	GCAAGTAAGAGGGTAAGGAGGTT	GTTTCTTCCTTACTGGAGAACATGCAAGAA	310-318	3	58TD30
	D7S550	178.41	13	0.81	CCTGATAGGCAGTTGGGTT	GTTTCTTCACTCCTAGGCATTTACTGACA	146-192	9	58TD30
	D7S2465	180.24	14	0.82	GCAGCTATCTAAGCACGCTT	GTTTCTTGGAGACTGAGGCAGGAGAAT	306-352	2	58TD30
	D7S2423	181.97	5	0.71	CCTTCTCTGGGAGTCTCTGTT	GTTTCTTCCTGATGAGCTGCGATCTT	210-222	7	58TD25

Table 5.2 contd.

MYP5	D17S1868	64.16	7	0.73	GAGGATGTCGAGGCAGA	GCTTGAGCCTAGGAGTCAA	127-139	14	58TD30
	D17S1795	68.44	7	0.71	CCATGAGGTTGATGGGA	GAGCAGGCAGTCTGCAA	219-231	14	58TD28
	D17S956	73.62	6	0.63	CCTGAAGGCTTGGAGCTT	GGAGTAATGAAGGAGGGAAAG	231-243	12	58TD30
	D17S1853	80.38	13	0.81	GACATGGTGAATGTTTTCCTAG	CCAGTTCTCAATGCTTCTTAGT	279-318	10	58TD30
	D17S1290	82.00	11	0.84	CTGTGCCCTCTAATCTTTAGGA	GTTGAGGCTTCAGTTAGCCA	152-192	11	58TD30
	D17S942	85.94	4	0.66	GTGACATGGTATCATGAGTACCT	GGTAAAGAGCTGATGCCAA	275-281	16	58TD30
	D17S2059	93.27	4	0.60	GCAGTGATTTTCACATTGCTT	CTGTATCTGAACCAACTGTATCACA	209-229	13	58TD28
	D17S1831	97.60	9	0.81	GGAAGCTCTGACTTCGGTTCT	GTTTCTTCTCAGCCTCCCAAAGTGCT	327-349	12	58TD30
	D17S1817	103.53	9	0.77	GGGTGATGAAAGCAATCTGAA	GTTTCTTGGTTGGCAGTGAGCCAAA	168-192	12	58TD30
PAX6	D11S2368	22.56	9	0.79	CCCCAACTATGATGCCTCCTT	GTTTCTTGGTGCAAGAATGTCTGTCTGTCTA	277-313	1	58TD32
	D11S1308	24.73	3	0.54	GGCAGAAGTGGTTTCAGACAGTA	GTTTCTTGGAGGAGAGTAGGAAACAGGAAA	197-201	6	58TD30
	D11S4114	26.98	10	0.77	GCAGACCCAGAGAGATGACAT	GTTTCTTGGAGCTGAGCTGTGTCATATCTT	279-307	9	58TD35
	D11S992	30.88	7	0.60	CCTGTCTCATAGGGTAGTTGTGAGA	GTTTCTTCCACAGATGCCAAAGACATAGAT	305-319	2	58TD32
	D11S904	33.57	8	0.82	GGCTGCTTGAATCTGAATCCTA	GTTTCTTGGCTCACACTATGACAAGCAAT	264-282	2	58TD30
	D11S4152	35.21	13	0.89	CCCTCCAGAATACTCTTCTATCCTT	GTTTCTTGGCACCAACATTCAGACCTT	113-145	1	58TD30
	D11S4154	37.62	8	0.79	CCCTAACCTCCATGTTGCTT	GTTTCTTGGCTTTCTCAGAGCACTTCAA	140-157	16	58TD30
	D11S1751	40.12	14	0.87	GCAGCCTTTAGTTTTGCCTA	GTTTCTTGGCTGGAATCAGCTCACTT	268-306	3	58TD32
	D11S4200	42.55	10	0.74	CCCTCTGAGTCTAATGCTACTCTTT	GTTTCTTGCCTTGCTGTTCTGTCTCTTT	232-256	6	58TD30
	D11S4203	45.94	11	0.76	GCAGCAGGCTGAACACATT	GTTTCTTCCAGGATGCTGGAATAGAGAA	156-207	5	58TD30
	D11S4083	47.06	14	0.86	GGACATTCTGCAGGGACTCTT	GTTTCTTGGTTCCTTAGACTCAGATTGGGAT	206-234	2	58TD30
	D11S1330	49.73	3	0.23	GCTTTACAGCCAGATGCTGTA	GTTTCTTGGATACGGACCACTATGTAGGTA	246-250	7	58TD30
	D11S905	51.95	9	0.72	CCTCACACACACCTCTGTCTATT	GTTTCTTGGTTTCAAAGACCTGGACAA	120-140	9	58TD30
	D11S1993	54.09	8	0.77	GGGATAATGATAACACCCTTCA	GTTTCTTGCTCTGGTCCTATAGCTGCTTA	225-249	16	58TD30
	D11S1361	56.76	5	0.50	GCATATCCTGAGTGGTGGAAA	GTTTCTTGCTTCCCTCTGCTGGACAA	212-227	17	58TD30
MYP6	D22S310	23.37	9	0.64	CCCGAAAGGAGTCTCAGTTT	GTTTCTTGTCTGTCTGTTTCCAGAGTTT	184-198	10	58TD30
	D22S1167	24.74	5	0.74	GGGCTTCAACAACATTCTTA	GTTTCTTCCTATAGTTTAGTCCCAGCTGCTT	213-225	9	58TD28
	D22S1144	27.48	9	0.76	CCAACTGTTGCGTTGAAGTT	GTTTCTTCCACAAACACACACATGCTT	139-159	3	58TD30
	D22S689	28.57	8	0.76	CCACCAGTCCTACTCCTCTTTT	GTTTCTTCTGCCTGCCTGCCTATCTAT	261-317	14	58TD30
	D22S1176	29.66	10	0.67	GCTTCCTGGAAGTCCTGAAA	GTTTCTTGCCAAAAGTAGATCTGTGCCAT	128-152	4	58TD30
	D22S280	31.30		0.82	CCAGTCTCTTAGGACAAGCAA	GTTTCTTGGGAAATGCAGGACACATT	257-281	10	58TD30
	D22S685	32.39	9	0.79	GCTCTCCAGCTCTCAGACTTT	GTTTCTTGGAAGGGATCCTAGTTATCACCTA	270-334	5	58TD30
	D22S424	33.76	5	0.55	GGAATGAGTGTCAGTCACTGGAA	GTTTCTTGGGACCAGCAAGCACAAA	93-109	2	58TD30
	D22S683	36.22	13	0.90	GGATAGAGCGAGACTCTGTCTCAA	GTTTCTTGGTGGAAATGCCTCATGTAGAA	197-233	3	58TD30
	D22S1173	37.82	3	0.46	GGTGTCCAACTTCATCCAGATT	GTTTCTTGCCAACAAGTGGATGAAGAAA	158-162	10	58TD30

Table 5.2 contd.

	D22S283	38.62	13	0.90	GCTGCTTGCTTCCTTCTGAA	GTTTCTTCCAGCATCATCATCTACCACTT	182-206	1	58TD30
	D22S426	41.42	6	0.61	GGGAGGTGAAAGTGGGAA	GTTTCTTCCGTTATCTCTAGCTAGTGGGTT	277-287	6	58TD35
	D22S1045	42.81	5	0.64	CTCTGTAGGTGGCCTGGTT	GTTTCTTCCAGTTCCTCTCCACCCTATA	360-378	9	58TD30
	D22S445	45.82	7	0.64	CCACTGTCTGTCCATCCGTT	GTTTCTTGGATGAACAGAATCAGGATGAA	165-187	14	58TD30
	D22S1157	47.31	8	0.85	GCACTTCAGCCTCGTCAA	GTTTCTTGGTCCCCTAGTCCTACATCAA	369-385	8	58TD30
	D22S1179	48.19	9	0.78	CCACTGCACCTGACCCTAGAT	GTTTCTTGCCTCAGGGAATCTATGACTCAT	172-188	16	58TD32
	D22S1168	50.09	5	0.42	GGGATACTTCGCTTGTGAGGTA	GTTTCTTGGCACGAGAAGCACTTGAA	235-251	1	58TD28
MYOC	MYOC1		5		GCACAGTGCAGGTTCTCAA	GTTTCTTCCGAGCTCCAGAGAGGTTTA	369-377	2	58TD30
	MYOC2		5		CCAACCATCAGGTAATTCCTT	GTTTCTTCCTCAAAACCAGGCACAA	116-125	7	58TD32
FBN1	D15S1003	47.85	6	0.74	TGGTAGTACCCCTGGATACCTG	AATCTTTGTGGATATGGCTCTGCT	195-212	14	58TD32
	D15S992	45.62	11	0.80	GGTTGTTTGTATATCTTCTCTCTTTC	GCCACCTTGATAGTATGTTAGCA	84-112	12	58TD30
COL2A1	D12S1701	62.54	5	0.75	GTAGAGGTCTTACAACTCCTTGGT	GTTTCTTGCTGAGAACTAAATCAAGAACTCA	125-137	13	58TD30
	D12S361	64.96	7	0.78	CTCTTAGNCCCACTCTGTAG	AAATATATGTGACCTAAATGGCAT	241-255	11	58TD30
COL11A1	D1S2626	136.34	9	0.72	GCAGATGACCTTCCGTTCAA	GTTTCTTCCCTGGTGTACTGTTCAGAGTACT	292-310	4	58TD28
	D1S2888	136.88	6	0.74	GAATGTCCCTTTCATTGCTA	CCACTTTGATCTAAAGTGCCT	278-288	13	58TD32
Chr 2	D2S2297	232.36	8	0.87	GCTTAGCTTCTAGTGCTGCTTTT	GTTTCTTCCACTGAGGCTTGAGGATTA	109-125	17	58TD30
	D2S2317	233.62	11	0.81	GCAATACAGCGAAACCCCAT	GTTTCTTGCTGGAAAAGACAAGGAAACAA	159-191	7	58TD30
	D2S172	235.07	19	0.92	GGCAAAGGCACAACTGCTTA	GTTTCTTCCATCCAGGTTGCTGTGAAT	262-302	8	58TD32
	D2S2193	236.70	13	0.90	GCTGTCATCAGAATCCAGTCTT	GTTTCTTGCTTCTGGTCATCTCATGGGTA	370-402	17	58TD30
	D2S2344	238.33	10	0.79	GACCTGCTAGGATGTCATTAGATTT	GTTTCTTGCTGCTTACAGAAGGAAAGGAT	247-273	18	58TD30
	D2S206	240.79	10	0.79	GGTCTGTTACATTGGTTCTCATT	GTTTCTTGGCTTTTGGTTAAGCACAGAA	200-222	18	58TD30
	D2S2348	242.17	18	0.90	CCTCAGTCTCTTGGAATGCTT	GTTTCTTCGAGTCCAGAACAGTGGAAA	257-313	7	58TD30
	D2S2205	243.01	5	0.69	CAGGAACTGGGAGACAAGATAGAT	GTTTCTTCCTCATCATAACACTCATCCCAA	150-162	2	58TD28
	D2S336	245.44	8	0.81	CCAGCCTGGTAAGCTTTTTCTA	GTTTCTTGGCACTTGGACTTCTTATTGTT	192-221	17	58TD28
	D2S2973	247.85	6	0.75	CACAAGGTACCACGACACAAGTA	GTTTCTTCCTTGCATCTTTACCAACACTT	123-144	18	58TD30
	D2S2202	249.22	5	0.66	CGTTACTACAGTGCTAGGACTGAA	GTTTCTTGGGAACTTGTGATTCCATGAA	149-157	18	58TD25

Key

Colour	Fluorescent Dye
	HEX
	FAM
	NED

Red= Presumed SNP in amplimer

Orange=not able to optimise

5.2.3. Checking for Mendelian errors

Prior to the linkage analysis, the data was analysed with the PEDCHECK program to identify any Mendelian errors. Any errors found were rechecked for allele mis-specification and if still found to be erroneous they were repeated or removed from the analysis for that marker.

5.2.4. Linkage analysis: SOLAR

SOLAR analysis was carried out using data regarding the pedigree structure along with age, sex and phenotypic data of all of the family members. Marker genotypes were inputted along with genotypic map information of the markers within the loci under investigation. Input files were created from this information and a LINUX workstation was used to carry out the analysis.

Due to the non-normality of the phenotype, a transformation was considered desirable to alter the frequency distribution to that as close as possible to normality. Standard transformations (e.g. log, square-root, Box-Cox) were not effective in removing non-normality. Therefore an alternative approach was adopted using normalised deviates as used by several studies for a non-normally distributed phenotype (Fox et al., 2004; Yalcin et al., 2004). Specifically, the transformation involved simulating a normal distribution of trait values with the mean of 0 and a standard deviation of 3. The original and simulated values were arranged by rank. An equivalent quantitative trait phenotype (q2) was then assigned to each subject dependent on their ranked position. Analysis was performed with both the original and the transformed phenotypes (q1 and q2, respectively).

The analysis was performed with different allele frequency settings. Equal allele frequencies were calculated and used in the initial analysis. Next, allele frequencies calculated using maximum likelihood were performed by SOLAR. Due to computational constraints, this process was extremely lengthy when there was genotypic data missing and for highly polymorphic markers (>5 days per marker in

some cases). Therefore an alternative method was used; the allele frequencies were calculated from all of the married-in individuals and pseudo-founders of the pedigree. The founders consisted of individuals as close in pedigree structure to the actual founders as possible. Married-in relatives were used to provide a representation of the population's allele frequencies and reduce bias due to the fact that the related individuals may inherit the alleles IBD and so the population is comparatively enriched with the alleles of its founders.

SOLAR is unable to analyse X chromosomal data using the same method as for the autosomal chromosome data. Unfortunately this alternative method could not handle the pedigree in full. To overcome this problem, the pedigree was divided into several smaller families and the IBD files were produced using the Curtis and Sham algorithm. A twopoint analysis was then completed.

5.3. Results

5.3.1. Pedigree and phenotype statistics

The trait distributions of q1 and q2 (the non-transformed and transformed values of refractive error, respectively) are shown in figure 5.2. As is expected for refractive error, the distribution of q1 is leptokurtotic at emmetropia (k=3.85) and skewed towards myopia (s=−0.25) (non-normal, Kolmogorov-Smirnov, p<0.01). The transformation by rank, q2, shows a normal distribution (k=−0.41, s=0, normal, Kolmogorov-Smirnov, p>0.15).

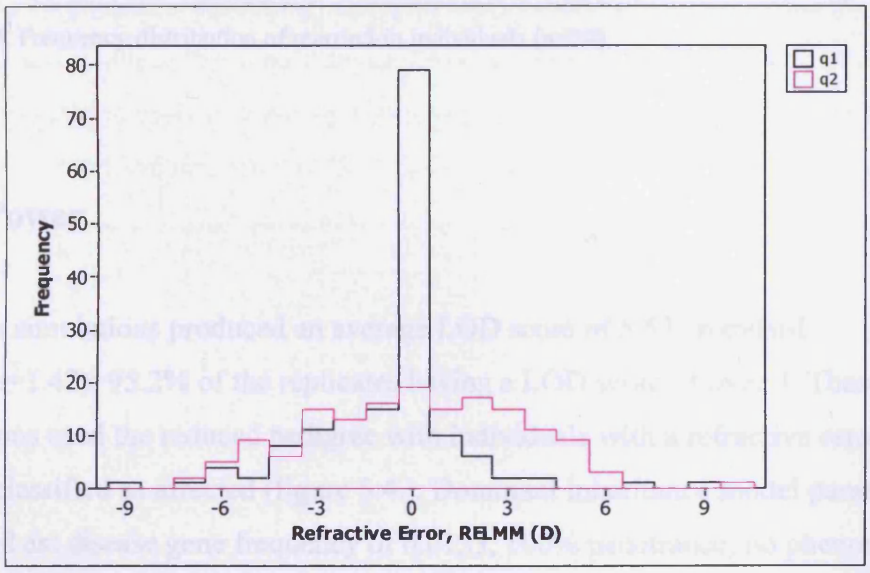


Figure 5.2. Frequency distribution of traits q1 and q2. Q1 represents the refractive error in the least minus meridian of the right eye. Q2 represents the rank transformed q1 data.

Considering the original data, q1, 37.8% of phenotyped individuals were myopic (≤−0.50D) and 3.2% were highly myopic (≤−6.00D). Emmetropes (≥−0.50D and ≤+0.50D) made up 46.8% of the pedigree and the remaining 15.4% were hyperopic (>+0.50D).

Married in individuals are highlighted in red in table A of the appendix. The trait values of whom appear to be skewed more towards hyperopia (γ_1=1.64) but remains leptokurtosed at emmetropia (figure 5.3.).

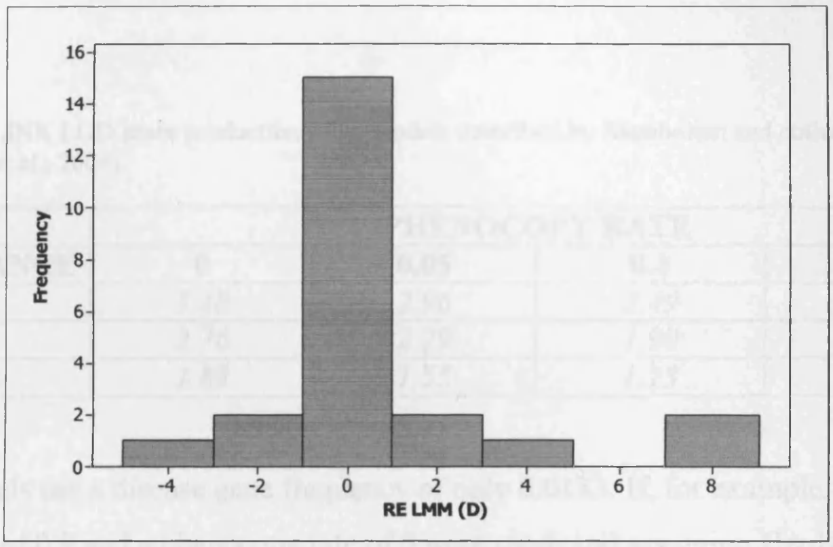


Figure 5.3. Frequency distribution of married-in individuals (n=24).

5.3.2. Power

SLINK's simulations produced an average LOD score of 5.53 (standard deviation=1.42); 95.2% of the replicates having a LOD score of over 3. These calculations used the reduced pedigree with individuals with a refractive error of −5.00D classified as affected (figure 5.4.). Dominant inheritance model parameters were used as: disease gene frequency of 0.0133, 100% penetrance, no phenocopies and eight alleles all occurring with equal frequency at a co-dominant marker. Mutation rate was 0, with no sex difference. These results suggest a high power to detect linkage from this reduced pedigree using the dichotomised trait. For the complete pedigree, there would be an expectation of increased power further with the increased size of the pedigree and the extra information of the quantitative phenotypes.

The simulations using the models described by Stambolian and colleagues produced LOD scores ranging from 0.99 to 3.48 (table 5.3.) for this pedigree (figure 5.4.). This

pedigree is less powerful for linkage when using common myopia as a phenotype, but would still potentially detect a linkage signal if the disease penetrance was 0.9 and the phenocopy rate 0 and be suggestive of linkage with penetrance of over 0.8 with phenocopy rates under 0.05.

Table 5.3. SLINK LOD score production using models described by Stambolian and colleagues (Stambolian et al., 2004).

	PHENOCOPY RATE			
PENETRANCE	**0**	**0.05**	**0.1**	**0.15**
0.9	*3.48*	*2.96*	*2.49*	*2.16*
0.8	*2.76*	*2.29*	*1.90*	*1.62*
0.584	*1.83*	*1.55*	*1.13*	*0.99*

These models use a disease gene frequency of only 0.0133. If, for example, a penetrance of 0.9 and a phenocopy rate of 0 were used, and assuming Hardy-Weinberg equilibrium, this would suggest a common myopia prevalence of only 2.4%, whereas it is known to be approximately 25% (chapter 1). These results are, therefore, only an approximate guide to the power of this study.

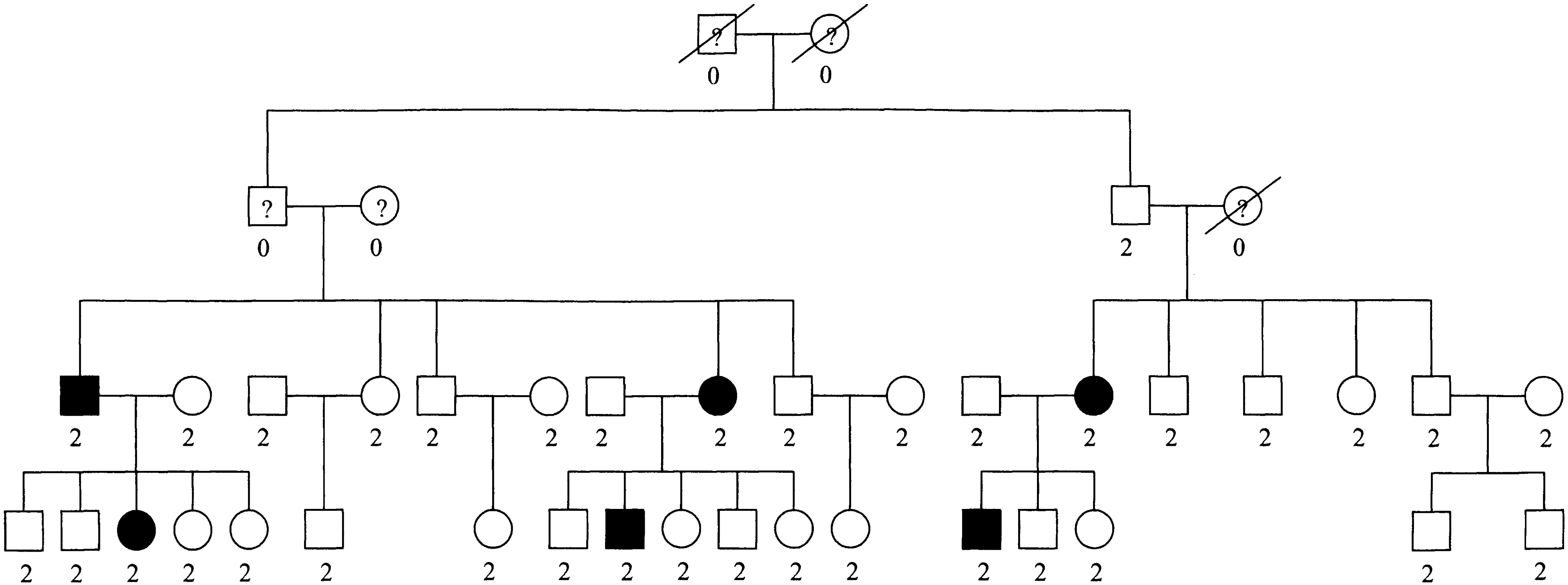


Figure 5.4. Reduced pedigree for power calculation. Filled symbols denote subjects with myopia of more than -5.00D in the least minus meridian of their right eye and those with a question mark have an unknown phenotype. Those individuals coded with a 0 have no available genotypic information and an unknown phenotype. Those individuals coded with a 2 have genotypic and phenotypic information available. Individuals with a diagonal line through the symbol are deceased.

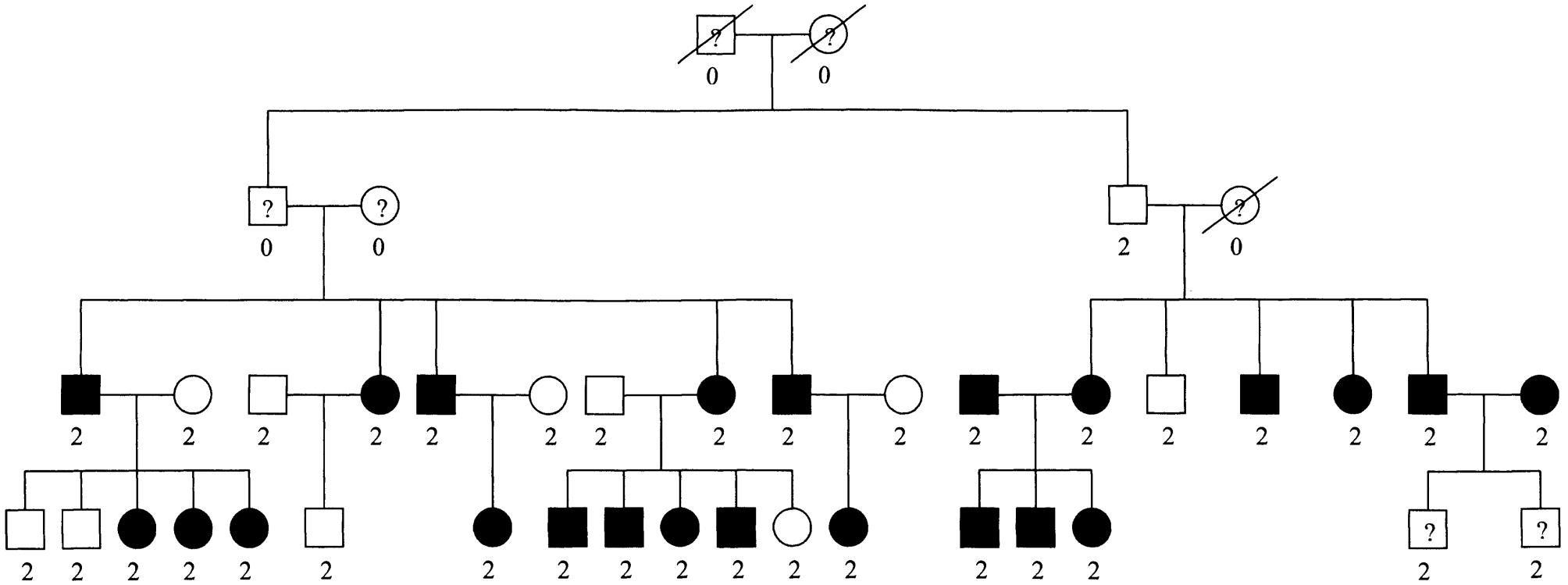


Figure 5.5. Reduced pedigree for power calculation (Stambolian model). Filled symbols denote subjects with myopia of more than -1.00D mean spherical equivalent and those with a question mark have an unknown phenotype. Those individuals coded with a 0 have no available genotypic information and an unknown phenotype. Those individuals coded with a 2 have genotypic and phenotypic information available. Individuals with a diagonal line through the symbol are deceased.

5.3.3. Genotyping

5.3.3.1. Genotyping procedures

From an initial microsatellite check performed by running subjects' PCR products on the sequencer, 60.0% of reactions needed no optimisation and produced clean signals and therefore accurate genotyping information. Other markers needed only minor changes to the PCR protocol: either more (10.6%) or less (11.8%) PCR cycles. Due to the failure to produce a clean signal from the fluorescently labelled PCR product, 12.9% of microsatellite markers were redesigned (final primer sequences are shown in table 5.2.). Only 4.7% had to be excluded due to presumed single nucleotide polymorphisms occurring within the amplimer and therefore potentially complicating the allele coding. After redesigning and reoptimising the previously unsuccessful markers, a further 4.7% of markers (approximately 36% of those redesigned) were excluded due to continued failure to obtain a PCR product of the expected length and therefore a clear signal for genotyping.

5.3.3.1.1. Genotyping success rates: subjects

Subject success rate for producing genotypes at all markers are shown in figure 5.6. where success rate is calculated by dividing the number of genotypes produced by the number of reactions carried out. Following the initial assessment of DNA quality (chapter 4), the expectation was of a high rate of success for all subjects. In reality, 79.5% of subjects could be genotyped for at least 75% of all markers.

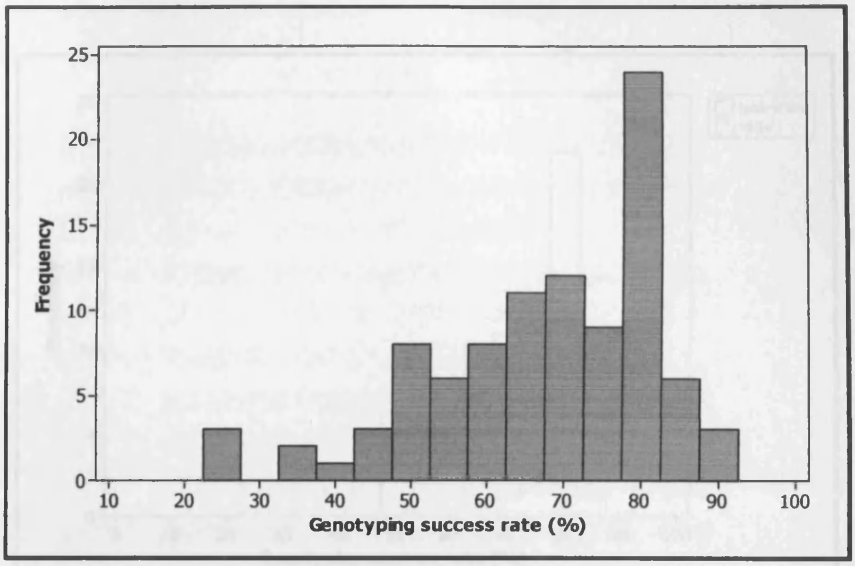


Figure 5.6. Success rates of subjects where success rate is equal to the number of genotypes produced divided by the number of reactions carried out.

5.3.3.1.2. Genotyping success rates: WGA samples

Of those 96 subjects genotyped, 29 underwent whole genome amplification by way of multiple displacement amplification (MDA, chapter 4). Figure 5.7. shows the comparison of the percentage of genotypes produced by those subjects who underwent whole genome amplification and those who's original DNA sample extracted from the mouthwash was high enough in human DNA content to be used as a suitable and plentiful DNA template. The distributions show a difference in success of genotyping (two sample t-test, $p<0.0001$), the mean percentage of WGA subjects being 53.22% (standard deviation=16.32) and that of the non-WGA individuals as 73.57% (standard deviation=9.17). Generally the WGA samples produced a lower percentage of genotypes, although there is some overlap.

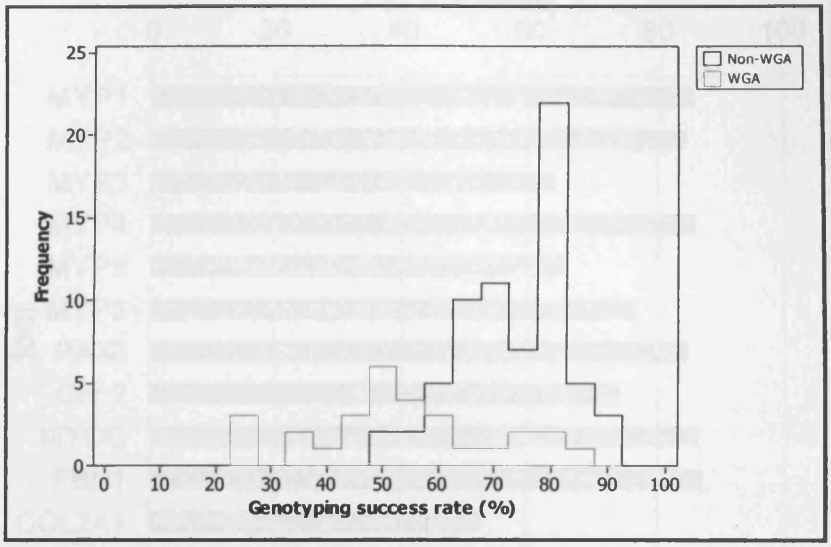


Figure 5.7. Comparing the genotype success rate produced by subjects using DNA from the original mouthwash (black outline) and those using WGA DNA (red outline).

5.3.3.1.3. Genotyping success rates: loci

Figure 5.8. shows the genotyping success rates of the individual marker loci. The best performances were for the FBN1 and MYOC loci, the worst for the collagen COL2A1 and COL11A1 loci.

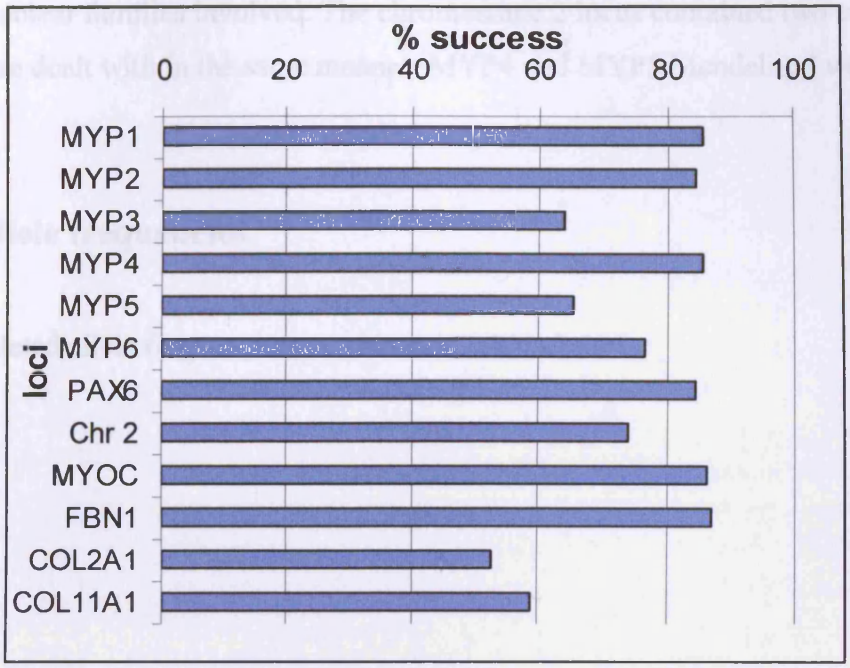


Figure 5.8. Success rates of groups of markers at the candidate loci investigated.

5.3.3.2. PEDCHECK

When examining the genotypes for Mendelian inconsistencies, each locus was investigated individually. MYP1 was found to have 11 errors, six of which were due to a male subject being scored as a heterozygote. This subject was subsequently removed from the analysis. One of the errors appeared to be due to preferential allele amplification and on re-examination of the genotyping data; this was corrected by re-assigning the genotype. Three errors appeared at the same marker and on re-examination of the genotyping traces identical results to the original analysis were found. This marker was therefore eliminated. A final error also showed no difference on reanalysis of the traces and therefore the immediate family group was removed for this marker.

MYP2 showed no Mendelian errors. One Mendelian inconsistency was found at each of the MYP3, MYP6, PAX6 and MYOC loci, which were corrected by removing the

specific nuclear families involved. The chromosome 2 locus contained two errors which were dealt with in the same manner. MYP4 and MYP5 Mendelised well.

5.3.4. Allele frequencies

The calculated allele frequencies are shown in tables 5.4a-l.

Table 5.4a-l. Allele frequencies calculated from founders and married in individuals.

Locus	Allele	DXS8106 Freq *Size*	DXS8028 Freq *Size*	DXS998 Freq *Size*	DXS8069 Freq *Size*	DXS8087 Freq *Size*	DXS1073 Freq *Size*
MYP	1	0.044 *156*	0.387 *320*	0.179 *340*	0.022 *162*	0.345 *322*	0.066 *246*
1	2	0.087 *158*	0.044 *322*	0.598 *342*	0.023 *164*	0.022 *326*	0.419 *248*
(mle)	3	0.337 *160*	0.022 *324*	0.201 *344*	0.483 *166*	0.461 *328*	0.428 *250*
	4	0.087 *162*	0.044 *326*	0.022 *346*	0.227 *168*	0.151 *330*	0.022 *260*
	5	0.022 *164*	0.066 *330*		0.224 *170*	0.022 *332*	0.066 *262*
	6	0.137 *166*	0.022 *334*		0.022 *172*		
	7	0.022 *168*	0.067 *338*				
	8	0.111 *170*	0.131 *342*				
	9	0.022 *172*	0.066 *344*				
	10	0.109 *174*	0.153 *346*				
	11	0.022 *176*					

Locus	Allele	D18S59 Freq *Size*	D18S476 Freq *Size*	D18S481 Freq *Size*	D18S63 Freq *Size*	D18S1132 Freq *Size*	D18S452 Freq *Size*	D18S1163 Freq *Size*	D18S464 Freq *Size*	D18S1158 Freq *Size*
MYP2	1	0.250 *157*	0.071 *283*	0.033 *189*	0.276 *260*	0.148 *120*	0.019 *102*	0.207 *201*	0.053 *276*	0.542 *103*
	2	0.054 *159*	0.107 *285*	0.317 *191*	0.017 *262*	0.111 *122*	0.056 *111*	0.034 *203*	0.053 *278*	0.188 *105*
	3	0.018 *161*	0.214 *287*	0.100 *193*	0.052 *264*	0.019 *124*	0.241 *117*	0.603 *205*	0.070 *280*	0.271 *107*
	4	0.107 *163*	0.054 *289*	0.083 *195*	0.069 *272*	0.463 *126*	0.093 *119*	0.017 *207*	0.632 *282*	
	5	0.286 *167*	0.429 *291*	0.083 *197*	0.052 *274*	0.222 *128*	0.037 *121*	0.017 *211*	0.140 *284*	
	6	0.054 *169*	0.054 *293*	0.233 *201*	0.259 *276*	0.037 *130*	0.259 *123*	0.086 *213*	0.052 *286*	
	7	0.178 *171*	0.071 *295*	0.033 *203*	0.138 *278*		0.111 *125*	0.034 *217*		
	8	0.053 *173*		0.084 *209*	0.069 *280*		0.074 *127*			
	9			0.034 *211*	0.017 *282*		0.019 *129*			
	10				0.051 *284*		0.093 *131*			

Locus	Allele	D12S80 Freq *Size*	D12S326 Freq *Size*	D12S64 Freq *Size*	D12S316 Freq *Size*	D12S348 Freq *Size*	D12S332 Freq *Size*	D12S1607 Freq *Size*	D12S78 Freq *Size*	D12S1605 Freq *Size*	D12S354 Freq *Size*
MYP3	1	0.093 *205*	0.234 *292*	0.059 *132*	0.500 *172*	0.074 *297*	0.047 *325*	0.186 *134*	0.328 *192*	0.273 *193*	0.290 *183*
	2	0.167 *207*	0.106 *304*	0.029 *134*	0.053 *174*	0.148 *299*	0.047 *333*	0.102 *144*	0.069 *194*	0.055 *195*	0.226 *187*
	3	0.037 *211*	0.064 *306*	0.029 *136*	0.053 *176*	0.296 *301*	0.581 *335*	0.271 *146*	0.052 *196*	0.309 *197*	0.032 *189*
	4	0.204 *213*	0.255 *308*	0.029 *138*	0.263 *178*	0.019 *303*	0.163 *337*	0.051 *148*	0.017 *198*	0.127 *199*	0.323 *193*
	5	0.407 *215*	0.085 *310*	0.441 *140*	0.105 *180*	0.370 *305*	0.023 *339*	0.322 *150*	0.069 *200*	0.218 *201*	0.097 *195*
	6	0.056 *217*	0.064 *312*	0.294 *142*	0.026 *184*	0.093 *307*	0.023 *341*	0.068 *152*	0.017 *202*	0.018 *203*	0.032 *197*
	7	0.036 *219*	0.149 *314*	0.089 *144*			0.116 *343*		0.052 *204*		
	8		0.043 *316*	0.030 *146*					0.121 *206*		
	9								0.121 *208*		
	10								0.069 *210*		
	11								0.034 *212*		
	12								0.017 *214*		
	13								0.034 *216*		

Table 5.4. continued

Locus	Allele	D7S798 Freq *Size*	D7S2546 Freq *Size*	D7S2423 Freq *Size*
MYP4	**1**	0.050 *246*	0.083 *310*	0.071 *210*
	2	0.017 *248*	0.188 *312*	0.429 *214*
	3	0.083 *250*	0.125 *314*	0.089 *216*
	4	0.217 *252*	0.354 *316*	0.375 *218*
	5	0.083 *256*	0.250 *318*	0.036 *222*
	6	0.117 *258*		
	7	0.433 *260*		

Locus	Allele	D17S1868 Freq *Size*	D17S1795 Freq *Size*	D17S956 Freq *Size*	D17S1290 Freq *Size*	D17S942 Freq *Size*	D17S2059 Freq *Size*	D17S1831 Freq *Size*	D17S1817 Freq *Size*
MYP5	**1**	0.021 *127*	0.053 *219*	0.286 *231*	0.043 *152*	0.216 *275*	0.083 *209*	0.022 *327*	0.058 *168*
	2	0.064 *129*	0.035 *221*	0.268 *233*	0.064 *156*	0.039 *277*	0.361 *221*	0.133 *331*	0.250 *170*
	3	0.043 *131*	0.175 *223*	0.393 *235*	0.043 *160*	0.196 *279*	0.417 *225*	0.022 *333*	0.327 *180*
	4	0.106 *133*	0.158 *225*	0.018 *237*	0.043 *164*	0.549 *281*	0.139 *229*	0.044 *335*	0.058 *182*
	5	0.660 *135*	0.035 *227*	0.018 *239*	0.021 *168*			0.200 *337*	0.019 *184*
	6	0.064 *137*	0.491 *229*	0.017 *243*	0.043 *172*			0.022 *339*	0.135 *186*
	7	0.042 *139*	0.053 *231*		0.128 *176*			0.133 *345*	0.058 *188*
	8				0.340 *180*			0.022 *347*	0.077 *190*
	9				0.084 *184*			0.400 *349*	0.018 *192*
	10				0.148 *188*				
	11				0.043 *192*				

Locus	Allele	D22S310 Freq *Size*	D22S1167 Freq *Size*	D22S1144 Freq *Size*	D22S1176 Freq *Size*	D22S424 Freq *Size*	D22S683 Freq *Size*	D22S1173 Freq *Size*	D22S283 Freq *Size*	D22S426 Freq *Size*	D22S1045 Freq *Size*	D22S445 Freq *Size*	D22S1157 Freq *Size*	D22S1179 Freq *Size*	D22S1168 Freq *Size*
MYP6	**1**	0.038 *184*	0.413 *213*	0.038 *139*	0.018 *128*	0.255 *93*	0.073 *197*	0.661 *158*	0.018 *182*	0.085 *277*	0.176 *360*	0.026 *165*	0.188 *369*	0.154 *172*	0.093 *235*
	2	0.269 *186*	0.138 *215*	0.019 *141*	0.127 *132*	0.018 *101*	0.145 *203*	0.264 *160*	0.053 *184*	0.191 *279*	0.039 *369*	0.158 *167*	0.063 *371*	0.019 *174*	0.703 *237*
	3	0.077 *188*	0.069 *221*	0.058 *147*	0.073 *134*	0.691 *103*	0.164 *205*	0.075 *162*	0.088 *186*	0.149 *281*	0.334 *372*	0.053 *171*	0.219 *375*	0.115 *176*	0.019 *243*
	4	0.115 *190*	0.328 *223*	0.443 *149*	0.073 *136*	0.036 *105*	0.164 *207*		0.070 *188*	0.426 *283*	0.412 *375*	0.158 *175*	0.031 *377*	0.058 *178*	0.148 *245*
	5	0.115 *192*	0.052 *225*	0.288 *151*	0.382 *138*		0.091 *209*		0.105 *190*	0.128 *285*	0.039 *378*	0.474 *179*	0.313 *379*	0.269 *180*	0.037 *251*
	6	0.077 *194*		0.058 *153*	0.255 *140*		0.073 *211*		0.088 *192*	0.021 *287*		0.105 *183*	0.094 *381*	0.077 *182*	
	7	0.173 *196*		0.077 *155*	0.018 *142*		0.036 *213*		0.070 *194*			0.026 *187*	0.031 *383*	0.231 *184*	
	8	0.115 *198*		0.019 *157*	0.018 *144*		0.073 *217*		0.070 *196*				0.063 *385*	0.058 *186*	
	9	0.019 *200*			0.018 *150*		0.018 *221*		0.070 *198*					0.019 *188*	
	10				0.018 *152*		0.036 *223*		0.193 *200*						
	11						0.091 *227*		0.105 *202*						
	12						0.018 *229*		0.053 *204*						
	13						0.018 *233*		0.017 *206*						

Table 5.4 continued

Locus	Allele	D11S2368 Freq *Size*	D11S1308 Freq *Size*	D11S4114 Freq *Size*	D11S992 Freq *Size*	D11S904 Freq *Size*	D11S4152 Freq *Size*	D11S4154 Freq *Size*	D11S4200 Freq *Size*	D11S4203 Freq *Size*	D11S4083 Freq *Size*	D11S1330 Freq *Size*	D11S905 Freq *Size*	D11S1993 Freq *Size*	D11S1361 Freq *Size*
PAX6	**1**	0.020 *277*	0.341 *197*	0.239 *279*	0.021 *305*	0.132 *264*	0.018 *113*	0.167 *140*	0.153 *232*	0.281 *156*	0.085 *206*	0.891 *246*	0.018 *120*	0.407 *225*	0.735 *212*
	2	0.040 *285*	0.585 *199*	0.022 *281*	0.043 *307*	0.132 *266*	0.091 *115*	0.167 *142*	0.051 *234*	0.386 *166*	0.220 *208*	0.036 *248*	0.105 *122*	0.017 *228*	0.122 *214*
	3	0.240 *289*	0.074 *201*	0.043 *287*	0.511 *309*	0.132 *270*	0.073 *119*	0.056 *144*	0.102 *238*	0.018 *187*	0.017 *210*	0.073 *250*	0.298 *124*	0.068 *231*	0.020 *216*
	4	0.140 *293*		0.022 *295*	0.064 *311*	0.057 *272*	0.055 *121*	0.185 *146*	0.424 *240*	0.105 *189*	0.102 *212*		0.018 *128*	0.102 *237*	0.082 *225*
	5	0.140 *297*		0.043 *297*	0.021 *313*	0.038 *274*	0.145 *123*	0.093 *149*	0.051 *242*	0.053 *191*	0.119 *214*		0.035 *132*	0.220 *240*	0.041 *227*
	6	0.300 *301*		0.217 *299*	0.149 *315*	0.208 *278*	0.091 *125*	0.222 *151*	0.034 *244*	0.018 *193*	0.017 *216*		0.035 *134*	0.119 *243*	
	7	0.060 *305*		0.239 *301*	0.191 *319*	0.283 *280*	0.018 *127*	0.093 *153*	0.034 *246*	0.018 *195*	0.068 *218*		0.175 *136*	0.034 *246*	
	8	0.020 *309*		0.065 *303*		0.019 *282*	0.036 *132*	0.019 *155*	0.034 *248*	0.053 *197*	0.119 *220*		0.211 *138*	0.034 *249*	
	9	0.040 *313*		0.110 *305*			0.109 *134*		0.102 *250*	0.018 *199*	0.136 *222*		0.105 *140*		
	10						0.161 *136*		0.017 *252*	0.035 *201*	0.051 *224*				
	11						0.161 *138*			0.018 *207*	0.017 *226*				
	12						0.018 *143*				0.017 *228*				
	13						0.018 *145*				0.034 *230*				

Locus	Allele	D2S2297 Freq *Size*	D2S2317 Freq *Size*	D2S172 Freq *Size*	D2S2193 Freq *Size*	D2S2344 Freq *Size*	D2S206 Freq *Size*	D2S2205 Freq *Size*	D2S336 Freq *Size*	D2S2973 Freq *Size*	D2S2202 Freq *Size*
Chr2	**1**	0.087 *109*	0.017 *159*	0.021 *262*	0.147 *370*	0.173 *247*	0.026 *200*	0.041 *150*	0.123 *192*	0.173 *123*	0.019 *149*
	2	0.043 *113*	0.053 *161*	0.021 *266*	0.029 *380*	0.058 *249*	0.154 *202*	0.061 *156*	0.053 *205*	0.017 *127*	0.245 *151*
	3	0.130 *115*	0.088 *163*	0.128 *268*	0.029 *382*	0.038 *251*	0.077 *204*	0.531 *158*	0.193 *211*	0.259 *131*	0.245 *153*
	4	0.065 *117*	0.088 *165*	0.021 *270*	0.147 *384*	0.193 *253*	0.077 *206*	0.347 *160*	0.158 *213*	0.034 *136*	0.434 *155*
	5	0.348 *119*	0.053 *167*	0.085 *272*	0.029 *386*	0.038 *255*	0.256 *208*	0.020 *162*	0.070 *215*	0.414 *140*	0.057 *157*
	6	0.022 *121*	0.228 *177*	0.106 *274*	0.206 *388*	0.366 *257*	0.308 *210*		0.298 *217*	0.103 *144*	
	7	0.283 *123*	0.211 *179*	0.043 *276*	0.059 *390*	0.038 *259*	0.026 *212*		0.070 *219*		
	8	0.022 *125*	0.017 *185*	0.021 *278*	0.088 *392*	0.019 *265*	0.026 *214*		0.035 *221*		
	9		0.193 *187*	0.085 *280*	0.059 *394*	0.058 *269*	0.050 *222*				
	10		0.035 *189*	0.064 *282*	0.088 *396*	0.019 *273*					
	11		0.017 *191*	0.064 *284*	0.029 *398*						
	12			0.043 *286*	0.029 *400*						
	13			0.064 *288*	0.059 *402*						
	14			0.021 *290*							
	15			0.043 *292*							
	16			0.085 *294*							
	17			0.043 *298*							
	18			0.021 *300*							
	19			0.021 *302*							

Table 5.4. continued

Locus	Allele	MYOC1 Freq *Size*	MYOC2 Freq *Size*
MYOC	**1**	0.019 *370*	0.053 *116*
	2	0.426 *372*	0.263 *118*
	3	0.296 *374*	0.070 *120*
	4	0.148 *375*	0.596 *122*
	5	0.111 *376*	0.018 *124*

Locus	Allele	D15S1003 Freq *Size*	D15S992 Freq *Size*
FBN1	**1**	0.148 *195*	0.018 *84*
	2	0.426 *197*	0.178 *94*
	3	0.333 *199*	0.089 *96*
	4	0.037 *201*	0.018 *98*
	5	0.019 *203*	0.018 *100*
	6	0.037 *212*	0.036 *102*
	7		0.054 *104*
	8		0.374 *106*
	9		0.054 *108*
	10		0.054 *110*
	11		0.107 *112*

Locus	Allele	D12S1701 (mle) Freq	Freq mle	Size	D12S361 (mle) Freq	Freq mle	Size
COL2A1	**1**	0.056	(0.054)	*125*	0.241	(0.242)	*241*
	2	0.444	(0.464)	*131*	0.037	(0.035)	*245*
	3	0.250	(0.259)	*133*	0.167	(0.159)	*247*
	4	0.167	(0.151)	*135*	0.018	(0.017)	*249*
	5	0.083	(0.072)	*137*	0.426	(0.442)	*251*
	6				0.093	(0.087)	*253*
	7				0.018	(0.018)	*255*

Locus	Allele	D1S2626 Freq *Size*	D1S2888 Freq *Size*
COL11A1	**1**	0.036 *292*	0.040 *278*
	2	0.018 *294*	0.200 *280*
	3	0.182 *296*	0.200 *282*
	4	0.218 *298*	0.400 *284*
	5	0.345 *300*	0.040 *286*
	6	0.073 *302*	0.080 *288*
	7	0.073 *304*	0.040 *290*
	8	0.055 *306*	

5.3.5. Linkage analysis: SOLAR

The linkage analysis results are shown in table 5.5. for twopoint analysis. Figures 5.9i. a-c and 5.9iia-g. show multipoint analysis for exclusion candidate gene regions and potential linkage regions respectively.

5.3.5.1. Syndromic Loci

The COL2A1 locus showed a suggestion of linkage with a maximum twopoint LOD score of 0.768 at D12S1701 using trait q2 and maximum likelihood estimate (MLE) allele frequencies. This was supported by a twopoint LOD score of 0.585 using the same analysis criteria but with trait q1. The other analyses showed lower LOD scores (0 to 0.298) therefore suggesting absence of linkage to this region.

The COL11A1 locus showed no linkage, the highest twopoint LOD score being 0.030 therefore it is possible to exclude linkage to this locus.

The FBN1 locus, again, showed little evidence of linkage with a maximum LOD score occurring when analysing the trait q1 with equal allele frequencies and obtaining a twopoint LOD score of 0.193.

5.3.5.2. Non syndromic loci

The MYP1 locus had its highest twopoint LOD score consistently at marker DXS1073, ranging from 0.260 to a maximum of 0.880 dependent on the criteria used. The maximum was calculated using q1 as the trait and equal allele frequencies.

Chromosome 18's locus, MYP2, had the highest LOD score of 0.088 when analysing trait q2 with equal allele frequencies and a multipoint analysis (figure 5.9iia). The scores for the other markers within this region varied only slightly from 0.

LOD scores at the MYP3 locus showed little variation from 0 using trait q1 as the phenotype. However, when using q2, the LOD scores increased. The multipoint LOD scores reached a peak of 1.370 when using equal frequencies; between markers D12S1605 and D12S354 (figure 5.9iib). Lower peaks occurred at markers D12S348 and D12S1607 (0.850 and 0.915 respectively) this pattern being followed by the analysis using the founder allele frequencies, but having lower LOD scores of approximately 0.3. Twopoint analysis gives a maximum LOD score of 0.459 at D12S354 and 0.417 at D12S348 with equal allele frequencies but only 0.048 at D12S354 for founder frequencies increasing to a maximum of 0.534 at D12S348.

MYP4, on chromosome 7 showed no evidence for linkage, the maximum LOD score being a multipoint score of 0.309 using q2 and equal frequencies (figure 5.9iic.), the twopoint LOD scores maximizing at 0.049.

MYP5 had a maximum twopoint LOD score of 0.806 at D17S1831 (q1, founder frequencies). The q1 twopoint analysis with equal allele frequencies also showed a peak at this marker and the multipoint analysis was at its maximum LOD score between this marker and D17S1817 (figure 5.9iid). Trait q2 showed a similar trend.

The analysis of the MYP6 region on chromosome 22 showed a 5cM region with multipoint LOD scores over 1 with trait q2 and equal allele frequencies (figure 5.9iie). The majority of the higher LOD scores occurred around markers D22S1173 and D22S1176 for the twopoint analysis and D22S1173 and D22S683 for the multipoint.

Chromosome 11 contains the candidate gene PAX6. Twopoint analysis produced a maximum peak of 0.598 at marker D11S1993 with trait q2 and equal allele frequencies. There were also consistently relatively high LOD scores at markers D11S904 and D11S4152. This was supported by the multipoint analysis showing peaks around these two markers, the maximum score of 0.796 for q2 with equal allele frequencies (figure 5.9iif.).

The chromosome 2 locus had a maximum twopoint LOD score of 0.831 at marker D2S2344, which was consistently higher than the adjacent markers, with q2 and

founder allele frequencies. Multipoint analysis showed a peak at the edge of the region around marker D2S2202, with a LOD score of only 0.686 (figure 5.9iig). The criteria of trait q1 with equal allele frequencies are those for which the MYOC locus had its highest LOD score, of 0.034 at the MYOC2 marker; therefore exclusion of linkage from this region of chromosome 1 is appropriate.

In summary the higher LOD scores occurred with the trait q2 and using equal allele frequencies. Generally the next highest scores were found using the same trait but with allele frequencies calculated from pedigree founders and married-in subjects. Allele frequencies did not seem to have such an influential effect on trait q1.

Table 5.5. Summary of SOLAR twopoint analysis

COL2a1 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D12S361	0.0028	-181.559	0.357657	0.028187
D12S1701	0.2530	-180.983	0.000000	0.380851

COL2a1 MLE frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D12S361	0.0117	-178.966	0.353856	0.066602
D12S1701	0.5849	-177.646	0.000000	0.424652

COL2a1 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D12S361	0.0000	-231.536	0.394830	0.000000
D12S1701	0.2982	-230.849	0.000000	0.406670

COL2a1 MLE frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D12S361	0.000	-231.536	0.395	0.000
D12S1701	0.768	-229.767	0.000	0.422

COL11a1 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D1S2626	0.0000	-181.565	0.386996	0.000000
D1S2888	0.0000	-181.565	0.386996	0.000000

COL11a1 MLE frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D1S2626	0.0000	-178.993	0.420816	0.000000
D1S2888	0.0296	-178.925	0.216415	0.209437

COL11a1 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D1S2626	0.0000	-231.536	0.394830	0.000000
D1S2888	0.0000	-231.536	0.394830	0.000000

COL11a1 MLE frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D1S2626	0.0000	-231.536	0.394830	0.000000
D1S2888	0.0065	-231.521	0.319747	0.074543

Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D15S992	0.0054	-181.553	0.363106	0.030562
D15S1003	0.1931	-181.121	0.196313	0.161063

Table 5.5. continued

MLE frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D15S992	0.0965	-181.343	0.279546	0.136732
D15S1003	0.0288	-181.499	0.307647	0.076148

Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D15S992	0.1120	-181.308	0.266479	0.152047
D15S1003	0.0232	-181.512	0.311637	0.072133

Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D15S992	0.0000	-231.536	0.394830	0.000000
D15S1003	0.1289	-231.239	0.224532	0.147984

MLE frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D15S992	0.0005	-231.535	0.385827	0.010784
D15S1003	0.0169	-231.497	0.336659	0.056447

Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D15S992	0.0046	-231.525	0.365211	0.035045
D15S1003	0.0098	-231.513	0.345445	0.047401

MYOC Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
myoc1	0.0000	-181.565	0.386996	0.000000
myoc2	0.0336	-181.488	0.311554	0.064754

MYOC MLE frequencies, q1

Model	LOD	Loglike	H2r	H2q1
myoc1	0.0000	-181.565	0.386996	0.000000
myoc2	0.0158	-181.529	0.335908	0.045474

MYOC Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
myoc1	0.0000	-231.536	0.394830	0.000000
myoc2	0.0008	-231.534	0.380306	0.013199

MYOC MLE frequencies, q2

Model	LOD	Loglike	H2r	H2q1
myoc1	0.0000	-231.536	0.394830	0.000000
myoc2	0.0000	-231.536	0.394830	0.000000

Table 5.5. continued

MYP2 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D18S59	0.0000	-181.565	0.386996	0.000000
D18S63	0.0000	-181.565	0.386996	0.000000
D18S452	0.0120	-181.538	0.326513	0.053234
D18S464	0.0000	-181.565	0.386996	0.000000
D18S476	0.0000	-181.565	0.386996	0.000000
D18S481	0.0000	-181.565	0.386996	0.000000
D18S1132	0.0000	-181.565	0.386996	0.000000
D18S1158	0.0000	-181.565	0.386996	0.000000
D18S1163	0.0000	-181.565	0.386996	0.000000

MYP2 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D18S59	0.0000	-231.536	0.394830	0.000000
D18S63	0.0000	-231.536	0.394830	0.000000
D18S452	0.0000	-231.536	0.394830	0.000000
D18S464	0.0000	-231.536	0.394830	0.000000
D18S476	0.0000	-231.536	0.394830	0.000000
D18S481	0.0000	-231.536	0.394830	0.000000
D18S1132	0.0000	-231.536	0.394830	0.000000
D18S1158	0.0000	-231.536	0.394830	0.000000
D18S1163	0.0000	-231.536	0.394830	0.000000

MYP2 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D18S59	0.0000	-231.536	0.394830	0.000000
D18S63	0.0000	-231.536	0.394830	0.000000
D18S452	0.0000	-231.536	0.394830	0.000000
D18S464	0.0000	-231.536	0.394830	0.000000
D18S476	0.0000	-231.536	0.394830	0.000000
D18S481	0.0000	-231.536	0.394830	0.000000
D18S1132	0.0000	-231.536	0.394830	0.000000
D18S1158	0.0150	-231.501	0.313582	0.075160
D18S1163	0.0000	-231.536	0.394830	0.000000

MYP3 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D12S64	0.0098	-181.543	0.304112	0.097100
D12S78	0.0000	-181.565	0.386996	0.000000
D12S80	0.0000	-181.565	0.386996	0.000000
D12S316	0.0000	-181.565	0.386996	0.000000
D12S326	0.1706	-181.173	0.189038	0.246972
D12S332	0.0000	-181.565	0.386996	0.000000
D12S348	0.0000	-181.565	0.386996	0.000000
D12S354	0.2405	-181.012	0.000000	0.424590
D12S1605	0.0065	-181.550	0.334297	0.047490
D12S1607	0.0000	-181.565	0.386996	0.000000

Table 5.5. continued

MYP1 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
DXS998	0.1417	-129.604	0.147053	0.118728
DXS1073	0.8803	-127.904	0.000000	0.312240
DXS8028	0.1224	-129.649	0.168954	0.101021
DXS8069	0.2238	-129.415	0.128810	0.127152
DXS8087	0.5398	-128.688	0.000000	0.261738
DXS8106	0.2407	-129.376	0.131641	0.127711

MYP1 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
DXS998	0.0831	-129.739	0.173120	0.092959
DXS1073	0.7457	-128.214	0.000000	0.306038
DXS8028	0.1182	-129.658	0.173069	0.098671
DXS8069	0.2119	-129.443	0.134869	0.122273
DXS8087	0.4402	-128.917	0.000000	0.267473
DXS8106	0.2351	-129.389	0.137866	0.128112

MYP1 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
DXS998	0.0006	-167.874	0.253114	0.008775
DXS1073	0.4021	-166.950	0.048197	0.252093
DXS8028	0.0121	-167.848	0.231693	0.034544
DXS8069	0.0693	-167.716	0.179366	0.082058
DXS8087	0.1341	-167.567	0.112117	0.150876
DXS8106	0.1095	-167.624	0.177666	0.089648

MYP1 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
DXS998	0.0000	-167.876	0.261366	0.000000
DXS1073	0.2598	-167.277	0.087693	0.207169
DXS8028	0.0096	-167.853	0.235511	0.030851
DXS8069	0.0590	-167.740	0.187950	0.074936
DXS8087	0.0540	-167.751	0.163452	0.102891
DXS8106	0.1043	-167.636	0.181240	0.090345

MYP2 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D18S59	0.0000	-181.565	0.386996	0.000000
D18S63	0.0000	-181.565	0.386996	0.000000
D18S452	0.0007	-181.564	0.376209	0.009300
D18S464	0.0000	-181.565	0.386996	0.000000
D18S476	0.0000	-181.565	0.386996	0.000000
D18S481	0.0000	-181.565	0.386996	0.000000
D18S1132	0.0000	-181.565	0.386996	0.000000
D18S1158	0.0000	-181.565	0.386996	0.000000
D18S1163	0.0009	-181.563	0.375720	0.010209

Table 5.5. continued

MYP3 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D12S64	0.2192	-181.061	0.000000	0.448210
D12S78	0.0000	-181.565	0.386996	0.000000
D12S80	0.0000	-181.565	0.386996	0.000000
D12S316	0.2081	-181.086	0.000000	0.383061
D12S326	0.1544	-181.210	0.212761	0.214475
D12S332	0.0000	-181.565	0.386996	0.000000
D12S348	0.0000	-181.565	0.386996	0.000000
D12S354	0.1034	-181.327	0.000000	0.402328
D12S1605	0.0000	-181.565	0.386996	0.000000
D12S1607	0.0000	-181.565	0.386996	0.000000

MYP3 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D12S64	0.0541	-231.411	0.232250	0.185540
D12S78	0.0097	-231.514	0.346576	0.051191
D12S80	0.0000	-231.536	0.394830	0.000000
D12S316	0.0000	-231.536	0.394830	0.000000
D12S326	0.2474	-230.966	0.058809	0.393989
D12S332	0.0000	-231.536	0.394830	0.000000
D12S348	0.4167	-230.576	0.000000	0.395935
D12S354	0.4594	-230.478	0.000000	0.435717
D12S1605	0.1871	-231.105	0.152422	0.236471
D12S1607	0.1559	-231.177	0.196390	0.194815

MYP3 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D12S64	0.1405	-231.212	0.012194	0.420963
D12S78	0.0158	-231.500	0.333284	0.065553
D12S80	0.0000	-231.536	0.394830	0.000000
D12S316	0.3102	-230.822	0.000000	0.419820
D12S326	0.2373	-230.990	0.084749	0.357705
D12S332	0.0000	-231.536	0.394830	0.000000
D12S348	0.5340	-230.306	0.000000	0.406803
D12S354	0.0481	-231.425	0.033875	0.366075
D12S1605	0.1175	-231.265	0.213652	0.182050
D12S1607	0.0984	-231.309	0.243240	0.151676

MYP4 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D7S798	0.0377	-181.479	0.294706	0.072133
D7S2423	0.0000	-181.565	0.386996	0.000000
D7S2546	0.0316	-181.493	0.308703	0.064979

MYP4 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D7S798	0.0485	-181.454	0.272152	0.090291
D7S2423	0.0188	-181.522	0.284374	0.095149
D7S2546	0.0271	-181.503	0.313116	0.059997

Table 5.5. continued

MYP4 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D7S798	0.0000	-231.536	0.394830	0.000000
D7S2423	0.0000	-231.536	0.394830	0.000000
D7S2546	0.0000	-231.536	0.394830	0.000000

MYP4 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D7S798	0.0000	-231.536	0.394830	0.000000
D7S2423	0.0000	-231.536	0.394830	0.000000
D7S2546	0.0000	-231.536	0.394830	0.000000

MYP5 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D17S942	0.0097	-181.543	0.349374	0.040799
D17S956	0.0000	-181.565	0.386996	0.000000
D17S1290	0.0000	-181.565	0.386996	0.000000
D17S1795	0.0000	-181.565	0.386996	0.000000
D17S1817	0.1473	-181.226	0.161901	0.227473
D17S1831	0.4839	-180.451	0.000000	0.394739
D17S1868	0.0000	-181.565	0.386996	0.000000
D17S2059	0.0000	-181.565	0.386996	0.000000

MYP5 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D17S942	0.0000	-181.565	0.386996	0.000000
D17S956	0.0000	-181.565	0.386996	0.000000
D17S1290	0.0000	-181.565	0.386996	0.000000
D17S1795	0.0000	-181.565	0.386996	0.000000
D17S1817	0.1781	-181.155	0.115713	0.270846
D17S1831	0.8055	-179.711	0.000000	0.407338
D17S1868	0.0000	-181.565	0.386996	0.000000
D17S2059	0.1965	-181.113	0.000000	0.393867

MYP5 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D17S942	0.0000	-231.36	0.394830	0.000000
D17S956	0.0000	-231.36	0.394830	0.000000
D17S1290	0.0057	-231.523	0.337465	0.056483
D17S1795	0.1158	-231.269	0.154392	0.223398
D17S1817	0.3966	-230.623	0.039523	0.363323
D17S1831	0.2290	-231.009	0.083277	0.313130
D17S1868	0.0000	-231.36	0.394830	0.000000
D17S2059	0.0000	-231.36	0.394830	0.000000

Table 5.5. continued

MYP5 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D17S942	0.0000	-231.536	0.394830	0.000000
D17S956	0.0000	-231.536	0.394830	0.000000
D17S1290	0.0000	-231.536	0.394830	0.000000
D17S1795	0.0803	-231.351	0.197365	0.183298
D17S1817	0.3843	-230.651	0.008077	0.402183
D17S1831	0.2446	-230.973	0.049486	0.337068
D17S1868	0.0000	-231.536	0.394830	0.000000
D17S2059	0.1249	-231.248	0.000000	0.383491

MYP6 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D22S283	0.0249	-181.508	0.313075	0.065913
D22S310	0.5432	-180.315	0.116805	0.226367
D22S424	0.0661	-181.413	0.198027	0.163426
D22S426	0.0000	-181.565	0.386996	0.000000
D22S445	0.0000	-181.565	0.386996	0.000000
D22S683	0.0000	-181.565	0.384889	0.001620
D22S1045	0.5700	-180.253	0.000000	0.353026
D22S1144	0.2475	-180.996	0.217164	0.162502
D22S1157	0.0000	-181.565	0.386996	0.000000
D22S1167	0.3813	-180.687	0.022615	0.309439
D22S1168	0.0881	-181.363	0.292096	0.091407
D22S1173	0.7175	-179.913	0.000000	0.355552
D22S1176	0.2455	-181.000	0.169502	0.229040
D22S1179	0.0011	-181.563	0.370337	0.016762

MYP6 Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D22S283	0.0408	-181.472	0.299236	0.078347
D22S310	0.4861	-180.446	0.119476	0.226889
D22S424	0.1224	-181.284	0.155883	0.198967
D22S426	0.0000	-181.565	0.386996	0.000000
D22S445	0.0289	-181.499	0.288168	0.099325
D22S683	0.0012	-181.563	0.366496	0.015359
D22S1045	0.5640	-180.267	0.000000	0.357935
D22S1144	0.0634	-181.420	0.290654	0.100314
D22S1157	0.0000	-181.565	0.386996	0.000000
D22S1167	0.3337	-180.797	0.040116	0.301267
D22S1168	0.0323	-181.491	0.313847	0.075838
D22S1173	0.7828	-179.763	0.000000	0.367137
D22S1176	0.4518	-180.525	0.096261	0.308386
D22S1179	0.0466	-181.458	0.274949	0.116941

Table 5.5. continued

MYP6 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D22S283	0.2921	-230.863	0.126526	0.243062
D22S310	0.9640	-229.316	0.000000	0.378567
D22S424	0.3761	-230.670	0.000000	0.356831
D22S426	0.1509	-231.188	0.196415	0.206529
D22S445	0.0000	-231.536	0.394830	0.000000
D22S683	0.3109	-230.820	0.051944	0.295378
D22S1045	0.5481	-230.274	0.000000	0.364541
D22S1144	0.2461	-230.969	0.172573	0.205371
D22S1157	0.0000	-231.536	0.394830	0.000000
D22S1167	0.5096	-230.363	0.000000	0.348338
D22S1168	0.0859	-231.338	0.276281	0.116272
D22S1173	1.0113	-229.207	0.000000	0.395799
D22S1176	1.0779	-229.054	0.000000	0.443599
D22S1179	0.0285	-231.470	0.320955	0.075644

MYP6 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D22S283	0.2662	-230.923	0.154054	0.220537
D22S310	0.9056	-229.451	0.000000	0.383274
D22S424	0.3613	-230.704	0.000000	0.361692
D22S426	0.1105	-231.282	0.219147	0.184116
D22S445	0.0274	-231.473	0.288907	0.106547
D22S683	0.2844	-230.881	0.087147	0.258405
D22S1045	0.3571	-230.714	0.000000	0.363728
D22S1144	0.0262	-231.475	0.319444	0.074855
D22S1157	0.0000	-231.536	0.394830	0.000000
D22S1167	0.3616	-230.703	0.000000	0.348477
D22S1168	0.1038	-231.297	0.223089	0.176102
D22S1173	0.9821	-229.274	0.000000	0.408003
D22S1176	1.1871	-228.803	0.000000	0.448901
D22S1179	0.0582	-231.402	0.288160	0.114359

Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D11S904	0.5616	-180.272	0.154651	0.266697
D11S905	0.0000	-181.565	0.386996	0.000000
D11S992	0.0000	-181.565	0.386996	0.000000
D11S1308	0.0000	-181.565	0.386996	0.000000
D11S1330	0.0000	-181.565	0.386996	0.000000
D11S1361	0.3386	-180.786	0.038302	0.311824
D11S1993	0.3418	-180.778	0.167890	0.226734
D11S2368	0.0394	-181.475	0.311555	0.080386
D11S4083	0.0000	-181.565	0.386996	0.000000
D11S4114	0.0236	-181.511	0.314428	0.057319
D11S4152	0.3626	-180.731	0.142700	0.217758
D11S4154	0.0262	-181.505	0.321356	0.066642
D11S4200	0.0000	-181.565	0.386996	0.000000
D11S4203	0.0000	-181.565	0.386996	0.000000

Table 5.5. continued

Founder frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D11S904	0.4102	-180.621	0.182200	0.232614
D11S905	0.0000	-181.565	0.386996	0.000000
D11S992	0.0062	-181.551	0.321528	0.066931
D11S1308	0.0000	-181.565	0.386996	0.000000
D11S1330	0.0000	-181.565	0.386996	0.000000
D11S1361	0.0043	-181.556	0.349234	0.035091
D11S1993	0.3370	-180.790	0.170739	0.230605
D11S2368	0.0158	-181.529	0.339868	0.050486
D11S4083	0.0000	-181.565	0.386996	0.000000
D11S4114	0.0000	-181.565	0.386996	0.000000
D11S4152	0.5590	-180.278	0.081614	0.270773
D11S4154	0.0714	-181.401	0.288496	0.103710
D11S4200	0.0000	-181.565	0.386996	0.000000
D11S4203	0.0000	-181.565	0.386996	0.000000

Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D11S904	0.3794	-230.662	0.211225	0.217215
D11S905	0.0000	-231.536	0.394830	0.000000
D11S992	0.0000	-231.536	0.394830	0.000000
D11S1308	0.0000	-231.536	0.394830	0.000000
D11S1330	0.0000	-231.536	0.394830	0.000000
D11S1361	0.1611	-231.165	0.167240	0.203760
D11S1993	0.5977	-230.160	0.000000	0.428529
D11S2368	0.0047	-231.525	0.369238	0.029420
D11S4083	0.0000	-231.536	0.394830	0.000000
D11S4114	0.0582	-231.402	0.275843	0.099097
D11S4152	0.3078	-230.827	0.162848	0.208711
D11S4154	0.0231	-231.483	0.319847	0.076194
D11S4200	0.0584	-231.401	0.294566	0.108686
D11S4203	0.0000	-231.536	0.394830	0.000000

Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D11S904	0.1793	-231.123	0.265633	0.149264
D11S905	0.0000	-231.536	0.394830	0.000000
D11S992	0.0000	-231.536	0.394830	0.000000
D11S1308	0.0000	-231.536	0.394830	0.000000
D11S1330	0.0045	-231.525	0.318219	0.078459
D11S1361	0.0000	-231.536	0.394830	0.000000
D11S1993	0.4091	-230.594	0.057184	0.368692
D11S2368	0.0000	-231.536	0.394830	0.000000
D11S4083	0.0000	-231.536	0.394830	0.000000
D11S4114	0.0270	-231.474	0.304644	0.074120
D11S4152	0.4956	-230.395	0.092715	0.270840
D11S4154	0.0348	-231.456	0.310131	0.087181
D11S4200	0.0078	-231.518	0.366559	0.031803
D11S4203	0.0000	-231.536	0.394830	0.000000

Table 5.5. continued

CHR2 Equal frequencies, q1

Model	LOD	Loglike	H2r	H2q1
D2S172	0.0624	-181.422	0.315921	0.073814
D2S206	0.0747	-181.393	0.113937	0.223683
D2S336	0.0512	-181.448	0.281281	0.095972
D2S2193	0.3452	-180.771	0.000000	0.351761
D2S2202	0.1751	-181.162	0.078512	0.261419
D2S2205	0.0000	-181.565	0.386996	0.000000
D2S2297	0.1188	-181.292	0.272059	0.134904
D2S2317	0.0342	-181.487	0.308776	0.069755
D2S2344	0.3647	-180.726	0.081017	0.293035
D2S2973	0.0004	-181.565	0.377401	0.007632

CHR2 Founder frequencies, q1

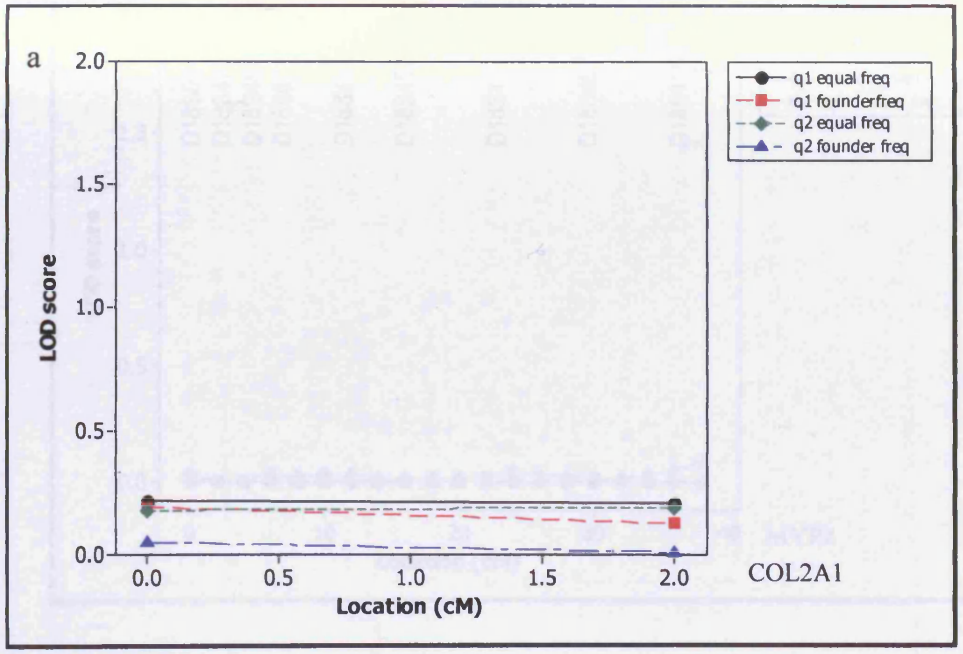
Model	LOD	Loglike	H2r	H2q1
D2S172	0.0808	-181.379	0.298720	0.089338
D2S206	0.0580	-181.432	0.164454	0.186842
D2S336	0.0738	-181.396	0.241961	0.127738
D2S2193	0.0241	-181.510	0.000000	0.337822
D2S2202	0.1133	-181.305	0.105050	0.251509
D2S2205	0.0000	-181.565	0.386996	0.000000
D2S2297	0.2244	-181.049	0.221268	0.189796
D2S2317	0.0403	-181.473	0.301869	0.077817
D2S2344	0.7319	-179.880	0.000000	0.380847
D2S2973	0.0022	-181.560	0.353708	0.028204

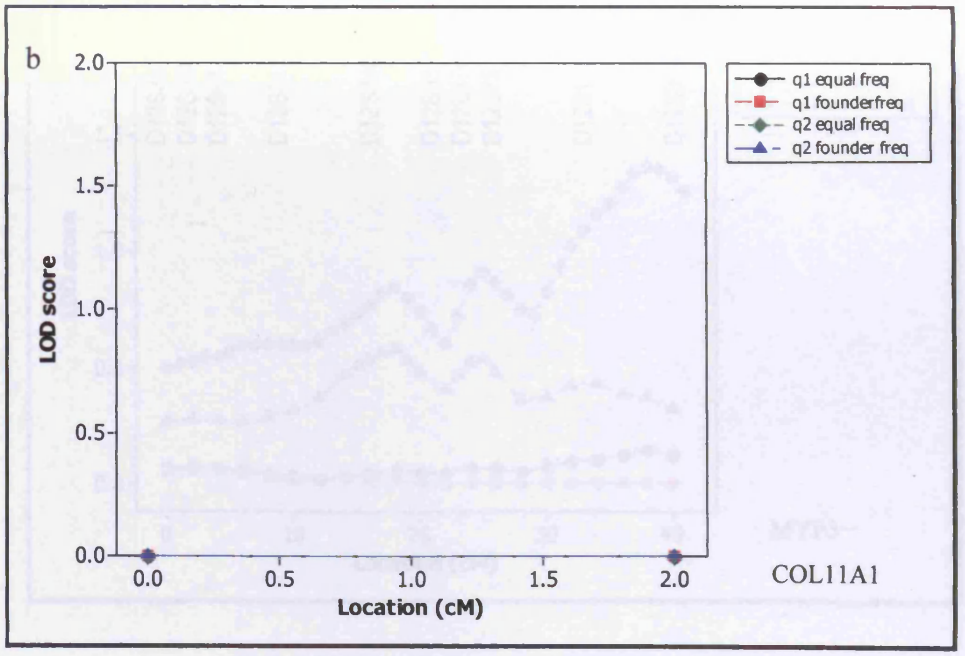
CHR2 Equal frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D2S172	0.0063	-231.521	0.362259	0.030491
D2S206	0.2014	-231.072	0.000000	0.336830
D2S336	0.0017	-231.532	0.371456	0.020658
D2S2193	0.1937	-231.090	0.000000	0.343940
D2S2202	0.0905	-231.328	0.215538	0.164189
D2S2205	0.0000	-231.536	0.394830	0.000000
D2S2297	0.0651	-231.386	0.265962	0.141142
D2S2317	0.0000	-231.536	0.394830	0.000000
D2S2344	0.7006	-229.923	0.000000	0.403225
D2S2973	0.0000	-231.536	0.394830	0.000000

CHR2 Founder frequencies, q2

Model	LOD	Loglike	H2r	H2q1
D2S172	0.0000	-231.536	0.393022	0.001661
D2S206	0.1729	-231.138	0.000000	0.336806
D2S336	0.0070	-231.520	0.335566	0.050631
D2S2193	0.0331	-231.460	0.154765	0.210754
D2S2202	0.0325	-231.461	0.272098	0.118520
D2S2205	0.0000	-231.536	0.394830	0.000000
D2S2297	0.0893	-231.330	0.258211	0.150021
D2S2317	0.0003	-231.535	0.383687	0.009691
D2S2344	0.8310	-229.622	0.000000	0.405963
D2S2973	0.0000	-231.536	0.394830	0.000000





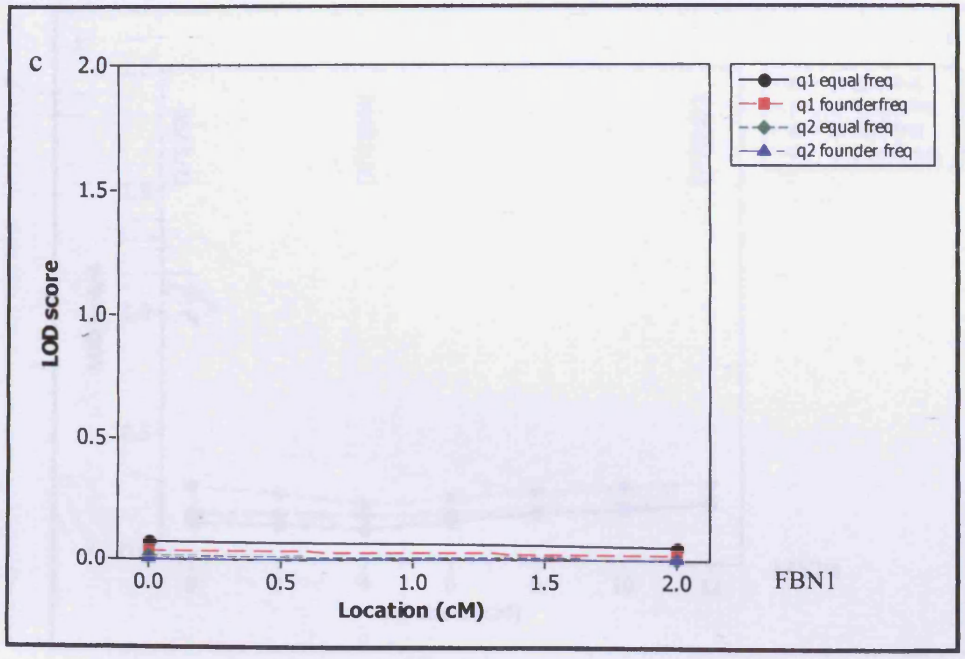
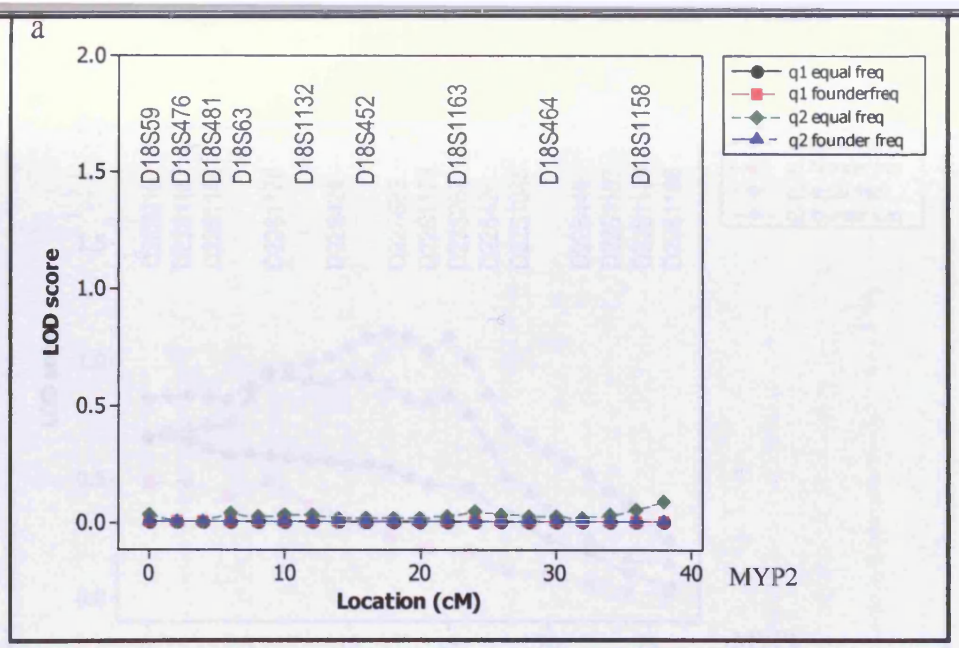
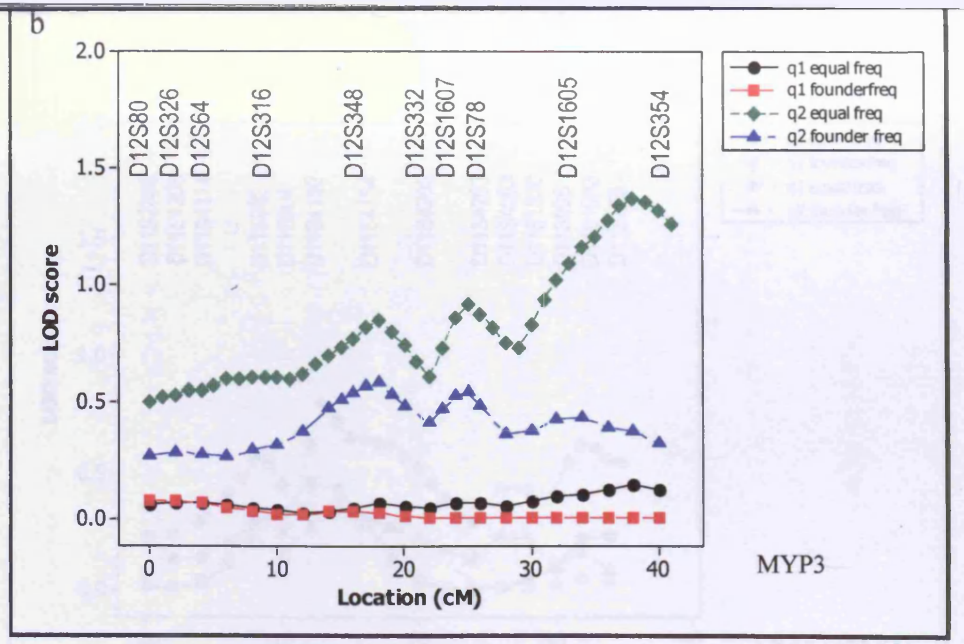
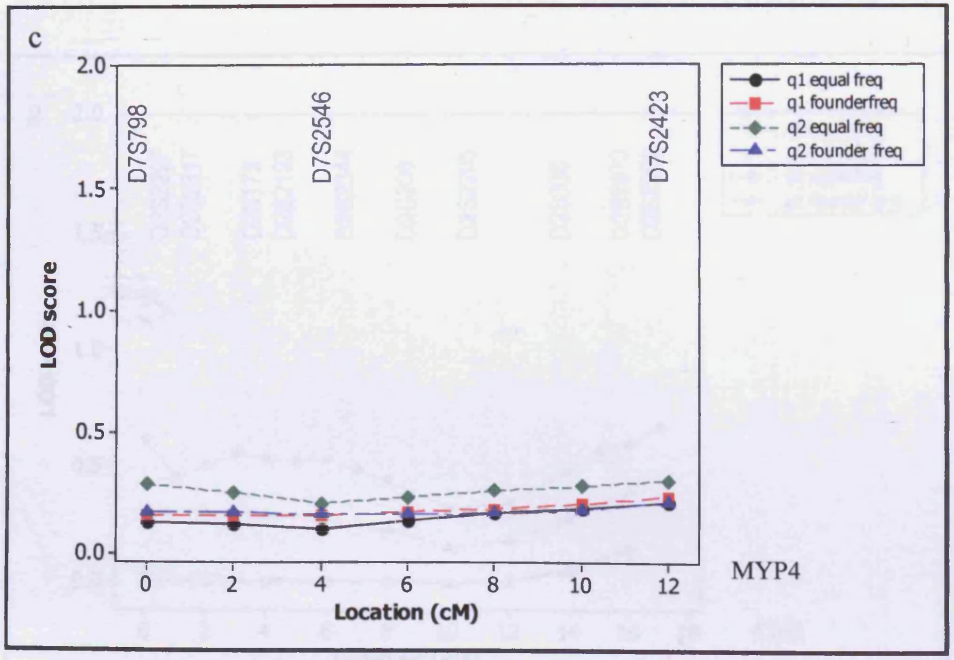
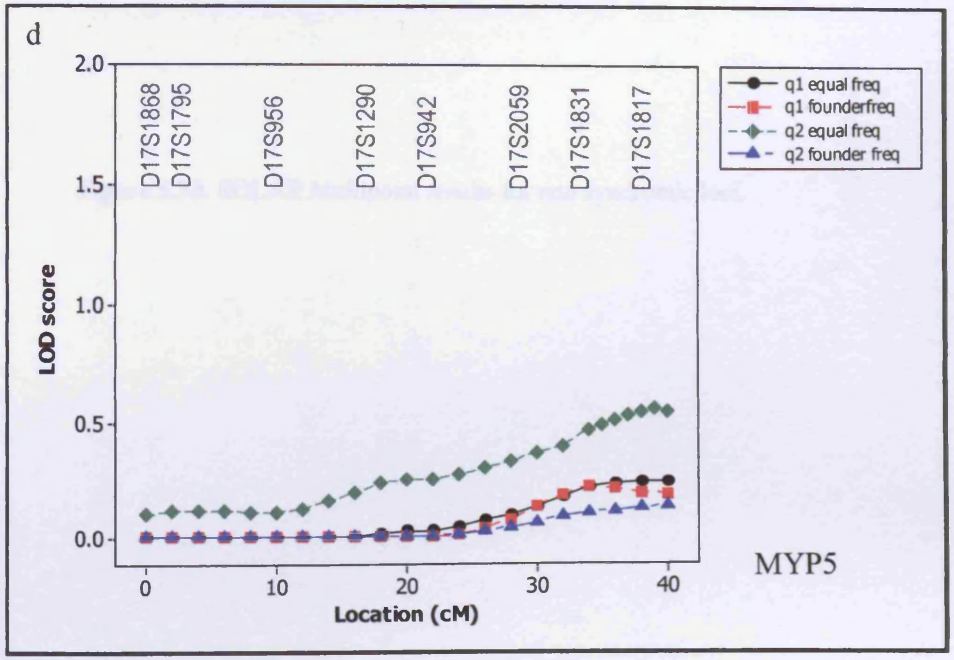


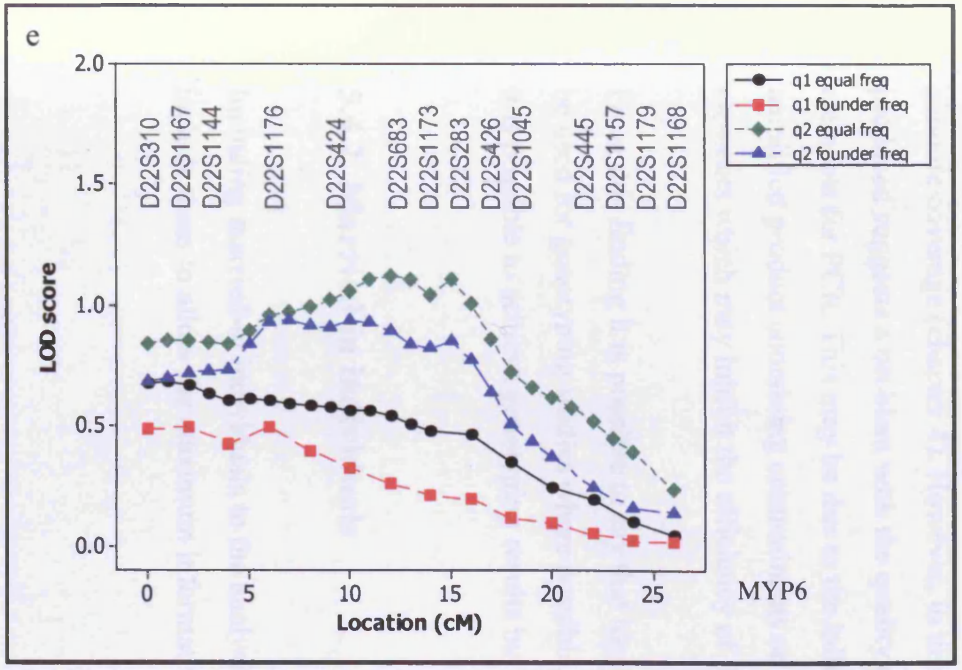
Figure 5.9i. SOLAR Multipoint scores for syndromic loci.

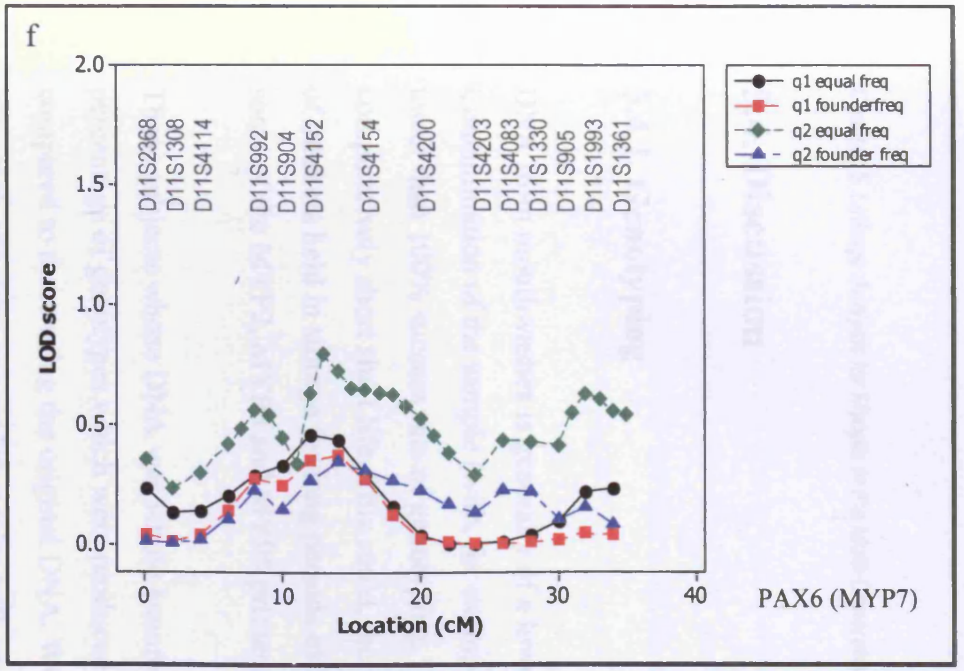

a
LOD score
Location (cM)
D18S59
D18S476
D18S481
D18S63
D18S1132
D18S452
D18S1163
D18S464
D18S1158
q1 equal freq
q1 founderfreq
q2 equal freq
q2 founder freq
MYP2


b
LOD score
Location (cM)
D12S80
D12S326
D12S64
D12S316
D12S348
D12S332
D12S1607
D12S78
D12S1605
D12S354
q1 equal freq
q1 founderfreq
q2 equal freq
q2 founder freq
MYP3


c
LOD score
Location (cM)
D7S798
D7S2546
D7S2423
q1 equal freq
q1 founderfreq
q2 equal freq
q2 founder freq
MYP4


d
LOD score
Location (cM)
D17S1868
D17S1795
D17S956
D17S1290
D17S942
D17S2059
D17S1831
D17S1817
q1 equal freq
q1 founderfreq
q2 equal freq
q2 founder freq
MYP5





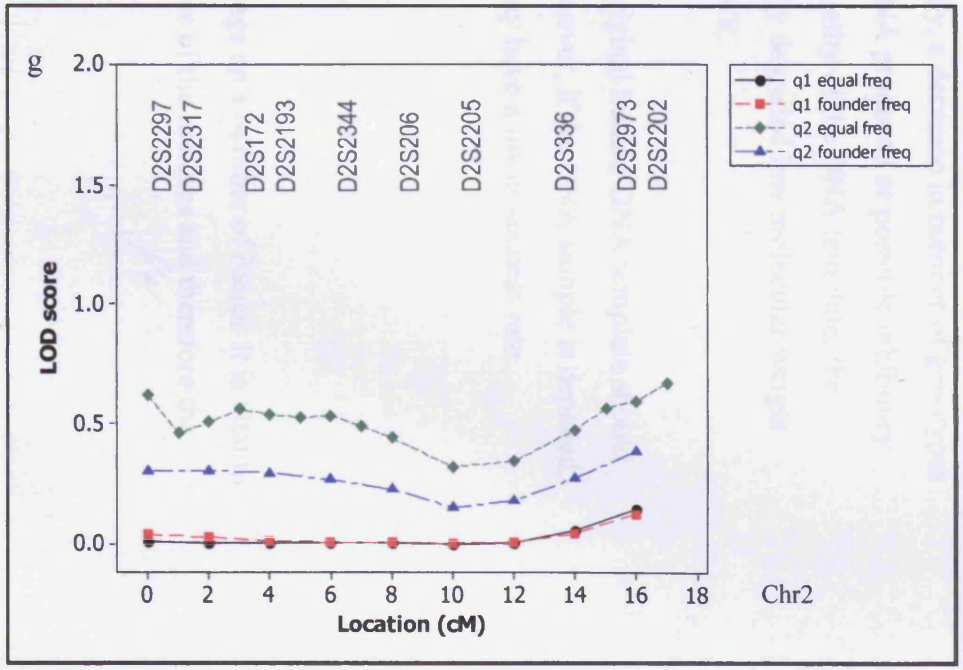


Figure 5.9ii. SOLAR Multipoint results for non syndromic loci.

5.4. Discussion

5.4.1. Genotyping

DNA from mouthwashes is generally of a lower quality than from tissues such as blood. Contamination of the sample (with, for example, non-human DNA) could account for a lower than 100% success rate of genotyping. Fluorescently labelled primers have a comparatively short shelf life. This could, therefore, also have reduced the success rates of markers held in storage for long periods of time such as COL2A1 and COL11A1 and some of the MYP2, MYP3 and MYP5 primers.

Those subjects whose DNA was whole genome amplified had, on average, a lower percentage of genotypes which were produced out of the number of reactions performed compared to those using the original DNA. WGA-amplified DNA has been shown to have high genotype concordance rates with non-amplified DNA and also complete genome coverage (chapter 4). However, in this study, a decrease in number of genotypes produced suggests a problem with the quality of DNA produced or possible inhibitory elements for PCR. This may be due to the initial quality of the DNA template, the amplified product containing contaminants or highly degraded low molecular weight elements which may inhibit the efficiency of the PCR.

From this finding it is possible to say that ideally original human DNA template should be used for genotyping studies where possible. However, if the DNA sample is depleted, it is possible to achieve genotyping results but it may have a lower success rate.

5.4.2. Married-in individuals

Including married-in individuals to the analysis brings up a number of issues. It is vital to include them to allow the maximum informativeness of the meioses and therefore the

gene flow pattern is able to be determined with more accuracy. However, the inclusion of myopic married-in individuals potentially increases the heterogeneity of the disease decreasing the power of the linkage analysis to detect linkage. The trait distribution of this group of individuals appears to show less bias towards myopia, having a prevalence of only 16.7% and includes no high myopes therefore the effect should be minimal.

5.4.3. Linkage Analysis

5.4.3.1. Syndromic loci

The pedigree showed some slight evidence of linkage to markers at the COL2A1 locus (Stickler syndrome). However, the highest LOD score is only approximately 0.8 and therefore not statistically significant.

COL11A1 and FBN1 loci could be excluded since the LOD scores at these loci were invariably close to zero.

5.4.3.2. Non syndromic loci

The most likely regions to be linked with myopia are MYP3 and MYP6 with LOD scores of over 1 and therefore suggestive of linkage but, again, not statistically significant. The twopoint analysis peaks of the linkage signal of MYP3 occur within the same 14cM region as was found for previous studies including that of a UK population cohort (Farbrother et al., 2004b) for the analysis using trait q1 and also q2 with founder allele frequencies. However, the highest LOD score occurred for the q2 trait with equal allele frequencies, which gave a peak outside this region.

MYP6 linkage analysis peaks also occurred within the same region as found by Stambolian and colleagues investigating linkage analysis with low myopia (Stambolian et al., 2004). The highest twopoint LOD scores (1.19) occurred only approximately 6cM away from the peak found in the former study. The same region also encompasses the higher multipoint LOD scores (highest multipoint LOD score of 1.13 with q2 and equal allele frequencies). The heritability at these regions suggest between 18 to 45% of the variance is due to that specific QTL.

Considering the high power produced by the SLINK simulations, a higher LOD score would be expected if linkage to a major QTL was present. Therefore, in this pedigree, these loci are unlikely to contain this major QTL.

5.4.3.3. Allele frequencies and trait description

The effect of the change in allele frequencies is important to note. The choice of using allele frequencies calculated from founders and married-in individuals or equal frequencies appears to vary the outcome of the analysis, particularly with trait q2, the transformed refractive error. The trait under investigation is also important in that q1 represents the real phenotype whereas q2 is a simulated trait based on the combination of a normally distributed phenotype and the original data, providing the software with the basis it needs for an accurate analysis. It is, therefore, difficult to favour one trait over the other, one being an accurate representation of the phenotype but potentially providing an inflated type I error rate, the other reducing the error rate but not showing an accurate representation, particularly around emmetropia.

5.4.4. Summary

The large size of the family in this study should provide an extremely powerful situation in which to detect linkage. Multiple testing and replication allow a decrease in the LOD

score threshold suggested for a significant linkage signal. This is because linkage is less likely to be detected at a point that has already been named as a region of linkage. Considering this and the results of this study, the LOD scores produced are too low to be certain of linkage especially for a pedigree with high power. Also, considering the results obtained vary with the trait and the allele frequencies used, these results do not provide evidence to support the previous myopia linkage analysis studies. Instead this study supports the theory that myopia is a highly heterogenous trait, these loci not being linked in this pedigree.

6. Association Analysis of Myocilin with High Myopia

6.1. Introduction

Linkage analysis investigates the co-segregation of two loci, generally with one being the disease locus and the other a marker locus. Association provides a different strategy: the investigation of over-representation or under-representation of a specific marker allele in diseased individuals compared to a control population (Baur and Knapp, 1997). Association and linkage analysis can be extremely valuable when used together to identify potential candidate genes. Once a chromosomal locus linked to a disease has been identified, association can be used to investigate the probability of a causal action of candidate genes within this region. However, the two strategies can also be used separately, association analysis providing greater power to detect genetic contributions to complex diseases (Cardon and Bell, 2001) particularly for those markers which are not very polymorphic, i.e. have fewer alleles therefore making it more difficult to identify from which heterozygous parent the allele is inherited from.

There are a number of methods used to investigate association, including case-control studies and family based association tests. Case-control studies compare the allele frequencies at a locus in a group of affected individuals compared to those of a control population. An increase in frequency of a specific allele within the case population suggests association with the disease. If a positive result is found, this association could have arisen due to a number of reasons. Firstly, the allele itself could directly affect the phenotype expressed. Secondly, the allele could be inherited along with a causative allele nearby on the chromosome (i.e. in linkage disequilibrium) or thirdly, that it is just a chance finding or an artefact of the test (Cardon and Palmer, 2003). Case-control methods are advantageous because they are widely used and understood, easy to recruit for and large sample sizes are easily collected. They are also beneficial if the disease studied has a late onset i.e. no parental or family phenotypic/genotypic information is required.

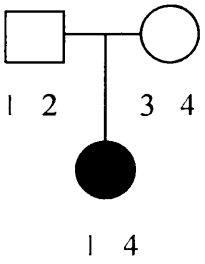
However there are limitations to this type of study. Inconsistency is the primary limitation and causes a major problem in the replication of association studies. This can be due to population stratification, which exists when there are differences in allele frequencies and disease prevalence due to diversity in background populations. Population stratification leads to false associations. This has been shown to be the main cause of non-replication and false positive association (Cardon and Palmer, 2003). One solution to avoiding population stratification is the difficult task of more careful matching of cases and controls.

An alternative method is to use family based analyses. These use an internal control, therefore avoiding population stratification entirely. One such method is called the transmission disequilibrium test (TDT) (Spielman et al., 1993). Family groups in the form of trios, an affected proband and their two parents, are collected. The equivalent "cases" are those alleles transmitted to the affected offspring from heterozygous parents and the "controls" are those alleles not passed on. No association exists when a specific allele is passed on 50% of the time whereas a value significantly different to 50% will indicate association.

6.1.1. Transmission Disequilibrium Test

The standard TDT uses only dichotomous traits and considers only heterozygous parents. Contingency tables can be produced by scoring the transmissions between parents and the affected offspring. For example, in figure 6.1, allele 2 is not transmitted to the offspring by the father in favour of the alternative allele, allele 1. The mother transmits allele 4 and allele 3 is not transmitted. These transmissions are fully informative. If a parent is homozygous, they are uninformative and so cannot contribute to the analysis. The test statistic compares the number of times the allele is transmitted from a heterozygous parent to an affected offspring to the number of times the alternative allele is transmitted. The test is designed for a marker with only two alleles. To extend its use to polymorphic

markers, the alternative alleles to that which is under investigation are grouped together. For example in figure 6.1. they are collectively labelled as "not 2".



Marker		Allele not transmitted	
		2	Not 2
Transmitted allele	2	a	b
	Not 2	c (+1)	d (+1)

Figure 6.1. Example of scoring for TDT contingency tables.

The test statistic used in this case follows a χ^2 distribution and is shown in equation 6.1.

$$T = \frac{(b-c)^2}{(b+c)}$$

Equation 6.1. Test statistic for TDT.

6.1.2. TDTEX

The TDT can be performed by the TDTEX program of the S.A.G.E. [2004] Statistical Analysis of Genetic Epidemiology, Release 5.1.0. This programme uses contingency tables as above, but investigates each allele simultaneously i.e. not grouping together alleles, and therefore producing a larger contingency table. The test statistics available include using two different null hypotheses as follows:

1. McNemar statistic: The distribution of alleles transmitted is not different to the distribution of those alleles not transmitted.
2. Marginal homogeneity: The number of alleles or genotypes transmitted is equal to the number not transmitted.

These hypotheses may be tested using allelic or genotypic transmission, the latter being preferential as there is no independence between allelic transmissions by one parent. However, the sample size would have to be much larger to avoid statistical problems caused by sparseness in the table.

The TDTEX programme will test these null hypotheses, but does not indicate which specific allele is causing the null hypothesis to be rejected. Therefore further tests can be carried out as described by Curtis (David Curtis website) and as implemented in ETDT software (Sham and Curtis, 1995). The number of times a specific allele is transmitted is compared to the number of times it is not transmitted and the test statistic is calculated as shown below, and is approximately equal to a χ^2 distribution.

$$\chi^2 = \frac{(n_t - \frac{1}{2}(n_p))^2}{\frac{1}{2}(n_p)} + \frac{((n_p - n_t) - \frac{1}{2}(n_p))^2}{\frac{1}{2}(n_p)}$$

Equation 6.2. Curtis' method of testing for association for polymorphic markers. Where n_t is the number of times that a specific allele was transmitted and n_p is the number of times that allele could potentially have been transmitted.

6.1.3. Family Based Association Test

To increase the amount of information used for the TDT, an extension to this test can be used, the Family Based Association Test (FBAT) (Laird et al., 2000). FBAT is advantageous because:

1. Quantitative traits can be analysed as well as dichotomous traits.
2. There is the potential to include covariates.
3. It is robust to missing parental genotypic data.
4. It is capable of dealing with multiple and non-normally distributed traits and
5. It is able to include unaffected offspring if available.

This test is an extension of the TDT which compares the genotype distribution observed in the cases to its expected distribution under the null hypothesis of no linkage and no association, taking into account the effect of the probands' phenotypes. The null distribution is based on the information provided by parental genotypes thereby avoiding population stratification biases (Van Steen and Laird, 2005). FBAT divides the test population into individual families, or in this case, trios, to evaluate their individual contribution to the test statistic. The number of informative families for each allele is calculated and so indicates the statistical relevance based on this value of informativeness for that particular test. If more than one locus is used, haplotypes can be built and their association evaluated. A Z statistic is calculated, producing a negative value if the transmission of alleles is decreased and a positive value if there is over transmission of that allele to the affected offspring. The test can also be performed based on genotypes as opposed to alleles.

The distribution of the test's score statistic, S, under the null hypothesis, is calculated based on the distribution of the affected offspring's genotype, conditional on the trait, and on the parental genotype. The actual observed genotype distribution is then compared to the null, the deviation providing the test statistic. Therefore if the distribution of the

affected offsprings' genotypes differs significantly from the null, association between that marker and the trait exists.

6.1.4. Power and sample size

The power of the test increases with sample size, but also depends on a number of other factors including the disease parameters and modes of inheritance. A software package within FBAT, called PBAT, provides tools for the design and analysis of association studies and can be used to calculate the power of the family based association test. By inputting the pedigree structure and ascertainment strategy along with genetic model parameters such as gene frequencies and disease prevalence, the power of the study can be calculated by simulation or approximate methods.

Haines and Pericak-Vance (Haines and Pericak-Vance, 1998) estimated the number of trios required to gain a power of 0.8 and a type 1 error rate of 0.05. For an autosomal dominant model with a disease allele frequency of 0.01, penetrance of 0.8 and disease prevalence of 0.0159, 194 trios would be required if the relative risk was 2.9, this sample size requirement increases with decreasing relative risk. For a recessive model where disease allele frequency is 0.20, penetrance is 0.5 and disease prevalence is 0.02, 86 trios are required for a study of equal power and type 1 error rate as above with a relative risk of 2.9. These figures demonstrate that the mode of inheritance and the disease allele frequency have a marked effect on the power of family based association tests, and provide a guide to the number of trios required for such studies.

6.1.5. Hardy-Weinberg Equilibrium

For association studies, the sample must be in Hardy Weinberg equilibrium. That is, the genotype frequencies are not changed between generations. In theory, genotype frequencies may be changed by evolutionary forces such as mutation, genetic drift or

migration into or out of a population (Haines and Pericak-Vance, 1998). However, the low rate of mutation suggests only a small role in evolution. Genetic drift occurs when allele frequencies change by chance due to factors such as mating patterns in small populations. Migration occurs when populations with a specific gene pool then breed outside their gene pool, introducing new alleles or altering current allele frequencies.

The Hardy-Weinberg theory has important applications including the ability to determine whether the trait has a strong genetic basis and also helps to determine carrier frequencies of diseases, disease prevalence and gross estimates of penetrance (Haines and Pericak-Vance, 1998).

To test whether the alleles within a population are in Hardy-Weinberg equilibrium, the following theory is applied. For a bi-allelic marker, with alleles **A** and **a** with frequencies of p and q respectively, where p+q=1, HW theory predicts that genotypic frequencies of **AA**, **Aa** and **aa** are p^2, $2pq$ and q^2 respectively. If the alleles are in Hardy-Weinberg equilibrium it follows that:

$$p^2+2pq+q^2=1$$

Equation 6.3. Hardy Weinberg equilibrium.

For polymorphic markers, either one allele is tested against a pooled collection of the other alleles or each allele is tested separately using equation 6.4., this can be extended to include n alleles.

For example, if a marker has alleles **P**,**Q** and **R** with allele frequencies of p, q and r respectively, genotype frequencies are as follows: **PP**=p^2

PQ=$2pq$

PR=$2pr$

$$\mathbf{QQ}=q^2$$
$$\mathbf{QR}=2qr$$
$$\mathbf{RR}=r^2$$

And it follows, according to HW theory that

$$p^2+2pq+2pr+q^2+2qr+r^2=1$$

Equation 6.4. Extension of Hardy-Weinberg theory for polymorphic markers.

The observed frequencies are compared to the expected frequencies which follow a χ^2 distribution. A deviation from Hardy Weinberg equilibrium suggests that the genetic basis of the trait is not correctly specified, that there is non random mating or that there is preferential selection for individuals with a specific genotype within the samples (Sham, 1998).

6.1.6. Association and Myopia

There have been many studies nominating candidate genes and investigating the possible associations with myopia (chapter 1). Only a few have returned significant results, hence no actual causative genes have yet been found. Most studies have taken candidate genes from within linked loci which have biological functions likely to be related to myopia e.g. structural functions and investigated the possible associations. For example, LIPIN2, within the MYP2 locus was investigated but no significant association was found (Zhou and Young, 2005). MYP3 region candidate genes such as lumican and fibromodulin also showed no association within a linked population (Paluru et al., 2004) although decorin and DSPG were also excluded as candidate genes for MYP3, a SNP in lumican suggested

possible association in a Taiwanese population (Wang et al., 2006). Replication of association studies appears to be less easy to accomplish than expected, possibly due to differences in ethnic backgrounds but also in study designs. The myocilin studies described below appear to have achieved replication with significant association which therefore strengthens the evidence for nominating this as a candidate gene for investigation.

6.1.7. Myocilin

Myocilin is a recently discovered protein, encoded by the MYOC gene (also known as GLC1A). The MYOC gene consists of three exons, upstream of which is the promoter region, which regulates its expression (Fingert et al., 2002). Myocilin is expressed most strongly in the trabecular meshwork of the eye (hence its alternative name of TIGR-trabecular meshwork-induced glucocorticoid protein) as well as in the sclera, iris, cornea, lens, ciliary body, retina and optic nerve. It is also expressed in other organs in the human body such as the heart, stomach and skeletal muscle. Mutations in MYOC have been identified as the cause of hereditary juvenile-onset open-angle glaucoma, there being 43 reported mutations in the gene in patients with open angle glaucoma. Myocilin mutations are also found in 3-4% of primary open angle glaucoma patients. Most of these mutations occur in exon three and are believed to be disease causing. However there are also polymorphisms that exist in the rest of the gene but they may not be disease causing (Fingert et al., 2002).

The exact mechanism by which mutations in the myocilin gene cause glaucoma has not yet been determined. Studies in the mouse suggest a number of theories:

(1) a loss of function of this protein. However, this is unlikely because haploinsuffient mice do not express a glaucomatous phenotype (Ricard and Tamm, 2005).

(2) an effect on the aqueous outflow facility of the eye, due to the retention of mutated myocilin by the rough endoplasmic reticulum. The break down of the process of

protein trafficking may, in turn, cause dysfunction of the trabecular meshwork and therefore affect outflow. (Direct outflow obstruction was argued against by alternative mouse studies which found a lack of myocilin induction when IOP was high, supported by the finding that over-expression of myocilin does not cause a rise in IOP (Ricard and Tamm, 2005)).

(3) a muscle-related ciliary body mechanism. This is due to the high expression of myocilin within muscle and the action of pilocarpine to reduce IOP (OMIM #601652).

The high homology (83%) between mice and humans for this gene (Fingert et al., 1998) suggests further studies with the mouse models are likely to provide important information regarding myocilin function and the mechanisms underlying glaucoma causation in humans (Ricard and Tamm, 2005).

Myocilin also has a connection to myopia. Specifically myopia is a clinical risk factor for glaucoma; myopes having an approximate threefold greater risk of developing glaucoma but no causal effect has yet been established (Saw et al., 2005). A previous study showed association between myopia and MYOC in a Singapore Chinese population of 97 unrelated high myopes (Wu et al., 1999). Attempts to replicate this finding with a Hong Kong Chinese population using a similarly designed case-control study (using 70 unrelated cases) (Leung et al., 2000), failed, but a positive result was found in 2004 by Tang and colleagues using a family based association test and a population of 157 Chinese nuclear families. Association was noted with a microsatellite polymorphism in the MYOC promoter region. The shortest allele (consisting of 13 dinucleotide repeats) showed increased transmission to high myopes, whereas the longest allele (15 dinucleotide repeats) showed decreased transmission, providing a possible protective effect (Tang et al., 2004).

6.2. Methods

6.2.1. Association analysis

Trios, consisting of a highly myopic proband and two parents, were recruited as described in Chapter 2 (Trio pedigrees, appendix III). Spectacle prescription details and DNA samples were collected for all subjects. The subjects were genotyped for two markers, MYOC1 and MYOC2; using the protocol described in chapter 5. MYOC1 is a microsatellite within the promoter region and MYOC2 is in the 3' untranslated region (table 5.2.). The parental alleles were then tested for Hardy Weinberg equilibrium (the offspring were recruited based on their phenotype and so will not provide a fair sample representation for the population and Hardy-Weinberg equilibrium would not be expected). The allele frequencies of the parental sample were determined and, by using equation 6.4., the expected genotypic frequencies were compared to those observed.

For the standard TDT, each allele's transmissions were scored using contingency tables as described above. TDTEX was then used to test the null hypothesis. An FBAT analysis was carried out considering high myopia as a dichotomous trait; those subjects with a refractive error of -6.00D in the least minus meridian of their least myopic eye were classed as affected (Young et al., 1998b). The FBAT analysis was then repeated analysing the trait quantitatively. Association was tested using alleles, genotypes and haplotypes.

Finally, a case-control analysis was performed by dividing the parental cohort into cases (those individuals with -6.00D or more myopia in the least minus meridian of both eyes) and controls and calculating the allele frequencies and also the genotype frequencies. The groups were compared using χ^2 test.

To identify which alleles matched those associated with myopia in previous studies, the microsatellite PCR products were sequenced and number of repeats counted, as discussed below.

6.2.2. Sample preparation for DNA Sequencing

Five replicate polymerase chain reactions (PCR, chapter 1) with the appropriate unlabelled primers were carried out. The replicate samples were then pooled and Qiaquick spin column (QIAGEN) purified to remove primers. The sample was eluted in a volume of 180µl 2mM Tris-HCl, pH8 and then 2µl of 20µg/µl glycogen and 20µl of 3M sodium acetate were added. The samples were then vortex-mixed and briefly centrifuged. A volume of 200µl of phenol/chloroform (phenol:chloroform:isoamyl alcohol-25:24:1) was added and the sample was vortexed vigorously for 30 seconds. The tube was then centrifuged at 14000rpm for 2 minutes. When the supernatant was clear it was transferred to a fresh tube and 400µl of ethanol was added, vortex-mixed and left to precipitate at −20°C for 20 minutes. The samples were centrifuged at 14 000rpm for 10 minutes, the supernatant was then discarded and 400µl of ice-cold 70% ethanol was added. The mixture was centrifuged at 14 000rpm for 2 minutes and the supernatant was discarded. The pellet was air dried in an inverted position for 5 minutes and then resuspended in 20µl of PCR water. The template was quantified by spectrophotometry (chapter 4) and diluted to a concentration of 25ng/µl. The sequencing section of the Biosciences department of Cardiff University were then sent 10µl of 25ng/µl template DNA and 5µM forward primer and the sequencing was performed on an ABI 3700 sequencer.

6.3. Results

See chapter 2 and trio pedigrees (Appendix III) for a full description of the subjects involved.

6.3.1. Association Analysis

6.3.1.1 Transmission Disequilibrium Test

Contingency tables were constructed by scoring the transmissions for one allele versus all other alleles. A summary of the results is shown in tables 6.1. and 6.2.

Table 6.1. TDT results for MYOC1.

MYOC1 Allele	χ^2	P value
1	2.27	0.32
2	3.45	0.17
3	4.27	0.12
4	3.63	0.16
5	0.33	0.85

Table 6.2. TDT results for MYOC2.

MYOC2 Allele	χ^2	P value
1	0.11	0.95
2	0.01	0.99
3	0.00	1.00
4	0.21	0.90
5	0.20	0.91

No statistically significant association is present between myopia as a dichotomous trait and MYOC1 and MYOC2 when analysing with a simple TDT and scoring transmissions with one allele versus all other alleles.

6.3.1.2. TDTEX

Contingency tables were constructed by scoring transmissions for each allele separately and simultaneously for MYOC1 (table 6.3.) and MYOC2 (table 6.4.).

Table 6.3. TDTEX contingency table for MYOC1 alleles.

Transmitted allele \ Allele not transmitted	1	2	3	4	5
1	*1*	*1*	*0*	*1*	*0*
2	*3*	*41*	*17*	*21*	*2*
3	*4*	*20*	*6*	*10*	*0*
4	*1*	*14*	*2*	*11*	*0*
5	*0*	*1*	*0*	*0*	*0*

Table 6.4. TDTEX contingency table for MYOC2 alleles.

Transmitted allele \ Allele not transmitted	1	2	3	4	5
1	*0*	*0*	*0*	*5*	*0*
2	*1*	*18*	*4*	*23*	*0*
3	*0*	*4*	*2*	*3*	*0*
4	*5*	*23*	*3*	*55*	*3*
5	*0*	*2*	*0*	*0*	*0*

Table 6.5. shows the summary of TDTEX results and significance levels of each null hypothesis tested for both alleles and genotypes.

Table 6.5. Summary of TDTEX results

Data	Number of informative pedigrees	McNemar test	Continuity corrected McNemar test	Marginal homogeneity test
MYOC1 alleles	78.79%	0.045*	0.191	0.019*
MYOC1 genotypes	78.79%	0.537	1.000	0.089
MYOC2 alleles	76.77%	0.540	0.945	0.986
MYOC2 genotypes	75.76%	0.613	0.996	0.779

**significant at $p \leq 0.05$*

TDTEX results show that there is a statistically significant lack of symmetry between transmitted and non transmitted alleles for MYOC1. The null hypothesis could not be rejected for MYOC2.

The analysis was repeated with genotypes but found no statistically significant difference from the null hypothesis for either MYOC1 or MYOC2.

This TDTEX analysis shows there is a difference in the distributions of the allele transmissions but does not pinpoint the specific allele involved. Therefore using the method suggested by Curtis as described above, the results shown in table 6.6. were calculated.

Table 6.6. Results determining specifically which allele has increased or decreased transmission for MYOC1.

Allele	χ^2	p	Transmission
1	3.60	0.06	-
2	0.62	0.43	-
3	4.16	0.04*	Increased
4	4.59	0.03*	Decreased
5	0.33	0.57	-

**significant at p≤0.05. Allowing for multiple testing p≤0.01, no results are significant.*

6.3.1.3. Family Based Association Test

The Z statistic shows the transmission state; a positive number means an increase in transmission, a negative value is equivalent to a decrease in transmission.

Table 6.7. FBAT results for MYOC1, allelic mode

Allele	Allele frequency	No. of families	Z statistic	
			Dichotomous	Quantitative (reversed sign)
1	0.04	9	-1.67	-1.73
2	0.52	65	0.85	0.86
3	0.19	52	1.81	1.96* *(p=0.05)*
4	0.24	45	-2.18* *(p=0.03)*	-2.25* *(p=0.02)*
5	0.02	3	-0.58	-0.78

**significant at p≤0.05*

Table 6.8. FBAT results for MYOC2, allelic mode

Allele	Allele frequency	No. of families	Z statistic	
			Dichotomous	Quantitative (reversed sign)
1	0.04	10	-0.30	-0.50
2	0.30	59	0.00	0.24
3	0.04	15	0.00	-0.12
4	0.61	62	0.23	0.06
5	0.02	5	-0.45	-0.23

**significant at p≤0.05*

FBAT results show no significant association of myopia with MYOC2. However, for MYOC1, there is a significant decrease in transmission of allele 4 when analysing the trait as both a dichotomous and a quantitative trait. Quantitative analysis also shows increased transmission of allele 3, but this is only just significant and does not occur with the dichotomous analysis.

Table 6.9. FBAT results for MYOC1, genotype mode.

MYOC1 Genotype	Genotype frequency	No. of families	Z statistic	
			Dichotomous	Quantitative (reversed sign)
1/1	0	0	-	-
1/2	0.04	9	-0.85	-0.53
2/2	0.28	49	1.29	1.37
1/3	0.03	2	-0.82	-0.80
2/3	0.18	46	0.94	1.16
3/3	0.04	9	1.13	1.04
1/4	0.01	2	-1.13	-1.14
2/4	0.23	44	-2.00* *(p=0.05)*	-2.20* *(p=0.03)*
3/4	0.08	28	0.73	0.98
4/4	0.07	11	-0.85	-0.75
1/5	0	0	-	-
2/5	0.03	3	0	-0.23
3/5	0	1	-0.58	-0.58
4/5	0	1	-0.58	-0.58
5/5	0	0	-	-

**significant at p≤0.05*

Table 6.10. FBAT results for MYOC2, genotype mode.

Genotype	Genotype frequency	No. of families	Z statistic	
			Dichotomous	Quantitative (reversed sign)
1/1	0	1	-0.58	-0.58
1/2	0	4	1.16	0.75
2/2	0.10	20	0.61	1.03
1/3	0	1	-0.58	-0.58
2/3	0.06	4	0.54	0.55
3/3	0	2	0.38	0.39
1/4	0.07	9	-0.35	-0.52
2/4	0.32	57	-1.01	-1.27
3/4	0.02	14	-0.14	-0.30
4/4	0.39	51	0.73	0.88
1/5	0	1	-0.58	-0.58
2/5	0.02	1	-0.58	-0.58
3/5	0	1	-1.00	-1.00
4/5	0.02	4	0.54	0.95
5/5	0	0	-	-

Using the FBAT analysis and genotypes as opposed to alleles, there was a statistically significant decrease in transmission of the genotype 2/4 for MYOC1 as for a dichotomous trait and a quantitative trait at significance values of p=0.05 and p=0.03 respectively, but no statistically significant result for MYOC2 genotypes (tables 6.9 and 6.10.).

FBAT also allows analysis of the haplotypes at the MYOC1 and MYOC2 loci using the Haplotype Family Based Association Test (HBAT). These results are shown in table 6.11.

Table 6.11. HBAT results showing only those haplotypes with more than 10 informative families.

Haplotype MY1 MY2	Haplotype frequency	No. of families	Z statistic	
			Dichotomous	Quantitative (reversed sign)
2 4	0.30	61.0	1.89	1.88
2 2	0.23	48.5	-0.99	-0.87
4 4	0.16	38.0	-2.08* *(p=0.04)*	-2.07* *(p=0.04)*
3 4	0.12	29.5	0.55	0.60
3 2	0.07	23.5	2.24* *(p=0.03)*	2.18* *(p=0.03)*
2 3	0.04	11.0	-0.30	-0.43
4 1	0.03	10.0	-0.30	-0.50

**significant at* $p \leq 0.05$

Only the results for the haplotypes with more than 10 informative families are shown in table 6.11. The results suggest a decrease in transmission of the haplotype 4 4 (allele 4 of MYOC1 marker and allele 4 of MYOC2 marker) to affected offspring and also an increase in transmission of haplotype 3 2 (allele 3 of MYOC1 and allele 2 of MYOC2).

6.3.2. PBAT power calculations

PBAT power calculations suggested this study has low power, of 17.7% by simulation. However, the method assumes a number of the model parameters which may not be accurate and therefore it can only be used as a guide. Alternatively, considering the trio phenotypes (Trio pedigrees, Appendix III), the disease gene frequency and disease prevalence, a situation similar to that proposed by Haines and Pericak-Vance (Haines and Pericak-Vance, 1998) for a recessive model is possible, suggesting approximately 86 trios consist a powerful enough cohort to detect a locus with a relative risk of disease of 2.9.

6.3.3. Case-control study

Although these cases and controls are not matched, it was interesting to carry out a comparison of allele and genotype frequencies between those parental trio members who are high myopes and those who are not. Allele frequencies were as shown in tables 6.12 and 6.15 for MYOC1 and MYOC2 respectively and represented by graphs in figures 6.2. and 6.3. Genotype frequencies are shown in tables 6.13. and 6.16. Case-control results are shown in table 6.14 for MYOC1 and table 6.17 for MYOC2.

Table 6.12. MYOC1 allele frequencies in cases and controls.

MYOC1 Alleles	Allele frequency for cases	Allele frequency for controls
1	0	0.044
2	0.578	0.568
3	0.189	0.189
4	0.222	0.196
5	0.011	0.007

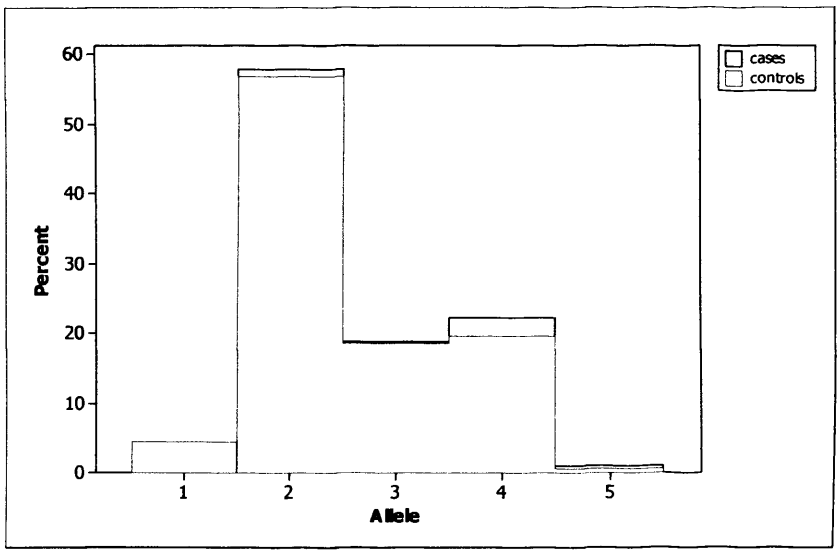


Figure 6.2. MYOC1 allele frequencies in cases (black outline) and controls (red outline).

Table 6.13. MYOC1 case-control genotype frequencies

MYOC1 genotypes	Genotype frequency for cases	Genotype frequency for controls
1/1	0	0.007
1/2	0	0.027
1/3	0	0.034
1/4	0	0.014
1/5	0	0
2/2	0.378	0.331
2/3	0.222	0.209
2/4	0.156	0.223
2/5	0.022	0.014
3/3	0	0.041
3/4	0.156	0.047
3/5	0	0
4/4	0.067	0.054
4/5	0	0
5/5	0	0

Table 6.14. MYOC1 case-control study results.

MYOC1 Alleles	χ^2 of alleles	χ^2 of genotypes
1	[Yates'] 2.85 (p=0.09)	4.24 (p=0.12)
2	0.03 (p=0.87)	0.74 (p=0.69)
3	0.00 (p=0.95)	3.65 (p=0.16)
4	0.30 (p=0.59)	0.27 (p=0.87)
5	[Yates'] 0.08 (p=0.78)	[Yates'] 0.00 (p=0.96)

Table 6.15. MYOC2 allele frequencies for cases and controls.

MYOC2 Alleles	Allele frequency for cases	Allele frequency for controls
1	0.043	0.021
2	0.404	0.271
3	0.064	0.048
4	0.479	0.644
5	0.011	0.014

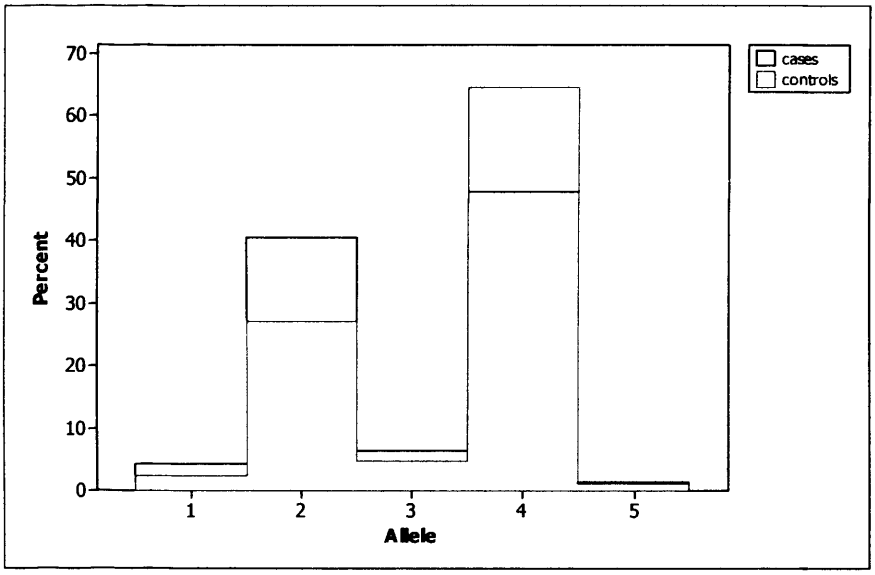


Figure 6.3. MYOC2 allele frequencies in cases (black outline) and controls (red outline).

Table 6.16. Case and control genotype frequencies for MYOC2

MYOC2 genotypes	Genotype frequency for cases	Genotype frequency for controls
1/1	0	0
1/2	0.021	0
1/3	0	0
1/4	0.064	0.047
1/5	0	0
2/2	0.170	0.100
2/3	0.106	0.033
2/4	0.340	0.280
2/5	0	0.013
3/3	0	0.013
3/4	0.021	0.033
3/5	0	0
4/4	0.255	0.440
4/5	0.021	0.013
5/5	0	0

Table 6.17. MYOC2 case-control study results.

MYOC2 Alleles	χ^2 for alleles	χ^2 for genotypes
1	[Yates'] 0.19 (p=0.44)	[Yates'] 0.38 (p=0.54)
2	5.78 (p=0.02)*	5.83 (p=0.05)*
3	0.37 (p=0.54)	2.23 (p=0.33)
4	7.88 (p=0.01)*	6.98 (p=0.03)*
5	[Yates'] 0.04 (p=0.85)	[Yates'] 0.10 (p=0.75)

**significant at $p \leq 0.05$*

There was no statistically significant difference between cases and controls when considering the alleles or genotypes separately for MYOC1 (χ^2=4.36, p=0.36 and χ^2=11.97, p=0.22 respectively). For MYOC2 no statistically significant difference was found between cases and controls when considering both allele frequencies and genotypes (χ^2=8.43, p=0.08 and χ^2=13.55, p=0.14 respectively) although by observation, there is a higher frequency of allele 2 and a lower frequency of allele 4 in cases with an

opposite increase and decrease in controls. Considering one allele separately and grouping all other alleles together, therefore reducing the number of degrees of freedom, produces χ^2 values as shown in tables 6.14 and 6.17 for MYOC1 and MYOC2 respectively. This shows no statistically significant difference between cases and controls for alleles and genotypes of MYOC1. However, an increase of allele 2 and a decrease of allele 4 are shown to be statistically significant for MYOC2.

6.3.4. Hardy-Weinberg Equilibrium

Alleles at both the MYOC1 and MYOC2 loci were in HWE. However it should be noted that this study would have low power to detect deviations from HWE.

6.3.5. Sequencing results

Figures 6.4., 6.5. and 6.6. show the sequencing traces produced for alleles 2, 3 and 4 respectively. Allele 2 consists of 13 dinucleotide repeats, allele 3 has 14 repeats and allele 4 has 15 repeats.

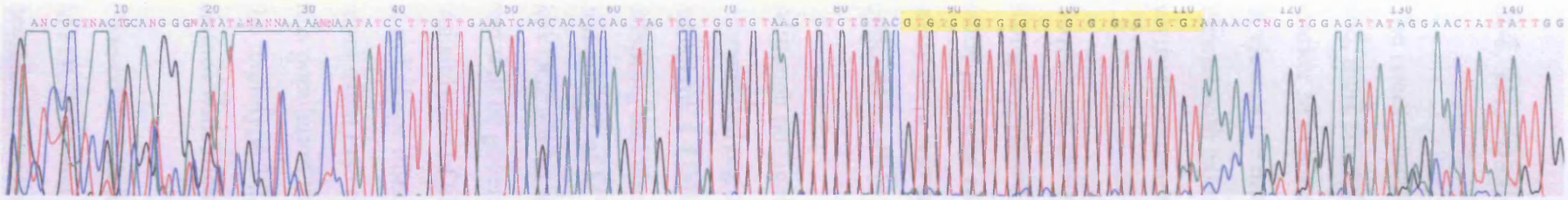


Figure 6.4. Sequence trace for MYOC1, allele 2 (13 dinucleotide repeats).

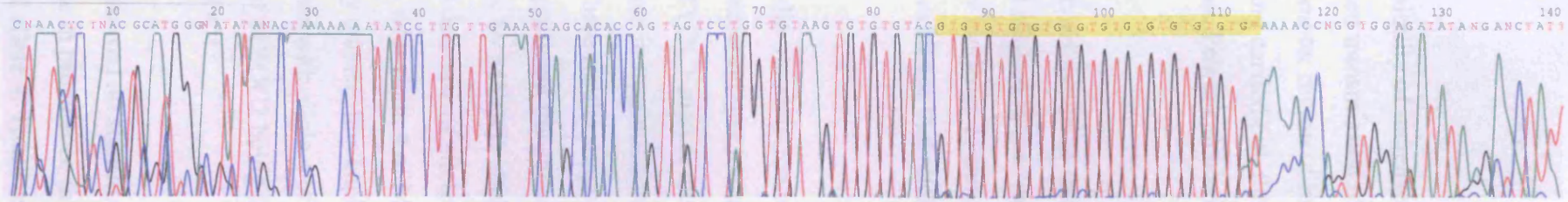


Figure 6.5. Sequence trace for MYOC1, allele 3 (14 dinucleotide repeats).

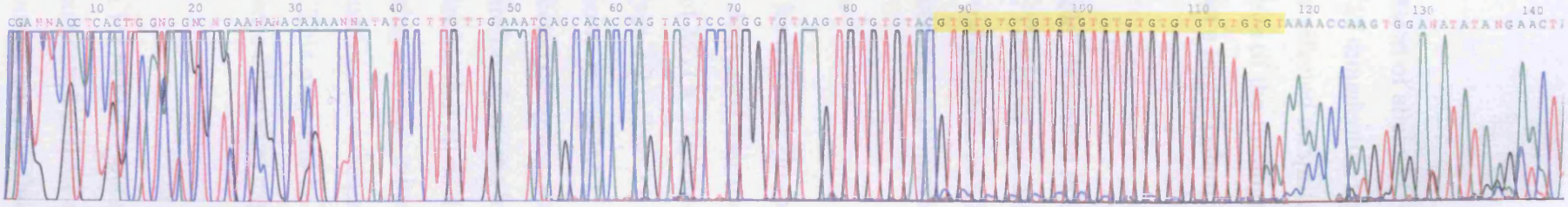


Figure 6.6. Sequence trace for MYOC1, allele 4 (15 dinucleotide repeats).

6.4. Discussion

The results suggest the possibility of increased transmission of allele 3 (14 repeats) and, more strongly, decreased transmission of allele 4 (15 dinucleotide repeats) of marker MYOC1 (in the promoter region of myocilin) to affected offspring. However, in view of the multiple testing carried out, the significance of the findings must be viewed with extreme caution. Power calculations suggest a larger cohort is required to verify these results.

Sequencing shows allele 2 corresponds to the "short allele" described in the Hong Kong association study and allele 4 to the "long allele". Therefore this association study provides confirmatory evidence of the decreased transmission of allele 4. In this study there was suggestive evidence that allele 3 has increased transmission as opposed to allele 2.

Although no significant allelic transmission was found for MYOC2 (in the MYOC 3' untranslated region), an association was found with the decreased transmission of haplotype 4 4 (allele 4 of MYOC1 marker and allele 4 of MYOC2 marker) to affected offspring and also an increase in transmission of haplotype 3 2 (allele 3 of MYOC1 and allele 2 of MYOC2). These results support those shown from FBAT, where, if the MYOC2 alleles show no significant transmission to affected offspring, the haplotype test will be governed mainly by MYOC1 alleles and therefore show the increased transmission of allele 3 and decrease of allele 4. It is possible that allele 4 of MYOC2 is in linkage disequilibrium with allele 4 of MYOC1, being not-transmitted together more often than expected by chance, this haplotype occurring more frequently within this population than any other with allele 4 of MYOC1. This may also be the case for the haplotype 3 2, allele 2 of MYOC2 being inherited more often than expected by chance with allele 3 of MYOC1.

Case-control study results showed no significant results for MYOC1 but allele 2 in MYOC2 showed an increase in prevalence in the case subjects and there was a decrease in the frequency of allele 4. These results are consistent with those found in the HBAT analysis further supporting the idea that these two alleles may be in linkage

disequilibrium with alleles 3 and 4 of MYOC1. Further tests would need to be carried out with a larger cohort that is more randomly sampled to confirm these findings, the parents being recruited because of their relationship with high myopes and to investigate further the distribution and possible association of the MYOC1 alleles.

Using the trait as a quantitative variable appears to give a slightly higher significance level and therefore produces a more powerful FBAT test.

Although each analysis is slightly different in its approach and test statistic, there is concordance of results such that there is a weak over-transmission of allele 3 of MYOC1 and under-transmission of allele 4, this decreased transmission being supported by the genotype analysis. The results provide evidence that myocilin may have a weak contribution to high myopia susceptibility in the UK population. However, as this is the first time that myocilin polymorphisms have been investigated in non East Asian subjects further replication studies are required before a causative role can be investigated.

Using this same method alternative genes could be investigated. Generally genetic association studies may make use of previous positive linkage results to identify candidate genes and attempt to detect association within these linkage regions. Alternatively as discussed here, an attempt to replicate association may be made in an alternate population. Considering this, a possible next gene to investigate would be Hepatocyte Growth Factor which has recently been found to be associated with myopia in a Han Chinese population (Han et al., 2006). Association was investigated using a family based association method and tag-SNPs and significant association was found with the polymorphism HGF5-5b on chromosome 7q21.1. A similar experimental design may be employed to attempt replication of association with this UK population of trios.

7. Final Discussion

7.1. Overview

As discussed in the preceding chapters, there is extremely strong evidence suggesting myopia is, in some way, genetically controlled. This is provided by linkage analysis and heritability studies. There is also a strong environmental component, more so in lower degrees of myopia. Farbrother et al (Farbrother et al., 2004a) found a difference in the relative risk to siblings in groups of high myopes and mild to moderate myopes. Relative risk to siblings is calculated as the risk of being myopic if a sibling is a high myope divided by the population prevalence of high myopia (chapter 1). For high myopes a value of λ_s=5.8 was calculated, but for low and moderate myopes a value of approximately λ_s=1.8 was found, suggesting an important role of genetic factors for high myopia compared to lower myopia and supporting the hypothesis that there may be more than one aetiologically distinct type of myopia. It is particularly difficult to allow for environmental variables such as near work due to the uncontrollable experimental design. The restrictions of working with human subjects include the requirement of the information collected to be observational (Rosenfield and Gilmartin, 1998).

To increase the power to detect potential linkage most genetic studies use populations or pedigrees enriched with extreme phenotypes, in this case, high myopia. This is also relevant in that high myopia is visually more debilitating; the chances of visual impairment being significantly higher due to optical qualities of refractive correction or related ocular pathologies such as retinal degeneration or detachment. The genetic basis for high myopia may then provide further information for the treatment of the condition which may also be applied to lower degrees of myopia.

Refractive error as a whole is highly heritable. However, there are many components which together make up the optical system and therefore determine the refractive error. Using the ocular component data, it may be possible to identify quantitative

trait loci for separate components of this system which may then have a major effect on the overall refractive error. The Irish-Welsh pedigree investigated showed high heritability of refractive error and corneal curvature. Despite the fact that axial length is highly correlated with refractive error and that previous studies have suggested that this is the determining component, a statistically insignificant heritability value for axial length was calculated for this pedigree. This could be due to a visual feedback mechanism, the growth of the axial length being controlled by the attempted maintenance of a clear retinal image which may contribute to the final refractive error.

By genotyping areas of chromosomes from buccal cell DNA and using variance components linkage analysis, the hypothesis of myopia as a strongly heterogenous disease was supported by the failure to replicate linkage to those loci previously found to be linked to high myopia for this large pedigree. Power studies showed a potentially high likelihood of detecting linkage if it was present and so the suggestion of linkage, having a LOD score over 1, was not convincing for this pedigree.

A battery of association tests showed the suggestion of weak association of the MYOC gene and high myopia. Initially a statistically significant difference in the transmission compared to the non-transmission of MYOC1 alleles was found. Further investigation of this showed an increase in the transmission of allele 3 of MYOC1, the microsatellite marker located in the 5' UTR of the MYOC gene, and a decrease of allele 4 of the same marker. However, the adjustment for multiple testing decreased the power and the results were no longer statistically significant. This trend of transmission of these alleles was supported by FBAT quantitative analysis but due to the inconsistency of the results, they should be taken as a suggestion and an extension of the analysis by way of increasing the sample size is necessary to establish an association of this gene with myopia.

This study has also supported the methodology behind the use of mouthwashes and buccal cells as a source of DNA for genetic studies. The quality of the DNA is sufficient to produce genotypic information and if the samples are depleted the use of

whole genome amplification methods can be applied but with some detrimental effect to the quality of the genotyping.

Detecting the possible genes in which causal polymorphisms exist will increase the understanding of the aetiology of the disease and therefore provide further clues as to the potential treatments to slow the progression of myopia and reduce the possible deterioration of vision by refractive or pharmaceutical methods.

7.2. Future work

Considering the potential power to detect linkage of the Irish-Welsh pedigree, a genome scan would provide invaluable information for the discovery of myopia susceptibility loci. Exclusion of the loci found to be previously linked to myopia suggests the disease to be highly heterogenous and therefore any potentially identifiable loci may be novel.

This genome scan could be performed either by using SNPs or microsatellite markers and genotyping the pedigree members across the whole genome. This can be done using the same method as performed here using SNPs spaced less than 1cM along the chromosomes or microsatellites markers spaced 10cM apart. Genotyping is performed and the analysis repeated using the genotypes at each chromosomal locus for all of the chromosomes. This can be done by hand or via a service provided by a number of companies in USA.

Possible linkage peaks may lead to further investigation of specific candidate genes. Currently no candidate genes for myopia have been identified but structural component genes seem to be the main suspects. Association studies are the key to these findings once the linkage peaks have been identified. Although no statistically significant results with these populations tested have been produced, MYOC seems to

have some association, if only weak, to myopia and therefore requires further investigation. An increase in study population is needed and exclusion criteria more carefully applied, potentially with the exclusion of any bilineal families.

If association is found and candidate genes are identified, the next stage is to prove biological meaningfulness by way of a functional assay. This can be done most effectively using in vivo and in vitro laboratory based methods. For example; by cloning the gene under investigation and altering its sequence to mimic the mutation, allows its effects to be examined in transgenic mouse models of the disease. Also in vitro cell culture experiments allow one to compare the protein activity of the altered gene to that of the gene with the original sequence, to observe the functional effects.

The continued recruitment of Family Study of Myopia pedigrees is an essential requirement for this investigation into myopia development. This cohort now comprises a large number of families with varying pedigree structures and also includes affected individuals with no other family members. To date there are a total of 150 families, 94 of which include two highly myopic siblings. These 150 families are made up of 1451 individuals and include 104 trios. There are also an additional 108 highly myopic individuals acting as "affected random probands". This cohort has the potential to contribute to many kinds of genetic analyses including wide spread heritability studies, linkage analysis, case-control and family based association studies. The size and increased prevalence of high myopia compared to the general population make it a powerful and near homogenous population for ongoing study.

Appendices

Appendix I

«Title» «Forenames» «Surname»
«Address1»
«Locality»
«Town»
«County»
«PostCode»

Date

Dear «Title» «Surname»

The Family Study of Myopia research project

Conlons Opticians is pleased to be helping researchers at Cardiff University in their work towards finding a cure for high myopia (high degree of short sightedness). Their research project, The Family Study of Myopia, is open to all high myopes and their families across the U.K. The aim of the research is to prevent the progression of high myopia by discovering the genes that cause it. The study is being funded by two Eye Research charities.

Further details about the research project are enclosed, along with a questionnaire and consent form to fill in if you are interested in participating. Importantly, if you would like to take part, you would not need to travel, all of the information can be obtained by post.

If you would like to discuss the research further, you can contact one of the researchers involved at the address below.

Yours sincerely

John W. Welsby

John Welsby
M.D. Conlons Opticians

On behalf of: Rosalind Creer
The Family Study of Myopia research group
Tel. (029) 2087 5063
email: *myopia@cardiff.ac.uk*

School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol

«Title» «First_name» «Surname»
«Address1»
«Address_2»
«Address3»
«Postcode»

The Family Study of Myopia

Date

Dear «Title» «Surname»

Having had your eyes examined at Cardiff University Eye Clinic, we would like to invite you and your family to participate in the Family Study of Myopia. Details about the research project are enclosed along with a questionnaire and consent form for you to complete should you decide to take part. Importantly all the information needed can be obtained by post so you would not need to travel.

The Family Study of Myopia is a research project being undertaken at Cardiff University. The study, which is funded by two Eye Research Charities, is working towards finding a cure for high myopia (a high degree of short sightedness) and is open to all high myopes and their families across the British Isles. The aim of the project is to identify the genes that cause high myopia, which will be the first step to finding a way to prevent its development.

If you would like to discuss the research further please do not hesitate to contact one of the researchers involved via the details on the information sheet.

Yours sincerely

Rosalind Creer MCOptom
The Family Study of Myopia research group

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3NB

The Family Study of Myopia

Information about the research project

We would like to invite you to take part in The Family Study of Myopia, a research project investigating the genetic factors that lead to the development of high myopia (also known as short-sightedness).

What is the purpose of the study?

The study is investigating how myopia is inherited from one generation to the next. Our aim is to discover the genes that make some people more likely to become short-sighted than others. This will help our understanding of why myopia occurs, and in the future may aid the development of treatments for the condition.

Why have I been chosen?

We are seeking the participation of families from across the U.K. and Ireland in which there are one or more individuals with high myopia. We are looking for the help of about 200 such families in total.

Who is organising the study?

The study is organised by researchers from the Department of Optometry and Vision Sciences at Cardiff University and the Medical Genetics Department at the University of Wales College of Medicine. The research is funded by two eye research charities, the National Eye Research Centre and the College of Optometrists.

What would it involve if I take part?

- We would ask you to fill in a short questionnaire about your eyesight and your general health, and also to identify other members of your family who might be prepared to take part in the study (the more members of your family who are willing to take part in the study the better, even if these relatives are not short sighted themselves).

- To enable us to trace myopia genes in your family, we would ask you to provide two mouthwash samples. These mouthwashes are easily done by swishing some saline around in your mouth for 20 seconds. The equipment and instructions will be posted to you if you agree to take part. We can assure you that these samples will only be used for studying myopia genes, and that all samples will be coded in order to protect your anonymity

Please turn over...

- We would ask for your permission to contact your Optometrist/Optician for details of your spectacle or contact lens prescription and your ocular health.

Will my confidentiality be maintained?
We take great care to ensure that the confidentiality of participating families is maintained. All personal details are kept securely, and the findings from this research will not identify individuals.

How do I participate?
If you would like to take part, please fill in the enclosed questionnaire and consent form and return them to us in the Freepost envelope provided. We will contact you with details about the mouthwash samples at a later date.

Contact for further information
If you have any further questions then we would be very happy to answer them either by telephone on 029 20875063, by post at the address overleaf or via email at *myopia@cardiff.ac.uk.*

Many thanks,

Rosalind Creer
The Family Study of Myopia

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3YJ

The Family Study of Myopia

Study Questionnaire

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....

First names .....

Date of birth .....

Address .....

.....

.....

Tel. Number .....

Please tick the box which you feel best describes your ethnic group:

- ☐ White European
- ☐ Other European
- ☐ African
- ☐ Asian
- ☐ American
- ☐ Afro-Caribbean
- ☐ Australasian
- ☐ Other (please specify)

1. At what age did you begin to wear spectacles? .....Yrs
2. What is the name and address of your current optician/optometrist?

.....

.....

3. Were you born prematurely? Yes ☐ No ☐ Don't know ☐
4. Did you have any eye condition/disease at birth or in childhood? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
5. Do you currently suffer from any eye condition or disease? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
6. Have you had any eye surgery? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
7. Do you take any medication for your eyes? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
8. Do you take any medication for any other health condition? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....

Please supply the names and addresses of any relatives who you think might be willing to participate in the research project. The participation of relatives who are not short-sighted is just as valuable as those that are. Similarly, the participation of your spouse would be very helpful, if appropriate.

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

Please continue on the next page...

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel. number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

Many thanks for your help

Conlons
— OPTICIANS —
"Experience the Difference"

The Family Study of Myopia research project

Conlons Opticians is pleased to be helping researchers at Cardiff University in their work towards finding a cure for high myopia (a high degree of short sightedness). Their research project, The Family Study of Myopia, is funded by two Eye Research Charities and is open to all high myopes and their families across the U.K. The aim of the research is to identify the genes that cause high myopia, which will be the first step to finding a way to prevent its development.

If you are interested in taking part in the project or would like more information, please complete the section overleaf and return this card (no stamp required) stating how you would like to be contacted. Importantly, if you would like to take part, you would not need to travel, all of the information can be obtained by post.

Yours sincerely

John W. Welsby

John Welsby
M.D. Conlons Opticians

On behalf of: Rosalind Creer
The Family Study of Myopia research group
Tel. (029) 2087 5063
email: *myopia@cardiff.ac.uk*

Fold along this line

no stamp required

The Family Study of Myopia

Rosalind Creer
The Family Study of Myopia
Optometry and Vision Sciences
FREEPOST SWC4376
CARDIFF
CF10 3GZ

PQ1C

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3YJ

The Family Study of Myopia

Study Questionnaire

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....

First names .....

Date of birth .....

Address .....

.....

.....

Tel. Number .....

Please tick the box which you feel best describes your ethnic group:

- ☐ White European
- ☐ Other European
- ☐ African
- ☐ Asian
- ☐ American
- ☐ Afro-Caribbean
- ☐ Australasian
- ☐ Other (please specify)

1. At what age did you begin to wear spectacles? .....Yrs
2. What is the name and address of your current optician/optometrist?

.....

.....

3. Were you born prematurely? Yes ☐ No ☐ Don't know ☐
4. Did you have any eye condition/disease at birth or in childhood? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
5. Do you currently suffer from any eye condition or disease? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
6. Have you had any eye surgery? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
7. Do you take any medication for your eyes? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....
8. Do you take any medication for any other health condition? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details .....

We are currently in contact with your:

☐ Sister .....

☐ Brother .....

☐ Mother .....

☐ Father .....

☐ Daughter .....

☐ Son .....

☐ Other .....

Aunt	Uncle	Grandmother	Grandfather	Other(please specify)

☐ Other .....

Aunt	Uncle	Grandmother	Grandfather	Other(please specify)

☐ Other .....

Aunt	Uncle	Grandmother	Grandfather	Other(please specify)

Please supply the names and addresses of any additional relatives who would be willing to participate in the research project. The participation of relatives who are not short-sighted is just as valuable as those that are.

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....

First names .....

Date of birth .....

Address .....

.....

.....

Tel.number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son

☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify): .....

Please continue on next page...

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....-
.....
Tel.number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify):

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel.number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify):

****

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....
First names .....
Date of birth .....
Address .....
.....
.....
Tel.number .....

Is this relative short-sighted? ☐ Yes ☐ No ☐ Don't know

Please indicate their relationship to you:

☐ Sister ☐ Brother ☐ Mother ☐ Father ☐ Daughter ☐ Son
☐ Uncle ☐ Aunt ☐ Husband ☐ Wife ☐ Grandfather ☐ Grandmother

☐ Other (please specify):

*Many thanks for your help, we will contact you in the near future about the mouthwash samples

School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol

The Family Study of Myopia

Consent Form for the Family Study of Myopia

Please tick boxes

I agree that my Optometrist/Optician can be contacted for further details about my eyes and health. ☐

I agree that other members of my family may be asked to take part in this study. ☐

I agree to provide mouthwash samples, which will be used to trace the passage of myopia genes through my family. ☐

I have been given an information sheet and have been given an opportunity to discuss the research. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time without my legal rights being affected. ☐

I agree to take part in this study. ☐

Name	Date	Signature
Name of parent/guardian (if applicable)	Date	Signature
ROSALIND CREER		
Researcher's name	Date	Signature

Form: CO(P)

School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol

The Family Study of Myopia

Dear

Your patient has kindly agreed to participate in the Family Study of Myopia. We would be very grateful if you would complete the enclosed form and return it in the envelope provided.

We would also be very grateful for your help with patient recruitment for our research, which aims to identify genetic factors influencing susceptibility to high myopia development.

Our research relies on discovering highly myopic brothers and sisters, ideally along with their extended family and your help will be extremely valuable to us.

You can help by informing any of your patients who meet the criteria about the study by providing them with one of the information packs provided. **The criteria for inclusion are:**

Prescription **(using negative cylinder form and applying to each eye)**

Least minus meridian - 6.00DS or over

Anisometropia less than 3.00DS

No *systemic pathology or syndromes* known to cause myopia

Over 7 years of *age*

We are collecting patient details in order to trace genes through families and will not be in any way involved in their future optometric care. Collection of DNA will be done by postal collection of a mouthwash sample, for which we will contact the patient directly. More details are displayed at our website http://www.cardiff.ac.uk/myopia/ .

If further information or more forms are required I can be contacted directly on tel. 00 44 2920 875063, at the above address or via email myopia@cardiff.ac.uk .

Many thanks in anticipation of your assistance with our research.

Yours sincerely

Rosalind Creer MCOptom

Form OP1

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3NB

The Family Study of Myopia

Optometric History Form (To be completed by optometrist)

Questionnaire and consent form returned by patient?	Yes	No

Patient details

Name .....

Address .....

.....

.....

.....

D.O.B .....

Ocular details

Date of last eye examination.....

Distance spectacle prescription

	SPH	*CYL*	*AXIS*	*VA*		*SPH*	*CYL*	*AXIS*	*VA*
RE					LE				

B.V.D.....

Patient ocular history (with particular reference to ocular problems or disease associated with myopia e.g. nuclear sclerotic cataract, keratoconus, retinal detachment etc.)

.....

.....

Patient general health (with particular reference to conditions associated with myopia)

.....

.....

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3YJ

The Family Study of Myopia

Thank you for your help with the Family Study of Myopia so far. We are now in the process of collecting mouthwashes and would be grateful if you would please follow the instructions below to provide your samples.

Mouthwash Instructions

The mouthwash is quick and painless, although it does taste quite salty. If you swallow any mouthwash it will not harm you in any way.

You should have received two plastic containers, each with your name written on the side, containing sterile mouthwash solution.

We suggest that you do the mouthwash before breakfast and before brushing your teeth.

PLEASE FOLLOW THESE INSTRUCTIONS

1. Pour the mouthwash solution from one of the containers into your mouth.
2. Vigorously swish the solution around your mouth for *at least* 20 seconds.
3. Carefully spit the solution back into the plastic container.
4. Repeat steps 1-4 for a second sample.
5. Please post your 2 mouthwash samples to us, *as soon as possible*, in the stamped addressed envelope provided. (Please screw the caps on tightly)

Usually two mouthwash samples are sufficient for our analyses. However each individual is different, therefore if we find that there is not enough DNA in the samples we may contact you again to provide some more.

If you have any questions please contact Rosalind Creer or Jez Guggenheim on (029) 20 875063

Thank you for your time

Appendix II

Optometry and Vision Sciences
Cardiff University
Redwood Building
King Edward VII Avenue
Cardiff
CF10 3YJ

The Family Study of Myopia

Study Questionnaire

Title	☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)
Surname	.....
First names	.....
Date of birth	.....
Address	.....
	.....
	.....
	.....
Tel. Number	.....

1. At what age did you begin to wear spectacles? .....Yrs
2. Would you agree to your optician/optometrist being contacted for further details about your eyes and health?
   Yes ☐ No ☐
3. What is the name and address of your current optician/optometrist?

.....

.....

4. Were you born prematurely? Yes ☐ No ☐ Don't know ☐
5. Did you have any eye condition/disease at birth or in childhood? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details.....
6. Do you currently suffer from any eye condition or disease? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details.....
7. Have you had any eye surgery? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details.....
8. Do you take any medication for your eyes? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details.....
9. Do you take any medication for any other health condition? Yes ☐ No ☐ Don't know ☐
   If Yes, please give details.....
10. Are there any other immediate family members who have not received a questionnaire? *(see attached family tree, those with a questionnaire are indicated)* Yes ☐ No ☐ Don't know ☐
11. If yes, would they be happy for us to contact them? Yes ☐ No ☐ Don't know ☐

If Yes, please list below and state their relationship to you .

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....

First names .....

Date of birth .....

Address .....

.....

.....

.....

Tel. Number .....

Please indicate their relationship to you:

Brother	Sister	Mother	Father	Son	Daughter	Aunt	Uncle	Other (Please specify)

If known, are they myopic (wear glasses to see in the distance)? Yes ☐ No ☐ Don't know ☐

Title ☐ Mr. ☐ Mrs. ☐ Ms. ☐ Miss ☐ Other (Please specify)

Surname .....

First names .....

Date of birth .....

Address .....

.....

.....

.....

Tel. Number .....

Please indicate their relationship to you:

Brother	Sister	Mother	Father	Son	Daughter	Aunt	Uncle	Other (Please specify)

If known, are they myopic (wear glasses to see in the distance)? Yes ☐ No ☐ Don't know ☐

I agree that the details I have provided may be used for this research study

Name:_____ ***Signature*** _____ ***Date*** _____

Name:_____ ***Signature*** _____ ***Date*** _____
of parent/guardian (if applicable)

School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol

The Family Study of Myopia

Dear

We would like to thank you very much for your help with our survey of the inheritance of short-sightedness (myopia). We're very pleased to have had such a fantastic response. We have been able to build up a detailed picture of the pattern of myopia inheritance in the complete extended family, which has shown that an important role is being played by both genetic factors and non-genetic factors (such as the amount of close work carried out, perhaps).

As we have had such a good response, we now hope to extend our research to try and identify the genes that have made some members of the family become myopic and others remain non-myopic. Your family is ideal for this study, firstly because it is so large, and secondly because over 50% of your relatives are myopic, double the amount we would expect in the general population.

With this in mind, we would like to invite you to one of two family gatherings we are organising between ourselves and Adrian Murphy, one in Kilkenny and the other in Dublin. Our aim is two-fold. To investigate the connection between spectacle prescriptions and genes we need to take some eye measurements from as many family members as we can. These tests will not only tell us how short-sighted you and your relations are, but also which precise feature of each person's eyes is responsible for causing any myopia they may have. All of the tests are very similar to those you may have had done at the opticians. They are all quick and painless, and there is no need for any eye drops. We also need a DNA sample from as many members of the family as possible. We routinely obtain such DNA samples from a "swish and spit mouthwash". This involves simply swishing a salt solution around your mouth and then spitting it back into a tube. We can assure you that all of the DNA samples we collect will be coded to protect your anonymity. We maintain strict confidentiality and give our complete assurance that your DNA samples will only be used to investigate myopia genes, and not for any other purpose.

The family gatherings will also be a chance to meet up with your relatives. We will provide a hot buffet and drinks, and there will be ample opportunity to chat.

You will find enclosed an invitation and reply envelope. We would be grateful if you would complete the RSVP form and return it to us in the envelope provided as soon as possible, indicating whether you wish to attend (and if so, which venue would be most convenient, any dietary requirements and who in your family is able to attend).

If you would like to attend the family get-together, we will send you a consent form, an instruction leaflet describing how to carry out the mouthwash procedure and two tubes of mouthwash solution, so that you can bring the mouthwash samples along on the day. It would be helpful if you could also bring your distance glasses with you.

We would very much like as many people within the family as possible to attend this social gathering but understand if you are unable to. Should this be the case but you would still like to be involved with the study, please indicate this on the form in the space provided. If you have not sent your original questionnaires but would still like to come and join the study you are very welcome.

Your support in this project is greatly appreciated, Adrian Murphy and ourselves are looking forward to seeing you on the 21st or 22nd February 2004. If you have a question about any aspect of the study, we would be delighted to hear from you.

Yours sincerely

Rosalind Creer
The Family Study of Myopia research group

The Family Study of Myopia invites you to an evening of food, drink & eye measuring!

Saturday 21st February
Kilkenny River Court Hotel, The Bridge, John St, Kilkenny
or
Sunday 22nd February
Royal Dublin Hotel, O'Connell St, Dublin

starting at: 5.30pm

RSVP

Please detach this section, tick box and return in envelope

Name and address: .....
.....
.....
.....

☐ *I am able to attend* ☐ *I am unable to attend*

☐ *I am unable to attend but would like to remain involved in the study*

Please indicate which gathering you would like to attend

☐ Kilkenny ☐ Dublin

Please state any dietary requirements here (e.g. Vegetarian) .....

School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol

The Family Study of Myopia

Dear

Thank you for your reply regarding the Family Study of Myopia and the family gathering. We are very pleased that you are able to attend. I have enclosed a map of the city centre showing the location of the Kilkenny River Court Hotel and a copy of the buffet menu. The party will be held in the Pembroke Suite. There is some parking available at the hotel and there are a number of car parks marked on the map, just in case you are not familiar with the city.

In order to trace the genes for myopia throughout the family we require DNA samples in the form of mouthwashes as explained in the previous information sheet. I have enclosed labelled mouthwash tubes, two for each person, and a set of instructions on how these samples should be taken. There is also a consent form for each person to sign. I would be grateful if you would take the samples first thing on the morning of the party and bring them with you along with the consent forms. If there is a family member who is unable to come but willing to provide some samples, please bring those along also. Again, we assure you that these samples will be coded for your anonymity and will be used for looking at genes for myopia only.

If you need any more information or have any questions please do not hesitate to contact us at the address above, by 'phone on 00 44 (0) 2920 875063 or via email: *myopia@cardiff.ac.uk*

Thank you again for your help, we are looking forward to seeing you on the 21st February.

Yours sincerely

Rosalind Creer
The Family Study of Myopia research group

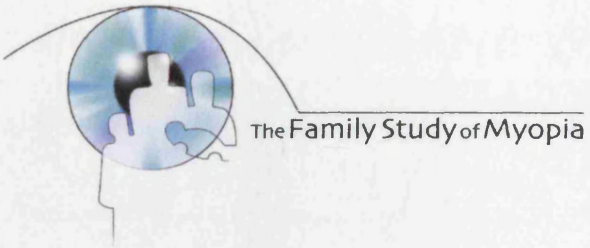
School of Optometry and Vision Sciences
Head of School Professor Tim Wess
Ysgol Optometreg a Gwyddorau Golygol
Pennaeth Yr Ysgol Yr Athro Tim Wess

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
Redwood Building
King Edward VII Avenue
Cathays Park
Cardiff CF10 3NB
Wales UK

Tel Ffôn +44(0)29 2087 4374
Fax Ffacs +44(0)29 2087 4859
www.cardiff.ac.uk

Prifysgol Caerdydd
Adeilad Redwood
Rhodfa Edward VII
Parc Cathays
Caerdydd CF10 3NB
Cymru Y Deyrnas Gyfunol



Consent Form

Please tick boxes

I agree to provide mouthwash samples, which will be used to trace the passage of myopia genes through my family. ☐

I agree to having eye measurements taken and the results being used for the study. ☐

I have been given information regarding the study and have been given the contact details of the researchers in case I need to discuss the research further. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time without my legal rights being affected. ☐

Name	Date	Signature
Name of parent/guardian *(if applicable)*	Date	Signature

Form CO(IW)

Appendix III

Table A. Phenotypic data for Irish-Welsh family. An x indicates that this measurement was not taken or an appropriate value was not able to be calculated. Subject numbers in red represent those married-in individuals.

SUBJECT NUMBER	RIGHT SPHERE (D)	RIGHT CYL (D)	RIGHT AXIS	LEFT SPHERE (D)	LEFT CYL (D)	LEFT AXIS	AUTOREF RIGHT SPH (D)	AUTOREF RIGHT CYL (D)	AUTOREF RIGHT AXIS	AUTOREF LEFT SPH (D)	AUTOREF LEFT CYL (D)	AUTOREF LEFT AXIS	AXIAL LENGTH RIGHT (MM)	AXIAL LENGTH LEFT (MM)	MEAN K RIGHT (D)	MEAN K LEFT (D)	GENOTYPED ?	SOLAR PHEN Q1	SOLAR PHEN Q2
26	-9.50	-1.00	150	-5.50	-0.75	30	-9.25	-3.05	164.49	-6.22	-1.34	19.71	27.82	26.18	43.97	43.65	y	-9.50	-7.18
27	-3.25	-0.25	25	-2.50	-0.50	135	-2.79	-0.99	33.70	-2.12	-1.36	139.70	x	x	x	x	y	-3.25	-4.42
28	-5.50	-1.50	150	-5.50	-1.25	15	-6.71	-2.46	159.58	-5.45	-1.33	6.96	27.17	26.63	42.09	42.57	y	-5.50	-5.62
29	-1.25	-	-	-1.25	-	-	-1.96	-0.46	40.72	-1.66	-0.08	147.67	x	x	x	x	y	-1.25	-2.16
30	x	x	x	x	x	x	-0.16	-0.83	90.89	-0.25	-0.98	98.13	23.69	23.77	42.94	42.67	y	-0.14	-1.04
31	-2.25	-0.50	120	-2.00	-	-	-2.53	-0.51	133.42	-2.47	-0.24	111.15	24.28	x	44.38	43.95	y	-2.25	-2.98
32	-1.25	-0.25	85	-0.75	-0.50	109	x	x	x	x	x	x	x	x	x	x	n	-1.25	-2.05
34	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	-0.53
35	-3.75	-1.00	180	-4.50	-	-	-4.00	-0.79	175.83	-4.29	-0.67	35.70	25.36	25.31	45.50	45.49	y	-3.75	-4.51
36	-3.00	-0.25	80	-2.75	-0.75	55	x	x	x	x	x	x	x	x	x	x	y	-3.00	-3.84
37	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.43
38	-2.50	-	-	-2.50	-	-	x	x	x	x	x	x	x	x	x	x	n	-2.50	-3.17
39	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.40
40	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.39
41	-6.00	-1.75	90	-5.50	-1.75	100	x	x	x	x	x	x	x	x	x	x	y	-6.00	-6.42
42	-2.00	-0.50	150	-2.00	-0.25	30	x	x	x	x	x	x	x	x	x	x	n	-2.00	-2.87
43	-5.00	-	-	-5.00	-	-	x	x	x	x	x	x	x	x	x	x	y	-5.00	-5.39
44	-1.00	-	-	-1.25	-	-	x	x	x	x	x	x	x	x	x	x	y	-1.00	-1.83
45	-2.00	-0.75	90	-2.25	-0.50	70	x	x	x	x	x	x	x	x	x	x	y	-2.00	-2.85
46	0	-	-	0	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	-0.36
47	-1.75	-0.75	102	-2.50	-0.25	90	-1.83	-1.00	108.75	-2.54	-0.49	84.71	x	x	43.38	43.16	y	-1.75	-2.59
48	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0	-0.34
49	-6.00	-0.50	120	-3.25	-1.75	90	-7.08	-1.00	118.66	-5.08	-1.57	87.60	x	x	43.92	43.50	y	-6.00	-5.89
50	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.25
51	-6.00	-0.25	75	-2.00	-0.50	70	x	x	x	x	x	x	x	x	x	x	n	-6.00	-5.88
52	-1.50	-0.75	110	-1.00	-1.25	80	-1.94	-0.26	52.55	-1.37	-0.64	77.96	23.16	23.03	45.86	45.61	y	-1.50	-2.18
53	-4.50	-0.50	15	-4.00	-2.00	175	x	x	x	x	x	x	x	x	x	x	y	-4.50	-5.21
54	-0.50	-	-	-0.50	-	-	x	x	x	x	x	x	x	x	x	x	y	-0.50	-1.31
55	-2.25	-2.00	115	-2.75	-1.25	50	x	x	x	x	x	x	x	x	x	x	y	-2.25	-2.95
56	-1.25	-	-	-1.00	-	-	x	x	x	x	x	x	x	x	x	x	y	-1.25	-1.87

57	-3.00	-	-	-0.75	-	-	x	x	x	x	x	x	x	x	x	x	y	-3.00	-3.81
58	2.50	-1.75	130	1.50	-	-	1.58	-2.27	126.86	2.19	-1.61	172.26	22.68	22.69	44.76	44.41	y	2.50	5.33
60	-1.75	-0.50	90	-1.75	-	-	-0.91	-0.78	109.28	-0.94	-0.90	48.59	24.35	24.41	42.48	42.67	y	-1.75	-2.46
61	x	-	-	x	-	-	x	x	x	x	x	x	x	x	x	x	y	x	x
62	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.22
63	-1.00	-	-	-1.50	-	-	-1.33	-0.70	93.54	-1.58	-0.58	96.78	24.34	x	42.24	42.33	y	-1.00	-1.64
64	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.19
65	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.17
69	-0.50	-0.50	160	-0.50	-0.75	45	-0.42	-0.82	127.86	-0.41	-1.62	59.49	23.94	24.17	43.58	43.39	y	-0.50	-1.23
70	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	-0.15
71	x	-	-	x	-	-	0.33	-0.50	117.16	0.12	-0.79	93.41	23.50	x	42.94	43.14	y	0.31	3.68
72	0.00	-	-	0.00	-	-	-0.33	-1.60	112.83	0.12	-1.58	78.21	24.42	24.34	42.51	42.20	y	-0.29	-1.17
73	7.00	-	-	6.50	-	-	4.66	-0.85	58.61	3.93	-1.53	145.03	x	20.49	x	x	y	7.00	6.23
74	-0.50	-	-	-0.50	-	-	x	x	x	x	x	x	x	x	x	x	y	-0.50	-1.19
91	0.50	-0.50	135	0.00	-0.50	90	0.04	-3.83	22.89	0.31	-1.04	146.95	x	x	x	x	y	0.50	3.73
102	x	-	-	x	-	-	0.25	-1.12	12.64	-0.04	-0.64	15.01	24.31	24.40	41.99	41.70	y	0.24	3.06
103	x	-	-	x	-	-	0.16	-0.79	9.24	-0.22	-0.11	56.60	22.60	22.52	43.72	44.15	n	0.16	3.03
104	x	-	-	x	-	-	-0.04	-0.08	76.41	0.29	-0.17	135.00	24.07	24.06	41.34	41.14	y	-0.03	-0.64
105	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	-0.03
106	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	-0.02
108	x	-	-	x	-	-	0.04	-1.00	134.43	-0.29	-0.53	7.47	23.76	x	43.48	43.45	y	0..05	2.96
112	-1.00	-2.25	12	-0.50	-3.50	164	-1.50	-2.46	21.49	-1.04	-3.28	157.12	24.62	24.44	44.25	44.68	n	-1.00	-1.62
113	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.18
114	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.19
115	-0.75	-0.75	180	-0.25	-1.25	5	x	x	x	x	x	x	x	x	x	x	y	-0.75	-1.48
118	x	-	-	x	-	-	0.70	-0.33	52.27	1.03	-0.64	76.44	23.71	23.51	42.35	42.56	y	0.65	4.03
119	1.50	-0.50	10	1.25	-0.25	180	1.20	-0.48	179.88	1.12	-0.87	176.70	x	x	x	x	y	1.50	4.60
125	1.25	-	-	1.25	-0.25	115	1.00	-0.87	161.46	0.66	-0.43	38.87	25.01	25.02	40.16	40.30	y	1.25	4.43
129	-3.25	-0.25	10	-3.25	-1.25	5	-3.12	-0.50	176.89	-3.21	-1.20	8.12	24.88	24.65	43.70	44.60	y	-3.25	-4.26
130	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.25
131	-4.00	-0.75	175	-4.50	-0.25	20	x	x	x	x	x	x	x	x	x	x	n	-4.00	-4.90
132	-3.00	-	-	-2.75	-	-	x	x	x	x	x	x	x	x	x	x	n	-3.00	-3.59
133	-4.25	-1.25	6	-5.50	-1.75	174	x	x	x	x	x	x	x	x	x	x	y	-4.25	-5.03
134	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.32
140	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	0.40
142	-3.25	-	-	-2.75	-0.75	75	-3.08	-0.50	118.63	-2.87	-0.58	82.60	x	26.03	42.62	42.81	y	-3.25	-3.87

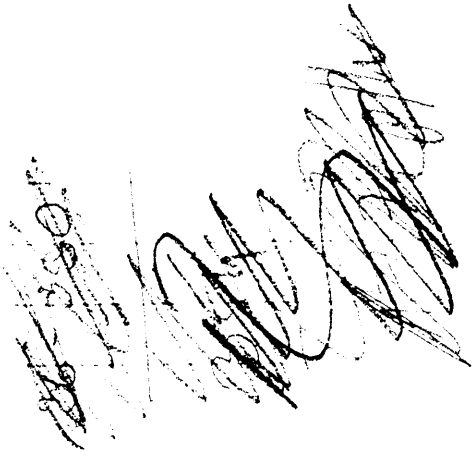
143	0.25	-0.25	180	0.25	-0.50	10	x	x	x	x	x	x	x	x	x	x	y	0.25	3.25
144	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	0.47
145	-1.00	-	-	-1.00	-	-	x	x	x	x	x	x	x	x	x	x	y	-1.00	-1.56
150	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.48
156	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.57
163	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	0.99
164	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.05
165	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.07
166	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.12
167	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.19
168	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.23
170	0.00	-	-	-0.50	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.25
173	-5.75	-0.25	13	-5.25	-0.50	55	-6.58	-0.45	169.83	-6.09	-1.05	6.44	24.82	24.77	45.49	45.52	y	-5.75	-5.70
174	-7.00	-	-	-7.00	-0.50	180	-7.00	-0.87	157.55	-7.65	-0.74	178.37	26.10	x	43.30	43.81	y	-7.00	-7.14
175	-1.75	-	-	-2.00	-	-	-2.62	-0.91	17.23	-2.69	-1.78	170.45	24.14	24.44	43.70	43.91	y	-1.75	-2.43
176	-1.75	-	-	-2.00	-	-	-3.62	-0.62	177.95	-2.87	-1.12	162.25	24.10	x	44.15	44.09	y	-1.75	-2.29
177	-2.75	-0.50	95	-0.75	-0.25	85	x	x	x	x	x	x	x	x	x	x	y	-2.75	-3.20
178	-2.00	-3.00	175	-2.25	-2.75	15	-2.91	-2.91	174.99	-3.54	-2.61	18.83	24.41	24.77	44.77	44.45	y	-2.00	-2.71
179	-3.00	-0.25	90	-2.75	-0.25	90	-3.54	-0.33	146.03	-3.33	-0.29	145.23	23.80	23.60	44.65	44.47	y	-3.00	-3.47
180	-2.00	-0.25	90	-2.50	-0.50	100	-2.33	-0.79	87.23	-2.25	-0.87	89.67	25.05	x	x	42.09	y	-2.00	-2.66
181	0.00	-	-	0.00	-	-	0.54	-0.70	71.09	-1.10	-0.32	95.14	24.29	24.87	40.91	41.14	y	0.50	3.83
182	0.00	-	-	0.00	-	-	0.33	-0.53	73.52	0.08	-0.47	111.81	24.73	24.90	39.27	39.59	y	0.31	3.71
187	0.25	-0.25	58	0.00	-0.25	92	x	x	x	x	x	x	x	x	x	x	n	0.25	3.44
189	-2.50	-1.50	150	-2.25	-0.25	110	x	x	x	x	x	x	x	x	x	x	n	-2.50	-3.05
190	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.28
193	-4.25	-1.00	60	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	-4.25	-4.92
194	x	-	-	0.00	-0.25	15	3.16	-1.06	166.72	0.47	-0.23	37.92	23.51	24.13	40.24	41.42	y	2.88	5.34
195	0.25	-0.50	180	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.25	3.44
196	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.31
197	0.00	-0.25	35	0.50	-0.25	90	x	x	x	x	x	x	x	x	x	x	y	0.00	1.51
198	-0.25	-1.25	97	-0.25	-0.50	84	0.50	-1.21	88.39	1.12	-2.08	93.40	23.76	23.72		42.54	y	-0.25	-1.13
199	-2.50	-2.50	168	1.75	-3.75	10	-3.46	-3.94	157.08	2.41	-4.16	20.67	x	22.78	45.33	45.23	y	-2.50	-3.05
201	x	-	-	x	-	-	2.72	-2.65	110.28	1.87	-1.00	87.81	x	23.63	44.42	44.51	y	2.48	5.24
202	x	-	-	x	-	-	0.71	-0.54	33.33	0.54	-0.16	98.00	x	x	x	x	n	0.65	4.18
203	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	1.63
204	0.00	-	-	0.00	-	-	4.22	-2.16	155.85	-12.25	-1.75	111.00	23.71	23.80	42.63	42.77	y	3.84	5.91

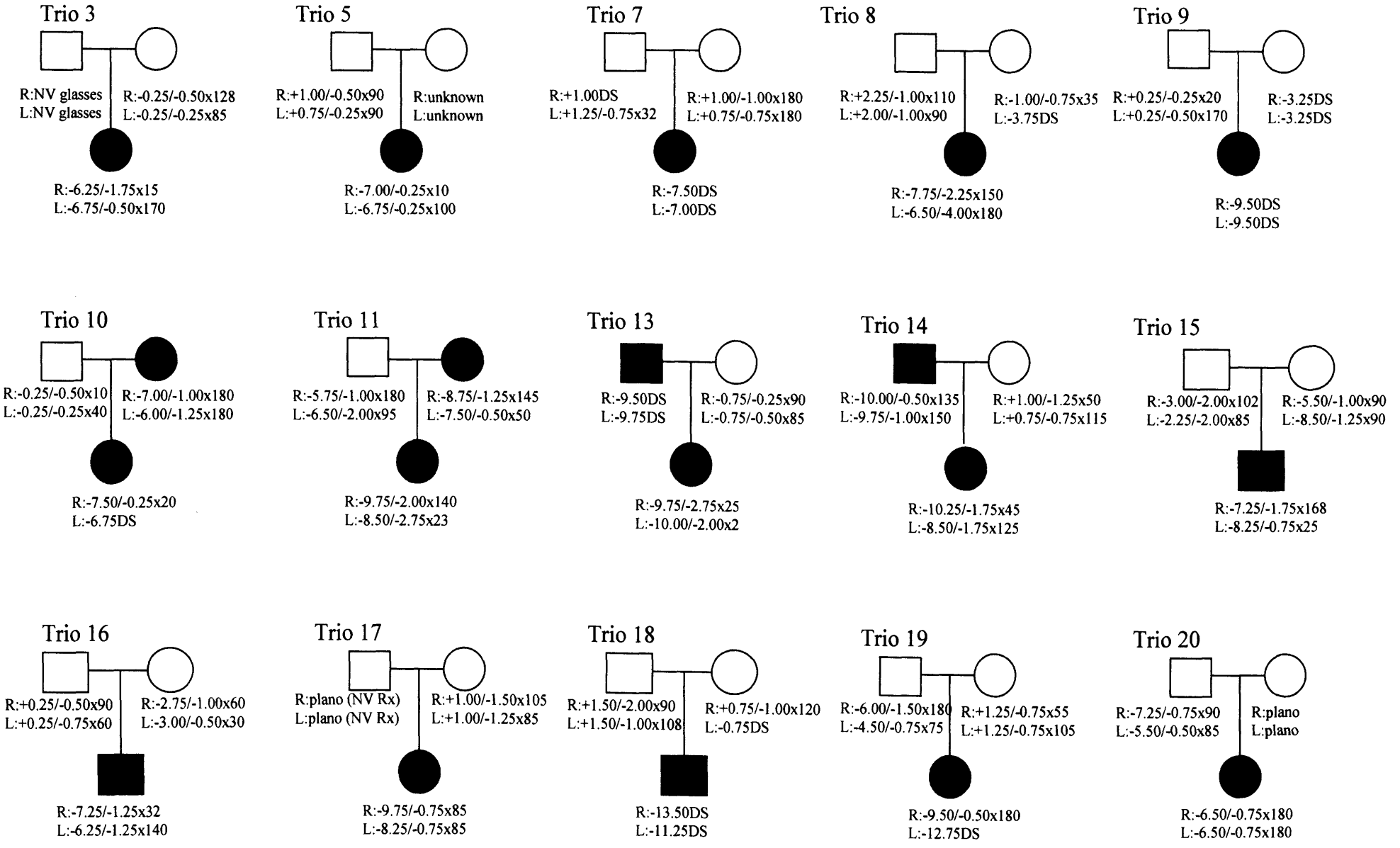
205	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.73
206	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.86
207	0	-	-	0	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.91
208	8.50	-1.50	20	9.00	-1.25	170	x	x	x	x	x	x	x	x	x	x	n	8.50	10.13
209	0	-	-	0	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	1.94
210	-4.50	-	-	-4.50	-	-	x	x	x	x	x	x	x	x	x	x	n	-4.50	-5.17
215	-3.00	-	-	-3.00	-	-	x	x	x	x	x	x	x	x	x	x	n	-3.00	-3.43
216	-1.00	-1.25	180	-1.25	-1.25	175	-1.12	-1.28	178.29	-1.25	-1.96	170.94	x	x	43.50	43.59	y	-1.00	-1.49
217	-1.00	-	-	-1.00	-	-	-1.19	-0.92	96.17	-1.37	-0.65	77.87	23.75	x	43.95	44.41	y	-1.00	-1.48
218	0	-	-	0	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	2.04
219	-3.00	-0.50	53	-2.00	-1.25	115	x	x	x	x	x	x	x	x	x	x	n	-3.00	-3.43
220	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	2.15
221	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	2.18
228	1.00	-0.50	90	0.75	-0.75	90	x	x	x	x	x	x	x	x	x	x	y	1.00	4.27
240	0.50	-1.00	180	0.25	-1.00	15	x	x	x	x	x	x	x	x	x	x	y	0.00	2.24
247	-10.00	-3.00	15	-11.50	-1.25	180	x	x	x	x	x	x	x	x	x	x	n	-10.00	-7.40
257	1.75	-	-	1.75	-	-	1.79	-0.33	97.48	2.00	-0.19	48.90	23.27	23.21	41.67	41.78	n	1.75	4.96
258	1.50	-1.00	10	2.00	-1.25	118	x	x	x	x	x	x	x	x	x	x	n	1.50	4.69
259	0.00	-	-	0.00	-	-	0.50	-0.79	104.94	0.33	-0.32	86.99	x	x	x	43.86	y	0.00	2.24
260	1.25	-1.25	105	1.00	-1.00	90	1.58	-1.33	59.00	1.33	-0.42	169.33	23.55	23.61	41.93	41.88	n	1.25	4.56
261	-4.00	-0.75	5	-3.50	-0.75	165	-4.25	-0.85	27.55	-3.46	-0.68	125.36	26.65	26.13	41.72	42.17	n	-4.00	-4.87
262	x	-	-	x	-	-	-1.29	-1.37	117.54	0.22	-2.24	72.51	24.00	23.73	42.94	42.81	y	-1.16	-1.85
263	0.00	-	-	0.00	-	-	0.07	-0.53	120.62	0.37	-0.09	86.28	23.63	23.67	42.01	42.01	y	0.06	3.00
264	0.00	-	-	0.00	-	-	-0.12	-0.17	136.36	0.21	-0.41	72.49	23.98	x	42.03	41.98	y	-0.10	-1.04
269	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.25
283	0.00	-	-	0.00	-	-	-0.08	-0.37	107.66	0.00	0.00	0.00	23.94	24.15	42.43	42.15	y	-0.06	0.70
285	0.00	-	-	0.00	-	-	-0.66	-0.25	28.48	-0.25	-0.15	130.94	23.11	22.87	43.70	44.01	y	-0.59	1.41
286	x	-	-	x	-	-	0.00	0.00	0.00	0.16	-0.03	45.00	23.43	23.30	42.75	42.86	y	0.01	2.88
287	0.00	-	-	0.00	-	-	0.58	-0.12	82.53	1.12	-0.08	95.39	23.40	23.35	40.72	41.29	y	0.54	3.87
288	0.00	-	-	0.00	-	-	0.62	-0.62	91.28	0.33	-0.28	145.98	22.80	22.75	43.33	43.53	y	0.57	3.89
289	-1.50	-0.75	155	-1.00	-0.25	170	-2.00	-0.45	126.14	-1.16	-0.46	126.89	x	24.81	42.92	42.67	y	-1.50	-2.16
290	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.32
291	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.44
292	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.46
293	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.49
294	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.49

295	2.00	-1.50	10	0.50	-0.50	160	2.37	-2.51	9.57	1.00	-1.09	161.94	22.38	x	42.97	42.92	y	2.16	5.08
296	x	-	-	x	-	-	-0.58	-0.81	156.80	-0.54	-1.12	159.78	23.73	x	44.28	44.31	y	-0.52	-1.32
297	x	-	-	x	-	-	1.25	-1.18	95.09	1.22	-0.84	99.30	x	x	x	x	y	1.14	4.33
298	x	-	-	x	-	-	4.78	-2.84	171.94	4.47	-3.00	4.24	x	x	x	x	y	4.35	5.99
302	0.00	-	-	0.00	-	-	0.21	-0.77	96.25	-0.13	-0.77	92.88	23.84	23.98	42.75	42.86	y	0.20	3.03
305	-3.75	-	-	-3.75	-0.25	135	-4.16	-0.52	142.23	-4.75	-0.90	19.79	x	x	42.49	42.30	y	-3.76	-4.71
312	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.65
313	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.66
314	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.72
315	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	y	0.00	2.72
316	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	2.75
317	0.00	-	-	0.00	-	-	x	x	x	x	x	x	x	x	x	x	n	0.00	2.84

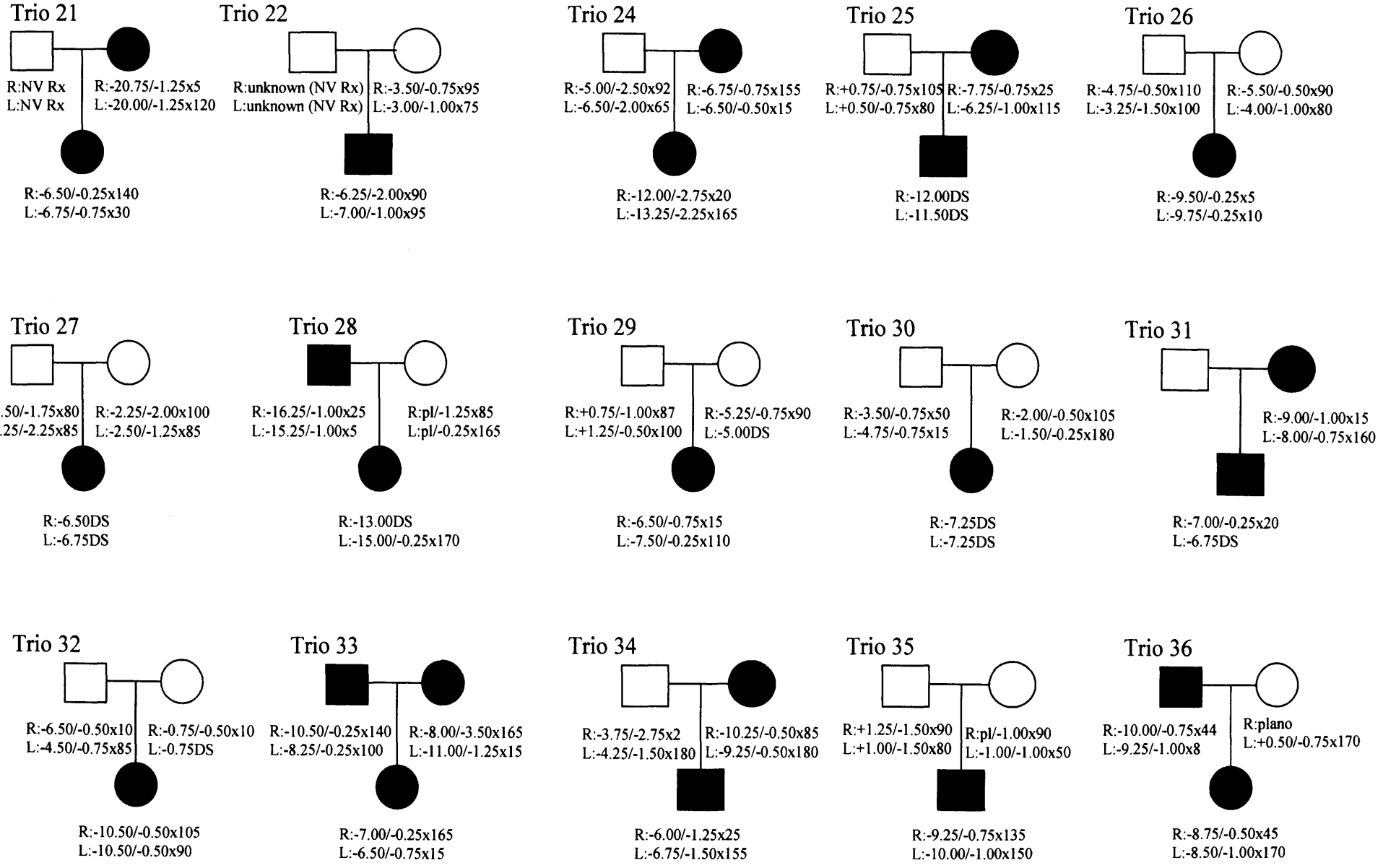
LARGE

FORMAT

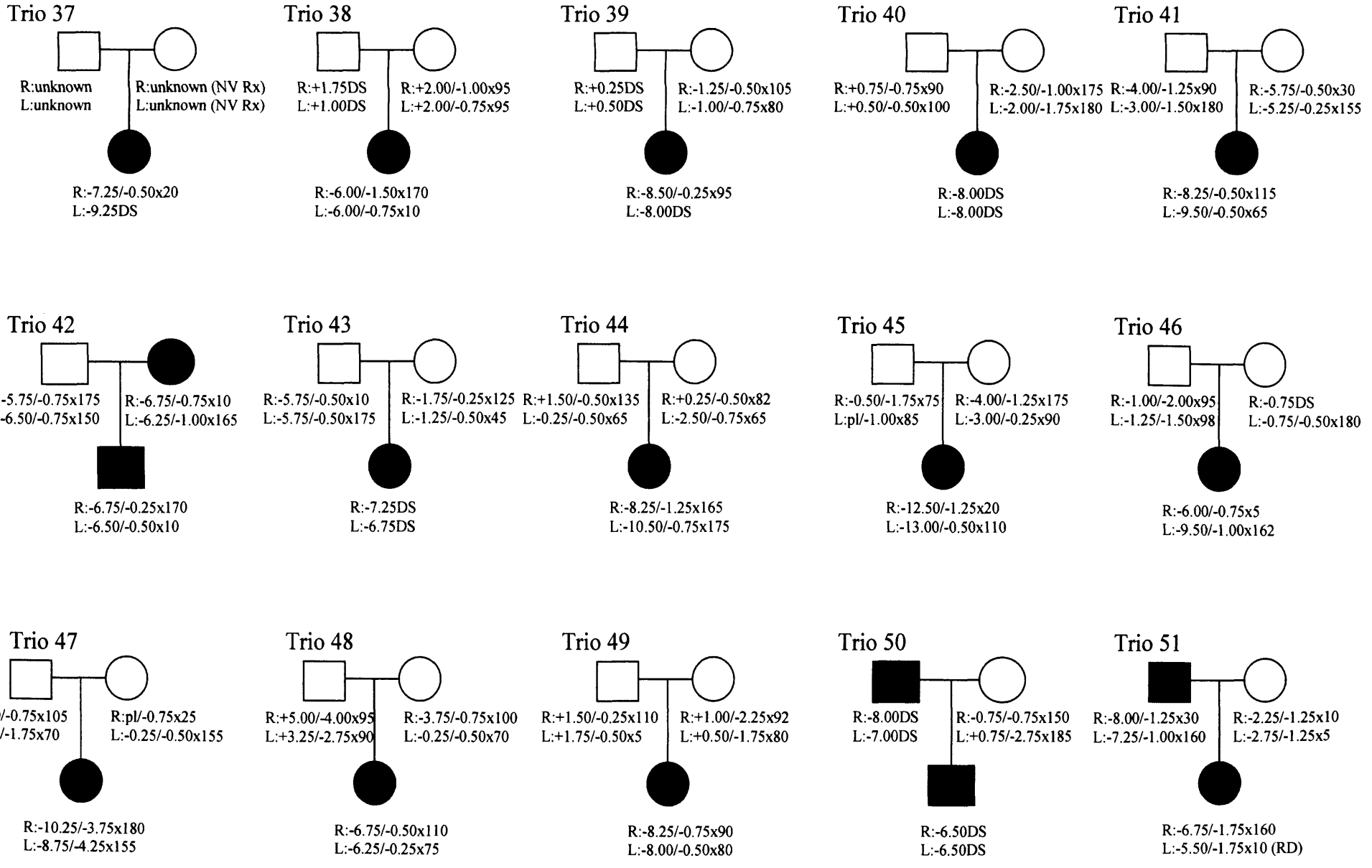




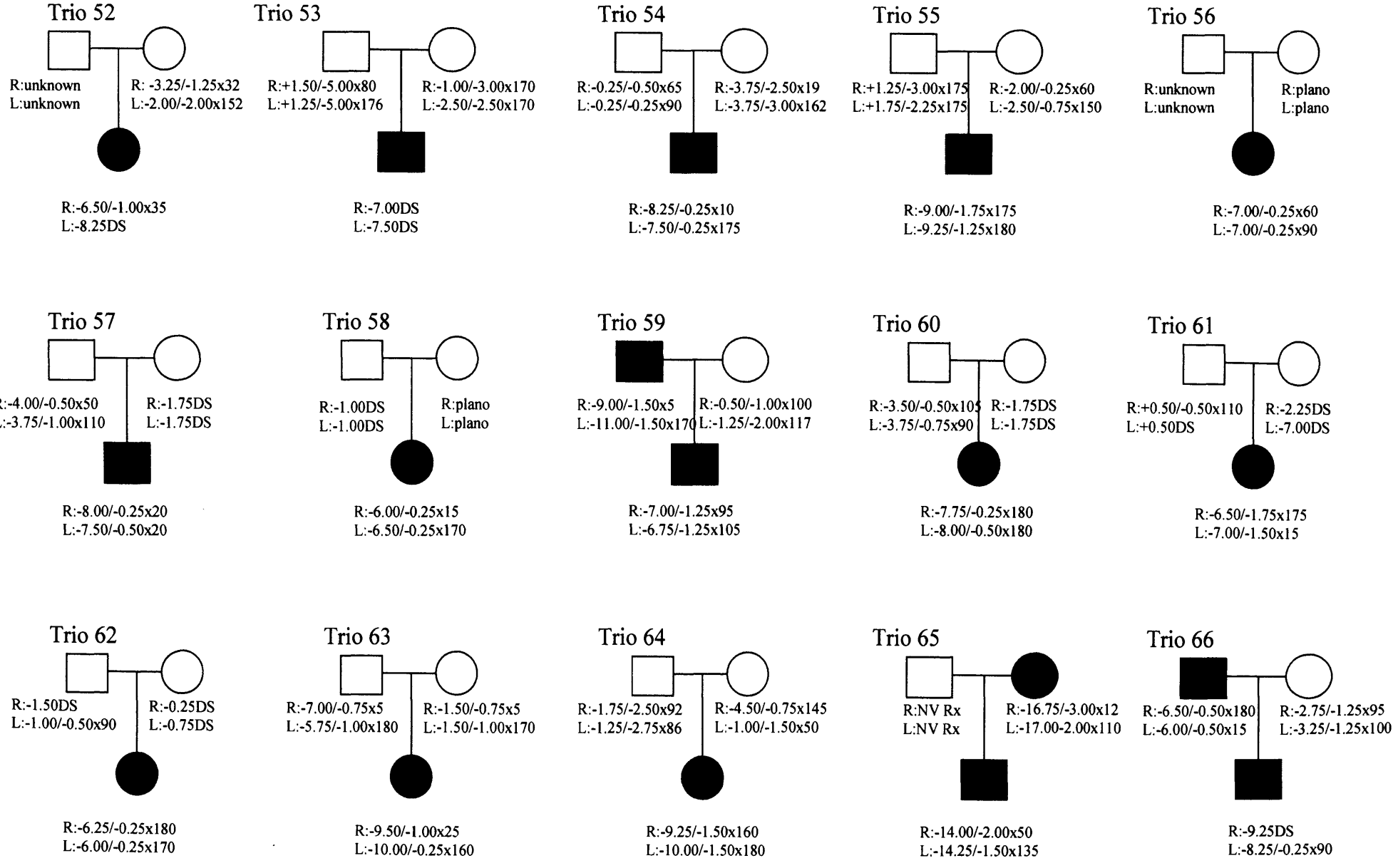
Trio Pedigrees. Trios recruited for association analysis.



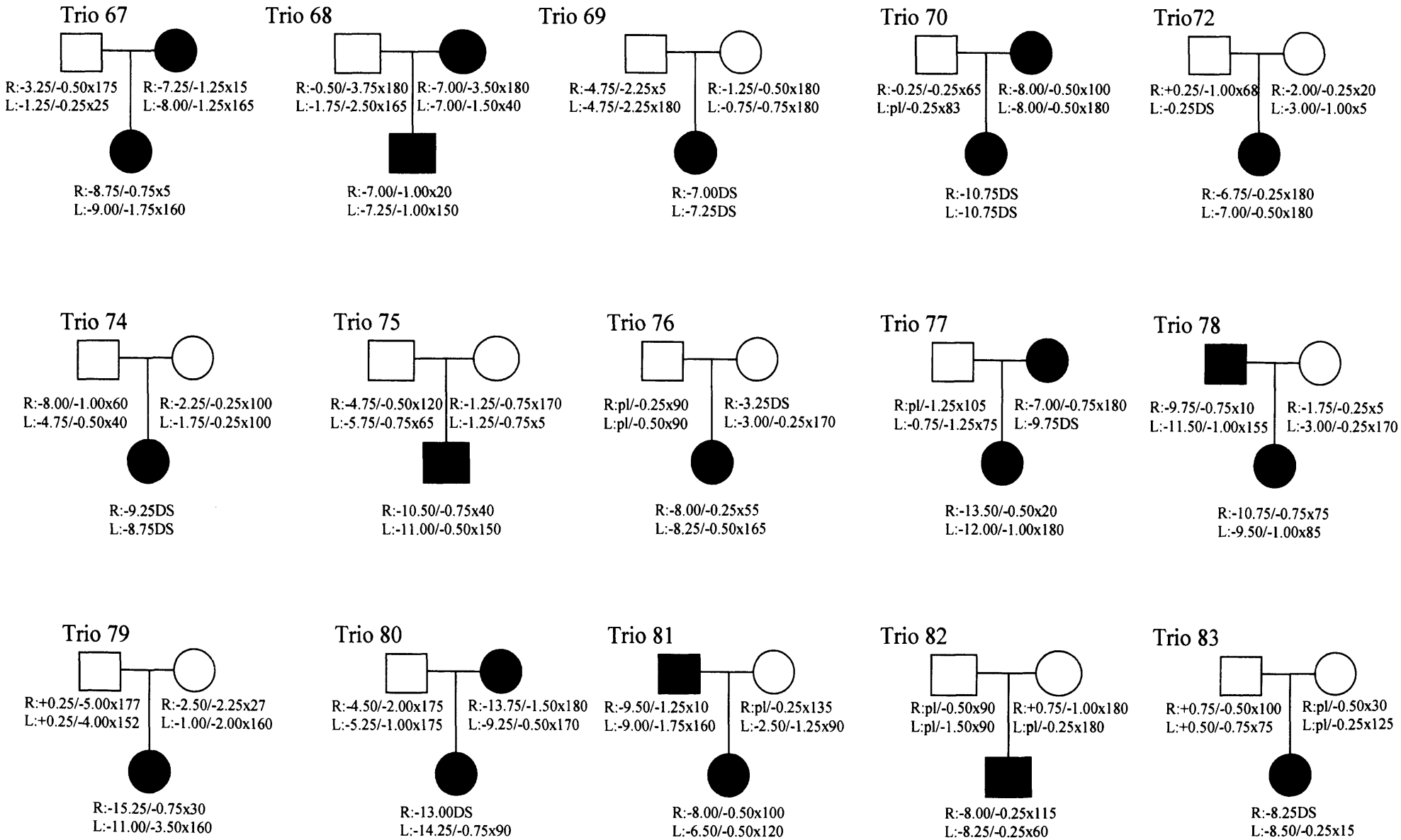
Trio pedigrees. contd.



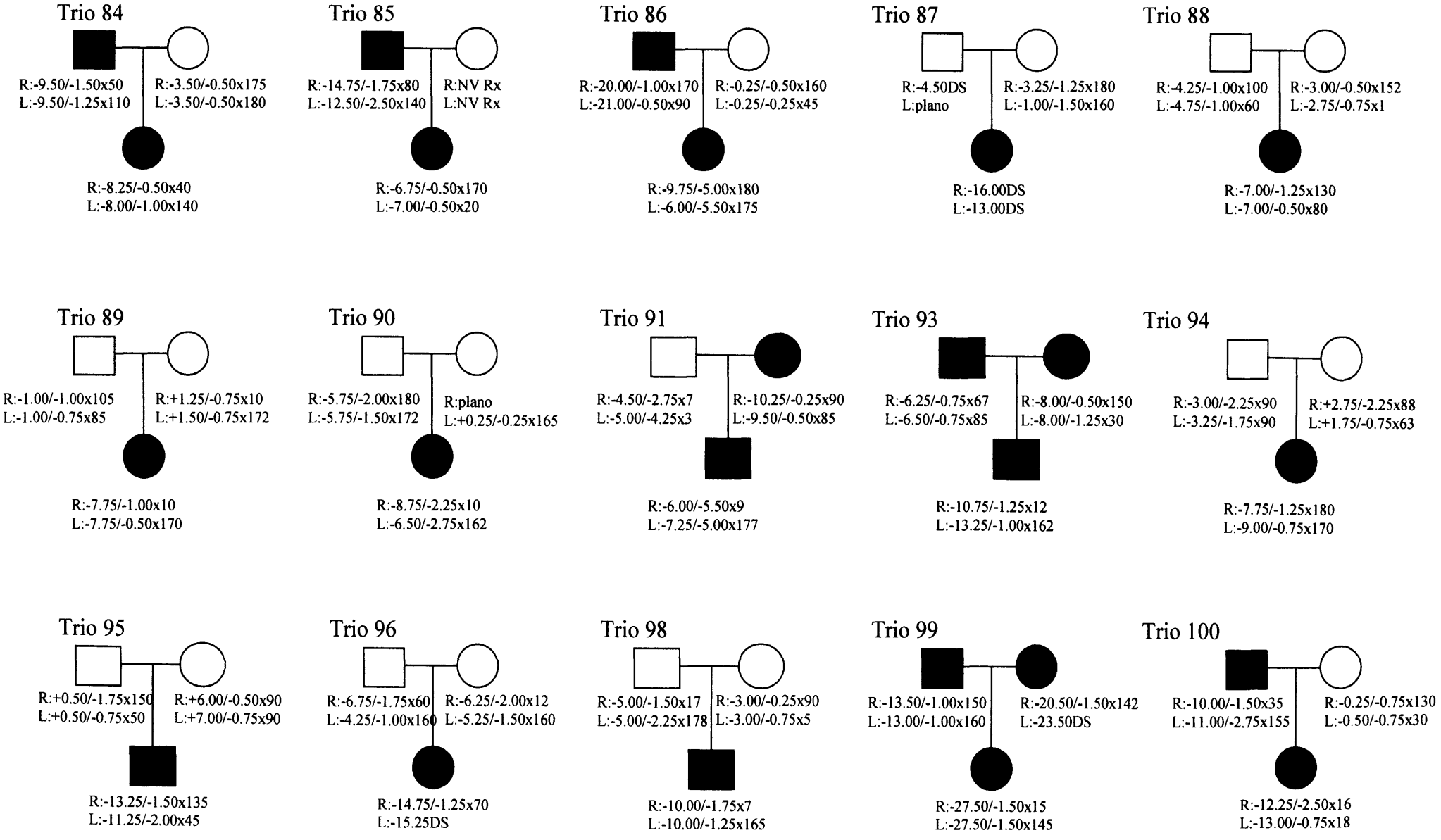
Trio pedigrees contd.



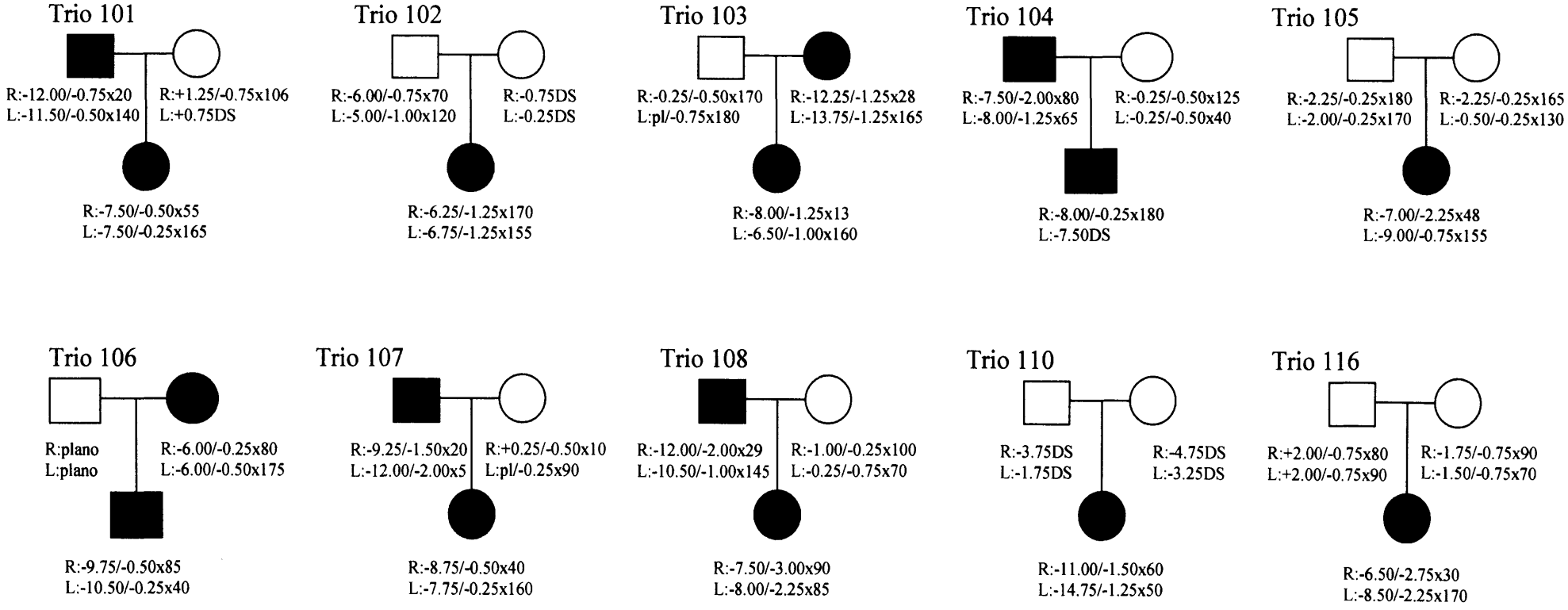
Trio pedigrees contd.



Trio pedigrees contd.



Trio pedigrees. contd.



Trio pedigrees. contd.

PUBLICATIONS

Akhtar, S., Bron, A.J., Qin, X., **Creer, R.C.**, Guggenheim, J.A., Meek, K.M. Gelatinous drop-like corneal dystrophy in a child with developmental delay: Clinico-pathological features and exclusion of the M1S1 gene. *Eye*; 2005 Feb; 19(2):198-204

Guggenheim, J.A., Qin, X., **Creer, R.C.** Postnatal refractive development in the Brown Norway rat: limitations of standard refractive and ocular component dimension measurement techniques. *Current Eye Research.* 2004 Oct-Nov;29(45):369-376.

Guggenheim, J.A., **Creer, R.C.** Myopia genetics update: Unlocking the secrets of myopia. *Optometry Today.* 2004 Sept 10th; 28-30.

PUBLISHED ABSTRACTS

Creer, R.C., Murphy, A., Pong-Wong, R., Haley, C.S., Guggenheim, J.A. High heritability of refractive error in a single multi-generational pedigree determined using three different methods. *Investigative Ophthalmology and Visual Science;* 2004; 45; E-Abstract 2724.

Creer, R.C., Murphy, A., Pong-Wong, R., Haley, C.S., Guggenheim, J.A. Heritability of refractive error within a multi-generational pedigree. *Ophthalmic and Physiological Optics;* 2004;24:151

Guggenheim, J.A., **Creer, R.C.**, Zayats, T., Williams, H., Kirov, G., Owen, M.J. A family-based test of association between alleles at the 5' UTR polymorphism of the myocilin (MYOC) gene and high myopia, in subjects from the U.K. Submitted March 2006 for XIth International Myopia Conference, Singapore.

References

Almasy L, and Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics* 62: 1198-1211.

Angi M R, Clementi M, Sardei C, Piattelli E, and Bisantis C (1993) Heritability of myopic refractive errors in identical and fraternal twins. *Graefes Archive for Clinical and Experimental Ophthalmology* 231: 580-585.

Badzioch M D, Thomas D C, and Jarvik G P (2003) Summary report: Missing data and pedigree and genotyping errors. *Genetic Epidemiology* 25 Suppl 1: S36-42.

Barker D L, Hansen M S, Faruqi A F, Giannola D, Irsula O R, Lasken R S, Latterich M et al. (2004) Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Research* 14: 901-907.

Barnes M, Williams C, Lumb R, Harrad R A, Sparrow J M, and Harvey I (2001) The prevalence of refractive errors in a UK birth cohort of children aged 7 years. *Investigative Ophthalmology & Visual Science* 42.

Baur M P, and Knapp M (1997) Association Studies in Genetic Epidemiology. In: Ivar-Harry Pawlowitzki, J H E, Elizabeth A. Thompson. [ed.] *Genetic mapping of disease genes*. San Diego: Academic Press, pp. 159-171.

Bear J C (1991) Epidemiology and Genetics of Refractive Anomalies. In: Grosvenor, T, and Flom, M C [eds.] *Refractive Anomalies: research and clinical applications*. Boston ; London : Butterworth-Heinemann, pp. 57-80.

Biino G, Palmas M A, Corona C, Prodi D, Fanciulli M, Sulis R, Serra A et al. (2005) Ocular refraction: heritability and genome-wide search for eye morphometry traits in an isolated Sardinian population. *Human Genetics* 116: 152-159.

Botstein D, and Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* 33 Suppl: 228-237.

Brown B, Edwards M H, and Leung J T (2002) Is esophoria a factor in slowing of myopia by progressive lenses? *Optometry & Vision Science* 79: 638-642.

Brownstein M J, Carpten J D, and Smith J R (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20: 1004-1006, 1008-1010.

Camp N J, and Cox A [eds.] (2002) *Quantitative Trait Loci: Methods and Protocols.* (1 edn.) Totowa, New Jersey: Humana Press.

Cardon L R, and Bell J I (2001) Association study designs for complex diseases. *Nature Review Genetics* 2: 91-99.

Cardon L R, and Palmer L J (2003) Population stratification and spurious allelic association. *Lancet* 361: 598-604.

Chong Y S, Liang Y, Tan D, Gazzard G, Stone R A, and Saw S M (2005) Association between breastfeeding and likelihood of myopia in children. *Journal of the American Medical Association* 293: 3001-3002.

Choo V (2003) A look at slowing progression of myopia. *The Lancet* 361: 1622-1623.

Chung K, Mohidin N, and O'Leary D J (2002) Undercorrection of myopia enhances rather than inhibits myopia progression. *Vision Research* 42: 2555-2559.

Curtin B J (1985) The Etiology of Myopia.*The Myopias: basic science and clinical management.* Philadelphia: Harper & Row Ltd, pp. 61-151.

Curtis D, and Sham P C (1994) Using risk calculation to implement an extended relative pair analysis. *Annals of Human Genetics* 58 (Pt 2): 151-162.

Davies J L, Kawaguchi Y, Bennett S T, Copeman J B, Cordell H J, Pritchard L E, Reed P W et al. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371: 130-136.

Dean F B, Hosono S, Fang L, Wu X, Faruqi A F, Bray-Ward P, Sun Z et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Science U S A* 99: 5261-5266.

Dean F B, Nelson J R, Giesler T L, and Lasken R S (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* 11: 1095-1099.

Dickson P A, Montgomery G W, Henders A, Campbell M J, Martin N G, and James M R (2005) Evaluation of multiple displacement amplification in a 5 cM STR genome-wide scan. *Nucleic Acids Research* 33: e119.

Edmunds R T (1949) Vision of albinos. *Archives of Ophthalmology* 42: 755-767.

Edwards M H (1996) Do variations in normal nutrition play a role in the development of myopia? *Optometry & Vision Science* 73: 638-643.

Falconer D S (1996) *Introduction to Quantitative Genetics*. Third ed. New York: Longman Scientific & Technical.

Farbrother J E (2003) *Genetic Epidemiology and Linkage Analysis of High Myopia.* Cardiff University.

Farbrother J E, Kirov G, Owen M J, and Guggenheim J A (2004a) Family aggregation of high myopia: estimation of the sibling recurrence risk ratio. *Investigative Ophthalmology & Visual Science* 45: 2873-2878.

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

Feigelson H S, Rodriguez C, Robertson A S, Jacobs E J, Calle E E, Reid Y A, and Thun M J (2001) Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiology, Biomarkers & Prevention* 10: 1005-1008.

Fingert J H, Stone E M, Sheffield V C, and Alward W L (2002) Myocilin glaucoma. *Survey of Ophthalmology* 47: 547-561.

Fingert J H, Ying L, Swiderski R E, Nystuen A M, Arbour N C, Alward W L, Sheffield V C et al. (1998) Characterization and comparison of the human and mouse GLC1A glaucoma genes. *Genome Research* 8: 377-384.

Fledelius H C (1996) Pre-term delivery and subsequent ocular development. A 7-10 year follow-up of children screened 1982-84 for ROP. 3) Refraction. Myopia of prematurity. *Acta Ophthalmologica Scandinavica* 74: 297-300.

Fox C S, Cupples L A, Chazaro I, Polak J F, Wolf P A, D'Agostino R B, Ordovas J M et al. (2004) Genomewide linkage analysis for internal carotid artery intimal medial

thickness: evidence for linkage to chromosome 12. *American Journal of Human Genetics* 74: 253-261.

Fulk G W, Cyert L A, and Parker D E (2000) A randomized trial of the effect of single-vision vs. bifocal lenses on myopia progression in children with esophoria. *Optometry & Vision Science* 77: 395-401.

Fulk G W, Cyert L A, Parker D E, and West R W (2003) The effect of changing from glasses to soft contact lenses on myopia progression in adolescents. *Ophthalmic and Physiological Optics* 23: 71-77.

Garcia-Closas M, Egan K M, Abruzzo J, Newcomb P A, Titus-Ernstoff L, Franklin T, Bender P K et al. (2001) Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiology, Biomarkers & Prevention* 10: 687-696.

Garoufalis P, Chen C Y, Dirani M, Couper T A, Taylor H R, and Baird P N (2005) Methodology and Recruitment of Probands and Their Families for the Genes in Myopia (GEM) Study. *Ophthalmic Epidemiology* 12: 383-392.

Gilmartin B (2004) Myopia: precedents for research in the twenty-first century. *Clinical Experimental Ophthalmology* 32: 305-324.

Goldschmidt E (1968) On the Etiology of Myopia. An epidemiological study. *Acta Ophthalmologica* Supplementum 98: 1-171.

Gordon D, Haynes C, Finch S J, and Brown A M (2006) Increase in linkage information by stratification of pedigree data into gold-standard and standard diagnoses: application to the NIMH Alzheimer Disease Genetics Initiative Dataset. *Human Heredity* 61: 97-103.

Goss D A, Hampton M J, and Wickham M G (1988) Selected review on genetic factors in myopia. *Journal of the American Optometric Association* 59: 875-884.

Grosvenor T (1987) A review and a suggested classification system for myopia on the basis of age-related prevalence and age of onset. *American Journal of Optometry & Physiological Optics* 64: 545-554.

Grosvenor T, and Goss D A (1999) *Clinical Management of Myopia*. Butterworth-Heinemann.

Gwiazda J, Hyman L, Hussein M, Everett D, Norton T T, Kurtz D, Leske M C et al. (2003) A Randomized Clinical Trial of Progressive Addition Lenses versus Single Vision Lenses on the Progression of Myopia in Children. *Investigative Ophthalmology and Visual Science* 44: 1492-1500.

Gwiazda J, Ong E, Held R, and Thorn F (2000) Myopia and ambient night-time lighting. *Nature* 404: 144.

Gwiazda J, Thorn F, Bauer J, and Held R (1993) Myopic children show insufficient accommodative response to blur. *Investigative Ophthalmology & Visual Science* 34: 690-694.

Haines J L, and Pericak-Vance M A [eds.] (1998) *Approaches to Gene Mapping in Complex Human Diseases.* New York: Wiley-Liss.

Hammond C J, Andrew T, Tat Mak Y, and Spector T D (2004) A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genomewide scan of dizygotic twins. *American Journal of Human Genetics* 75: 294-304. Epub 2004 Jun 2024.

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

Han W, Yap M K, Wang J, and Yip S P (2006) Family-based association analysis of hepatocyte growth factor (HGF) gene polymorphisms in high myopia. *Investigative Ophthalmology & Visual Science* 47: 2291-2299.

Harty L C, Garcia-Closas M, Rothman N, Reid Y A, Tucker M A, and Hartge P (2000) Collection of buccal cell DNA using treated cards. *Cancer Epidemiology, Biomarkers & Prevention* 9: 501-506.

Hauser E R, Watanabe R M, Duren W L, Bass M P, Langefeld C D, and Boehnke M (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genetic Epidemiology* 27: 53-63.

Heath E M, Morken N W, Campbell K A, Tkach D, Boyd E A, and Strom D A (2001) Use of buccal cells collected in mouthwash as a source of DNA for clinical testing. *Archives of Pathology and Laboratory Medicine* 125: 127-133.

Hoffman J I, and Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology* 14: 599-612.

Holbrook J F, Stabley D, and Sol-Church K (2005) Exploring whole genome amplification as a DNA recovery tool for molecular genetic studies. *Journal of Biomolecular Techniques* 16: 125-133.

Hosono S, Faruqi A F, Dean F B, Du Y, Sun Z, Wu X, Du J et al. (2003) Unbiased whole-genome amplification directly from clinical samples. *Genome Research* 13: 954-964.

Ibay G, Doan B, Reider L, Dana D, Schlifka M, Hu H, Holmes T et al. (2004) Candidate high myopia loci on chromosomes 18p and 12q do not play a major role in susceptibility to common myopia. *BioMed Central Medical Genetics* 5: 20.

Irving E L, Sivak J G, and Callender M G (1992) Refractive plasticity of the developing chick eye. *Ophthalmic Physiol Opt* 12: 448-456.

Katz J, Schein O D, Levy B, Cruiscullo T, Saw S M, Rajan U, Chan T K et al. (2003) A randomized trial of rigid gas permeable contact lenses to reduce progression of children's myopia. *American Journal of Ophthalmology* 136: 82-90.

Laird N M, Horvath S, and Xu X (2000) Implementing a unified approach to family-based tests of association. *Genetic Epidemiology* 19 Suppl 1: S36-42.

Lam D S, Lee W S, Leung Y F, Tam P O, Fan D S, Fan B J, and Pang C P (2003) TGFbeta-Induced Factor: A Candidate Gene for High Myopia. *Investigative Ophthalmology & Visual Science* 44: 1012-1015.

Larsson E K, Rydberg A C, and Holmstrom G E (2003) A population-based study of the refractive outcome in 10-year-old preterm and full-term children. *Archives of Ophthalmology* 121: 1430-1436.

Lench N, Stanier P, and Williamson R (1988) Simple non-invasive method to obtain DNA for gene analysis. *Lancet* 1: 1356-1358.

Leung J T, and Brown B (1999) Progression of myopia in Hong Kong Chinese schoolchildren is slowed by wearing progressive lenses. *Optometry & Vision Science* 76: 346-354.

Leung Y F, Tam P O, Baum L, Lam D S, and Pang C C (2000) TIGR/MYOC proximal promoter GT-repeat polymorphism is not associated with myopia. *Human Mutation* 16: 533.

Lim R, Mitchell P, and Cumming R G (1999) Refractive associations with cataract: the Blue Mountains Eye Study. *Investigative Ophthalmology & Visual Science* 40: 3021-3026.

Lithander J (1999) Prevalence of myopia in school children in the Sultanate of Oman: a nation-wide study of 6292 randomly selected children. *Acta Ophthalmologica Scandinavica* 77: 306-309.

Logan N S, Gilmartin B, Marr J E, Stevenson M R, and Ainsworth J R (2004) Community-based study of the association of high myopia in children with ocular and systemic disease. *Optometry & Vision Science* 81: 11-13.

Lyhne N, Sjolie A K, Kyvik K O, and Green A (2001) The importance of genes and environment for ocular refraction and its determiners: a population based study among 20-45 year old twins. *British Journal of Ophthalmology* 85: 1470-1476.

Majumder P P, and Ghosh S (2005) Mapping quantitative trait loci in humans: achievements and limitations. *Journal of Clinical Investigation* 115: 1419-1424.

Marr J E, Halliwell-Ewen J, Fisher B, Soler L, and Ainsworth J R (2001) Associations of high myopia in childhood. *Eye* 15: 70-74.

Martin S, Richards A J, Yates J R, Scott J D, Pope M, and Snead M P (1999) Stickler syndrome: further mutations in COL11A1 and evidence for additional locus heterogeneity. *European Journal of Human Genetics* 7: 807-814.

Mitchell P, Hourihan F, Sandbach J, and Wang J J (1999) The relationship between glaucoma and myopia: the Blue Mountains Eye Study. *Ophthalmology* 106: 2010-2015.

Naiglin L, Gazagne C, Dallongeville F, Thalamas C, Idder A, Rascol O, Malecaze F et al. (2002) A genome wide scan for familial high myopia suggests a novel locus on chromosome 7q36. *Journal of Medical Genetics* 39: 118-124.

Nakatsuka C, Hasebe S, Nonaka F, and Ohtsuki H (2005) Accommodative lag under habitual seeing conditions: comparison between myopic and emmetropic children. *Japanese Journal of Ophthalmology* 49: 189-194.

O'Connell J R, and Weeks D E (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *American Journal of Human Genetics* 63: 259-266.

Ott J (1989) Computer-simulation methods in human linkage analysis. *Proceedings of the National Academy of Science of the United States of America* 86: 4175-4178.

Ott J (1991) *Analysis of human genetic linkage*. Baltimore: John Hopkins University.

Paez J G, Lin M, Beroukhim R, Lee J C, Zhao X, Richter D J, Gabriel S et al. (2004) Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Research* 32: e71.

Paluru P, Ronan S M, Heon E, Devoto M, Wildenberg S C, Scavello G, Holleschau A et al. (2003) New Locus for Autosomal Dominant High Myopia Maps to the Long Arm of Chromosome 17. *Investigative Ophthalmology and Visual Science* 44: 1830-1836.

Paluru P C, Nallasamy S, Devoto M, Rappaport E F, and Young T L (2005) Identification of a novel locus on 2q for autosomal dominant high-grade myopia. *Investigative Ophthalmology & Visual Science* 46: 2300-2307.

Paluru P C, Scavello G S, Ganter W R, and Young T L (2004) Exclusion of lumican and fibromodulin as candidate genes in MYP3 linked high grade myopia. *Molecular Vision* 10: 917-922.

Park T W, Winawer J, and Wallman J (2003) Further evidence that chick eyes use the sign of blur in spectacle lens compensation. *Vision Research* 43: 1519-1531.

Phillips J R (2005) Monovision slows juvenile myopia progression unilaterally. *British Journal of Ophthalmology* 89: 1196-1200.

Pompanon F, Bonin A, Bellemain E, and Taberlet P (2005) Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6: 847-859.

Ricard C S, and Tamm E R (2005) Focus on Molecules: Myocilin/TIGR. *Experimental Eye Research* 81: 501-502.

Richards A J, Yates J R, Williams R, Payne S J, Pope F M, Scott J D, and Snead M P (1996) A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in alpha 1 (XI) collagen. *Human Molecular Genetics* 5: 1339-1343.

Rose K A, Morgan I G, Smith W, and Mitchell P (2002) High heritability of myopia does not preclude rapid changes in prevalence. *Clinical and Experimental Ophthalmology* 30: 168-172.

Rosenfield M, and Gilmartin B (1998) *Myopia and Nearwork.* Butterworth-Heinemann.

Saw S M, Chia K S, Lindstrom J M, Tan D T, and Stone R A (2004a) Childhood myopia and parental smoking. *British Journal of Ophthalmology* 88: 934-937.

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

Saw S M, Shih-Yen E C, Koh A, and Tan D (2002) Interventions to retard myopia progression in children: an evidence-based update. *Ophthalmology* 109: 415-421.

Saw S M, Tan S B, Fung D, Chia K S, Koh D, Tan D T, and Stone R A (2004b) IQ and the Association with Myopia in Children. *Investigative Ophthalmology & Visual Science* 45: 2943-2948.

Scavello G S, Jr., Paluru P C, Zhou J, White P S, Rappaport E F, and Young T L (2005) Genomic structure and organization of the high grade Myopia-2 locus (MYP2) critical region: mutation screening of 9 positional candidate genes. *Molecular Vision* 11: 97-110.

Scavello G S, Paluru P C, Ganter W R, and Young T L (2004) Sequence Variants in the Transforming Growth beta-Induced Factor (TGIF) Gene Are Not Associated with High Myopia. *Investigative Ophthalmology & Visual Science* 45: 2091-2097.

Schmidt K, Kraft H G, Parson W, and Utermann G (2005) Genetics of the Lp(a)/apo(a) system in an autochthonous Black African population from the Gabon. *European Journal of Human Genetics.*

Schwartz M, Haim M, and Skarsholm D (1990) X-linked myopia: Bornholm eye disease. Linkage to DNA markers on the distal part of Xq. *Clinical Genetics* 38: 281-286.

Sham P (1998) *Statistics in Human Genetics*. London: Arnold Publishers.

Sham P C, and Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals of Human Genetics* 59 (Pt 3): 323-336.

Shapiro M B, and France T D (1985) The ocular features of Down's syndrome. *American Journal of Ophthalmology* 99: 659-663.

Shi M M (2001) Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clin Chem* 47: 164-172.

Shih Y F, Hsiao C K, Chen C J, Chang C W, Hung P T, and Lin L L (2001) An intervention trial on efficacy of atropine and multi-focal glasses in controlling myopic progression. *Acta Ophthalmologica Scandinavica* 79: 233-236.

Shimizu N, Nomura H, Ando F, Niino N, Miyake Y, and Shimokata H (2003) Refractive errors and factors associated with myopia in an adult Japanese population. *Jpn J Ophthalmol* 47: 6-12.

Snir M, Friling R, Weinberger D, Sherf I, and Axer-Siegel R (2004) Refraction and keratometry in 40 week old premature (corrected age) and term infants. *British Journal of Ophthalmology* 88: 900-904.

Sorensen D, and Gianola D (2002) *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. New York: Springer.

Sorsby A, Benjamin B, Davey J B, Sheridan M, and Tanner J M (1957) *Emmetropia and its abberations: A Study in the Correlation of the Optical Components of the Eye*. London: Privy Council: Medical Research Council Special Report Series No. 293.

Sorsby A, Leary G A, and Fraser G R (1966) Family studies on ocular refraction and its components. *Journal of Medical Genetics* 3: 269-273.

Sorsby A, Sheridan M, and Leary G A (1962) *Refraction and its Components in Twins*. London: Privy Council: Medical Research Council Special Report Series No.303.

Spielman R S, McGinnis R E, and Ewens W J (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 52: 506-516.

Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M, Holmes T et al. (2004) Genomewide Linkage Scan for Myopia Susceptibility Loci among Ashkenazi Jewish Families Shows Evidence of Linkage on Chromosome 22q12. *American Journal of Human Genetics* 75: 448-459. Epub 2004 Jul 2023.

Strachan T, and Read A P (1999) *Human Molecular Genetics 2*. Second ed. Oxford: BIOS Scientific Publishers Ltd.

Tabor H K, Risch N J, and Myers R M (2002) Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Review Genetics* 3: 391-397.

Tan D T, Lam D S, Chua W H, Shu-Ping D F, and Crockett R S (2005) One-year multicenter, double-masked, placebo-controlled, parallel safety and efficacy study of 2% pirenzepine ophthalmic gel in children with myopia. *Ophthalmology* 112: 84-91.

Tang W, Yip S, Yap M, Lo K, Ng P, Choi P, and Lee S (2004) Testing for association between MYOC and myopia susceptibility in Hong Kong Chinese population. In: 10th International Myopia Conference 2004.

Terwilliger J D, and Goring H H (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Human Biology* 72: 63-132.

Tunnacliffe A H (1993) *Introduction to Visual Optics*. London: Unwin Brothers Ltd., The Gresham Press.

Ulgen A, and Li W (2005) Comparing single-nucleotide polymorphism marker-based and microsatellite marker-based linkage analyses. *BioMed Central Genetics* 6 Suppl 1: S13.

Van Allen M I, Fung J, and Jurenka S B (1999) Health care concerns and guidelines for adults with Down syndrome. *American Journal of Medical Genetics* 89: 100-110.

Van Steen K, and Laird N (2005) Family Based Association Tests and the FBAT-toolkit. User's Manual.

Vogel F, and Motulsky A G (1996) *Human Genetics: Problems and Approaches.* Third ed. Heidelberg: Springer-Verlag.

Walline J J, Jones L A, Mutti D O, and Zadnik K (2004) A randomized trial of the effects of rigid contact lenses on myopia progression. *Archives of Ophthalmology* 122: 1760-1766.

Wang I J, Chiang T H, Shih Y F, Hsiao C K, Lu S C, Hou Y C, and Lin L L (2006) The association of single nucleotide polymorphisms in the 5'-regulatory region of the lumican gene with susceptibility to high myopia in Taiwan. *Molecular Vision* 12: 852-857.

Weale M E, Depondt C, Macdonald S J, Smith A, Lai P S, Shorvon S D, Wood N W et al. (2003) Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping. *American Journal of Human Genetics* 73: 551-565.

Weale R A (2003) Epidemiology of refractive errors and presbyopia. *Survey of Ophthalmology* 48: 515-543.

Weaver R F, and Hedrick P W (1997) *Genetics.* Third ed. Dubuque: Wm. C. Brown Publishers.

Weeks D E, Ott J, and Lathrop G M (1990) SLINK: a general simulation program for linkage analysis. *American Journal of Human Genetics* 47: A204 (Supplement).

Williams J T, and Blangero J (1999) Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. *Genetic Epidemiology* 16: 113-134.

Winawer J, and Wallman J (2002) Temporal constraints on lens compensation in chicks. *Vision Research* 42: 2651-2668.

Winawer J, Zhu X, Choi J, and Wallman J (2005) Ocular compensation for alternating myopic and hyperopic defocus. *Vision Research* 45: 1667-1677.

Wojciechowski R, Congdon N, Bowie H, Munoz B, Gilbert D, and West S K (2005) Heritability of refractive error and familial aggregation of myopia in an elderly american population. *Investigative Ophthalmology & Visual Science* 46: 1588-1592.

Wojciechowski R, Moy C, Ciner E, Ibay G, Reider L, Bailey-Wilson J E, and Stambolian D (2006) Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36. *Human Genetics*.

Wu H, Yu X, and Yap E (1999) Allelic association between trabebular meshwork-induced glucocorticoid response (TIGR) gene and sporadic severe myopia. *Investigative Ophthalmology and Visual Sciences*. p. S600. Abstract 3148

Xu L, Wang Y, Wang S, Wang Y, and Jonas J B (2006) High Myopia and Glaucoma Susceptibility The Beijing Eye Study. *Ophthalmology*.

Yalcin B, Willis-Owen S A, Fullerton J, Meesaq A, Deacon R M, Rawlins J N, Copley R R et al. (2004) Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nature Genetics* 36: 1197-1202.

Young F A, and Leary G A (1972) The inheritance of ocular components. *American Journal of Optometry and Archives of American Academy of Optometry* 49: 546-555.

Young F A, Leary G A, Baldwin W R, West D C, Box R A, Harris E, and Johnson C (1969) The transmission of refractive errors within eskimo families. *American Journal of Optometry and Archives of American Academy of Optometry* 46: 676-685.

Young T L, Atwood L D, Ronan S M, Dewan A T, Alvear A B, Peterson J, Holleschau A et al. (2001a) Further refinement of the MYP2 locus for autosomal dominant high myopia by linkage disequilibrium analysis. *Ophthalmic Genetics* 22: 69-75.

Young T L, Deeb S S, Ronan S M, Dewan A T, Alvear A B, Scavello G S, Paluru P C et al. (2004a) X-linked high myopia associated with cone dysfunction. *Archives of Ophthalmology* 122: 897-908.

Young T L, Paluru P, Heon E, Bebchuck K, Armstrong C, Ronan S M, Holleschau A et al. (2001b) A new locus for autosomal dominant high myopia maps to chromosome 17q21-23. *American Journal of Human Genetics* 69: 526.

Young T L, Paluru P, Nallasamy S, Cole W G, and Heon E (2004b) Exclusion of TBX2 and TBX4 as candidate genes in MYP5-linked high myopia. In *The American Society of Human Genetics*. Toronto.

Young T L, Ronan S M, Alvear A B, Wildenberg S C, Oetting W S, Atwood L D, Wilkin D J et al. (1998a) A second locus for familial high myopia maps to chromosome 12q. *American Journal of Human Genetics* 63: 1419-1424.

Young T L, Ronan S M, Drahozal L A, Wildenberg S C, Alvear A B, Oetting W S, Atwood L D et al. (1998b) Evidence that a locus for familial high myopia maps to chromosome 18p. *American Journal of Human Genetics* 63: 109-119.

Zadnik K, Jones L A, Irvin B C, Kleinstein R N, Manny R E, Shin J A, and Mutti D O (2000) Myopia and ambient night-time lighting. CLEERE Study Group. Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error. *Nature* 404: 143-144.

Zhang Q, Guo X, Xiao X, Jia X, Li S, and Hejtmancik J F (2005) A new locus for autosomal dominant high myopia maps to 4q22-q27 between D4S1578 and D4S1612. *Molecular Vision* 11: 554-560.

Zhou J, and Young T L (2005) Evaluation of Lipin 2 as a candidate gene for autosomal dominant 1 high-grade myopia. *Gene* 352: 10-19.

Zhu X, Winawer J A, and Wallman J (2003) Potency of myopic defocus in spectacle lens compensation. *Investigative Ophthalmology & Visual Science* 44: 2818-2827.

Website Bibliography

David Curtis: http://www.smd.qmul.ac.uk/statgen/dcurtis/lectures/etdtlect.html

Zeiss IOLmaster: http://www.doctor-hill.com/iol-master/iolmaster_main.htm

The Family Study of Myopia: http://www.cardiff.ac.uk/myopia