CLINICAL AND EXPERIMENTAL
OPTOMETRY

INVITED REVIEW

A review of current approaches to identifying human genes involved in myopia

Clin Exp Optom 2008; 91: 1: 4–22 DOI:10.1111/j.1444-0938.2007.00181.x

Wing Chun Tang[†‡] BSc(Hons)
Maurice KH Yap[†] PhD MCOptom FAAO
Shea Ping Yip[‡] MPhil PhD FIBMS
[†]School of Optometry and the
[‡]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China
E-mail: shea.ping.yip@polyu.edu.hk

Submitted: 13 March 2007
Revised: 9 May 2007
Accepted for publication: 17 May 2007

The prevalence of myopia is high in many parts of the world, particularly among the Orientals such as Chinese and Japanese. Like other complex diseases such as diabetes and hypertension, myopia is likely to be caused by both genetic and environmental factors, and possibly their interactions. Owing to multiple genes with small effects, genetic heterogeneity and phenotypic complexity, the study of the genetics of myopia poses a complex challenge. This paper reviews the current approaches to the genetic analysis of complex diseases and how these can be applied to the identification of genes that predispose humans to myopia. These approaches include parametric linkage analysis, non-parametric linkage analysis like allele-sharing methods and genetic association studies. Basic concepts, advantages and disadvantages of these approaches are discussed and explained using examples from the literature on myopia. Microsatellites and single nucleotide polymorphisms are common genetic markers in the human genome and are indispensable tools for gene mapping. High throughput genotyping of millions of such markers has become feasible and efficient with recent technological advances. In turn, this makes the identification of myopia susceptibility genes a reality.

Key words: genetics, linkage, linkage disequilibrium, myopia, polymorphism

Myopia is the most common eye disorder in the world. The prevalence of myopia is estimated to be 17, 26 and 27 per cent in Australia, the United States and Western Europe, respectively.[1–7] It is especially high (71 to 96 per cent) in Asian regions and countries including Hong Kong, Taiwan and Singapore.[8–11] Severely myopic eyes are prone to degenerative changes, such as retinal degeneration and glaucoma, which can lead to visual impairment. In the long term, such problems will impose a heavy burden on the health care system and the economy of the society concerned.

The development of myopia is usually due to the excessive growth of the eyeball, as evident from many biometric studies on the eyes of humans and other animals.[12] The aetiology or mechanism underlying this abnormal ocular growth is still unclear, while the debate about the role of nature and nurture in myopic development continues.[13,14] From a simplistic viewpoint of genetics, myopia can appear as part of a rare disease syndrome, as a rare monogenic trait (or disease), or more commonly, as a complex trait (or disease) (Figure 1).

Myopia, usually high myopia, is sometimes presented as one of the features in a wide variety of rare heritable disease syndromes.[15] Many of these uncommon syndromes are single-gene disorders, such as Stickler syndrome and Marfan syndrome, and some are chromosomal in nature such as Down syndrome. There are also rare cases of myopia, typically high myopia, that show simple Mendelian inheritance patterns (dominant, recessive or sex-linked).[16] By definition, such cases probably represent the monogenic form of myopia, which is caused by mutations in a single gene. It is very likely that there is a handful of such myopia genes and mutations in each of these genes are expected to cause directly the monogenic form of myopia.

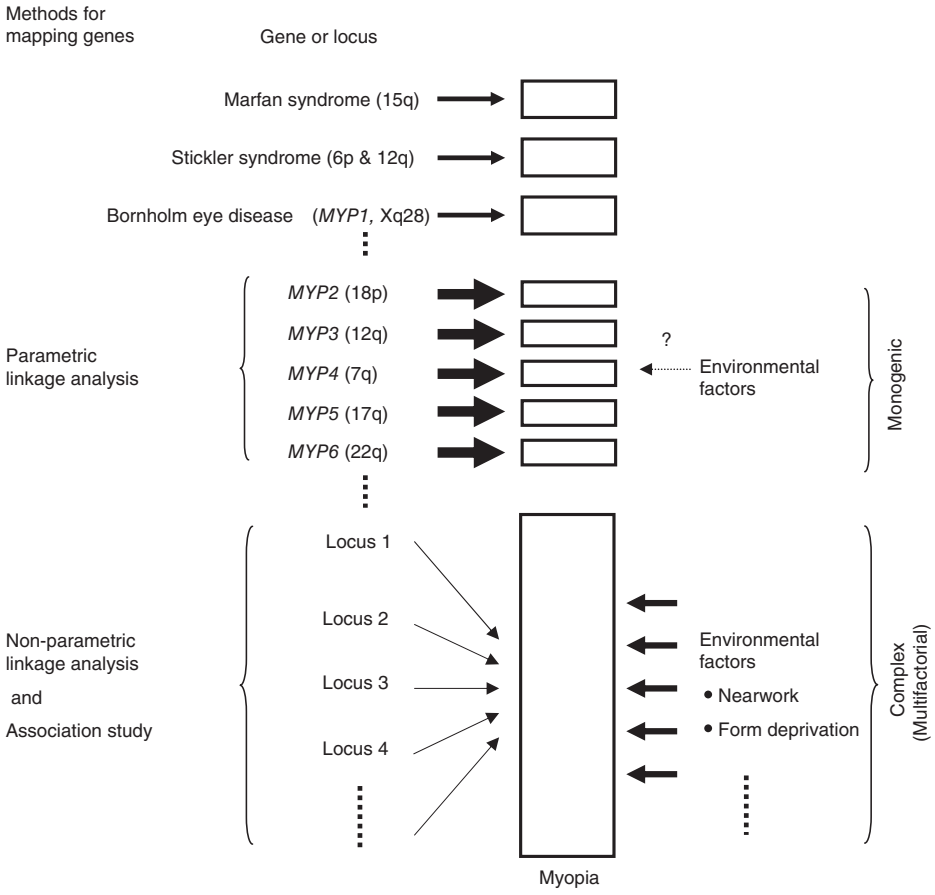

Figure 1. Myopia can appear as part of a syndrome, in rare monogenic form or more commonly as a complex trait/disease. The boxes represent syndromes (with myopia as one of the presenting features) or myopia (monogenic or complex forms). The size of the boxes serves as a very crude indication of the frequency of the syndromes or different forms of myopia. The relative thickness of the arrows serves to roughly indicate the relative contribution of genetic and environmental factors to the genesis of different forms of myopia. The influence of environmental factors on the monogenic form is probably minimal, as indicated by the question mark. The complex form of myopia is produced by the action of many genes (locus 1, locus 2 et cetera) and many environmental factors. Also indicated on the left are the appropriate methods for mapping genes involved in monogenic and complex forms of myopia. Note that the first myopia locus *MYP1* was assigned to the clinical syndrome Bornholm eye disease manifested as amblyopia, myopia and deuteranopia. *MYP1*, *MYP2* et cetera are standard symbols of myopia genes approved by the HUGO Gene Nomenclature Committee. The approximate chromosomal locations (15q, 18p et cetera) are indicated within brackets after the names of the syndromes or myopia genes.

On the other hand, there is evidence that both heredity and environment play a role in the development of common myopia (Figure 1). Common myopia is like many other common diseases that are 'complex' or 'multifactorial' in nature and result from the effects of both genetic and environmental factors.[17–20] Environmental factors such as excessive near work and form-deprivation may interact with genetic factors to produce abnormal eye growth. By definition, a complex disease/trait is caused by multiple genes, multiple environmental factors and possibly gene-gene and gene-environment interactions. Therefore, a complex disease/trait tends to run in families but does not show a typical Mendelian inheritance pattern. A single susceptibility gene is neither necessary nor sufficient to cause a complex disease/trait and genetic effects involve probabilistic predisposition rather than predetermined programming. Complex diseases are exemplified by diabetes, hypertension and coronary heart disease, while common examples of complex traits include height, body weight and blood cholesterol concentration. As an analogy with myopia, non-insulin-dependent (type 2) diabetes shows the characteristics of a complex disease. It also has a rare monogenic form showing typical Mendelian inheritance, which is caused by mutations in one of several different genes.

Twin studies have presented strong evidence for genetic inheritance in the control of eye size and refractive errors. Monozygotic twins have much higher concordance of myopia and its related ocular components, such as axial length, anterior chamber depth and corneal curvature, than dizygotic twins.[21–26] Family studies indicate that children are more likely to develop myopia and have longer anterior and vitreous chambers even before becoming myopic, if their parents are myopic.[18,27,28] In general, estimates of heritability indicate that axial length, anterior chamber depth and corneal curvature are predominantly genetically determined.[29–33] This suggests a role for genetic influences in eye shape and development of myopia.

Most myopia is due to excessive eye growth and individuals with an apparently inherited form of the condition exhibit excessive elongation of the eye.[26,34] Thus, axial length and refractive error are similar to height, which is a continuous and complex trait and is determined by both genetic factors and environmental factors (like diet and exercises). Instead of a single gene, multiple genes exhibit effects on the variations in both eyeball size and height.

Heterogeneity and complexity in phenotypic expression make the study of complex diseases much more challenging.[35] Heterogeneity refers to the situation in which a disease is contributed by more than one gene or different genetic variations in the same gene or different genes. The expression of phenotype can be variable in that the same genetic variation can apparently result in different degrees of severity for a disease. In addition, some individuals with a particular genetic variation may not express the phenotype, while others without the genetic variation may somehow express the phenotype because of some other causes. Despite these complexities, many researchers make serious attempts at mapping the genes of myopia in humans by different approaches of genetic analysis. Gene mapping refers to the process of locating or identifying a gene or a group of genes within our genome for subsequent studies. As the study of complex diseases identifies a significant contribution of heredity to their development, it is likely that more genes will be found to influence susceptibility to complex diseases rather than to cause disease directly. Genetic susceptibility refers to the ability of genetic factors to affect the risk of contracting a disease. Exposure to appropriate environmental risk factors may trigger the expression of the genetic effects on a complex disease, alter its manifestation or exacerbate its severity.

Identification of susceptibility genes of myopia will shed light on the pathophysiological mechanism underlying their genesis. Undoubtedly such information is important for the design of preventive and therapeutic measures in the future. One can imagine that, in the not too distant future, it may be possible to use genetic tests to identify genetically susceptible individuals at a very young age even before myopia develops. Such individuals may be advised to modify their lifestyle (such as reading habits) or treated to delay the onset of myopia, slow down the progression or even prevent it from developing. Susceptible individuals with different backgrounds of genetic predisposition may develop myopia on exposure to different environmental triggers and need different kinds of advice and different treatment modalities. One treatment may be effective for those of one particular genetic background but not those of another. This form of long-awaited personalised medicine can be realised only with the elucidation of the molecular pathways involved in the trigger and development of myopia. In turn, this relies on the identification of myopia genes and the subsequent study of their functions.

The purpose of this article is to elucidate the approaches of genetic epidemiology to identifying the genes involved in myopia. This review is intended to help optometrists in the understanding and interpretation of reports on genetics of myopia and other complex eye diseases. Before going through these approaches, different types of genetic markers that are useful tools in disease gene mapping are first introduced. Such information is essential for understanding the basic concepts of genetic analysis.

GENETIC MARKERS

Deoxyribonucleic acid (DNA) contains the genetic instructions for the biological processes of a cellular form of life and is responsible for the genetic propagation of inherited traits. In humans, these traits range from eye colour to disease susceptibility. Mutation is a heritable alteration of this genetic material. It can result from errors that occur during the replication of DNA before a cell divides and that escape the repair mechanisms of the cell. It can also be induced by substances known as mutagens and can be transmitted to subsequent generations. Mutation is the ultimate source of genetic variations among individuals of the same species and among difference species. If a variation in DNA sequence occurs by chance and reaches a frequency of at least one per cent in a population, it is classically called a polymorphism. Not all polymorphisms have phenotypic effects or direct effects on the genesis of disease. Some may have no effect at all and some may just play a role in whether a person has high or low risk of a particular disease. The term 'polymorphism' is a neutral term and can simply mean sequence variation. By contrast, particularly in a clinical setting, mutation is used typically in a more restricted and negative sense, implying disease-causing genetic variations.

A locus is a specific physical location in a chromosome of the genome. A gene is a stretch of DNA sequence that instructs a cell to produce a particular product, usually a protein. Alleles are alternative forms (two or more forms) of any defined gene or DNA sequence at a given locus, and usually differ from one another by one or a few bases in their DNA sequences. In humans, all cells except sex cells (sperm cells and eggs) are diploid and hence they each contain two sets of chromosomes, one from each parent. The terms homozygous and heterozygous are used to describe the presence of the same allele and different alleles, respectively, at a locus for a given pair of homologous chromosomes. Genotype is the specific genetic make-up of an individual, in the form of DNA. Together with the environmental factors that influence the individual, it codes for the phenotype (observable trait) of that individual.

A genetic marker is an inherited genetic trait with different possible forms that are readily recognisable and help researchers tell apart different individuals in an experiment. Any DNA sequences existing as two or more distinguishable alternative forms (alleles) can be used as genetic markers. If a disease is inherited in a simple pattern, the inheritance of disease alleles or genetic variants can be followed in pedigrees. Known polymorphisms are useful genetic markers for genetic studies or gene mapping. Rare mutations usually contribute to rare Mendelian genetic diseases. In recent years, many researchers believe that common genetic variations contribute to the genetic risk of developing complex diseases.[36,37] Many studies relate common genetic variations to the clinical phenotypes of complex diseases, drug responses and environmental factors. There is no doubt that the development of a high-resolution map of genetic variations is a very important part of the Human Genome Project (HGP), in which the DNA sequences and the genes of the

entire human genome are determined. The complete sequence is now known and this provides several million polymorphisms spreading across the genome for genetic studies.[37–40]

Two types of genetic markers are used commonly in gene mapping of complex diseases: microsatellites and single nucleotide polymorphisms (SNPs). They are abundantly present in the human genome. The single nucleotide polymorphism database (dbSNP) at the National Centre for Biotechnology Information (NCBI) provides detailed information of these genetic variations in humans and many other different organisms. The information includes the surrounding sequence context of the polymorphism, the allele frequency of the polymorphism and the experimental methods, protocols and conditions used to assay the variation.

(A) Microsatellite (GT repeats)

..TAC(GTGTGTGTGTGTGTGTGTGT)ACC..

..TAC(GTGTGTGTGTGTGTGTGTGTGT)ACC..

..TAC(GTGTGTGTGTGTGTGTGTGTGT)ACC..

..TAC(GTGTGTGTGTGTGTGTGTGTGTGTGT)ACC..

..TAC(GTGTGTGTGTGTGTGTGTGTGTGTGT…)ACC..

(B) Single nucleotide polymorphism (SNP)

……GCCTCCGTCAGTGCTGCCT……

……GCCTCCATCAGTGCTGCCT……

Figure 2. Examples of microsatellite and single nucleotide polymorphism. A: The top panel shows a multi-allelic microsatellite with each allele carrying at least 11 GT dinucleotide repeats (placed within brackets for the sake of clarity). B: The bottom panel shows a di-allelic single nucleotide polymorphism with a G allele and an A allele (underlined). Short stretches of DNA sequences flanking these genetic markers are also shown.

Microsatellites

Microsatellites are also called short tandem repeats or simple sequence length polymorphisms and are arrays of short repeat sequences (usually di-, tri- and tetra-nucleotide repeats) that display length variations with different alleles carrying different numbers of repeat units (Figure 2A). They are multi-allelic, usually between five and 10 or more alleles at each locus. For example, there may be five alleles at a certain locus of GT dinucleotide tandem repeat polymorphism and each allele has GT repeat for 11, 12, 13, 14 or 15 times, giving length differences in the multiples of two bases (or nucleotides). Such DNA length differences can be detected by automatic DNA sequencers via capillary gel electrophoresis, which separates microsatellite-containing fragments that are first amplified by a Nobel prize-winning technique called polymerase chain reaction. Genetic maps, which are composed of several hundred evenly spaced microsatellites with known positions, allow locations of unknown genes to be determined in relation to those of microsatellites. Such information is easily accessed in NCBI databases. Microsatellites are usually applied to linkage analysis, which can map a gene of interest to a large region of chromosomal segment extending a few thousand kilobases (kb). Many loci for high and low myopia were successfully mapped by linkage analysis using microsatellites (Table 1).[41–58]

Single nucleotide polymorphisms (SNPs)

A SNP (pronounced as snip) is a DNA sequence variation due to the change in a single base (Figure 2B). About 99.9 per cent of the DNA sequences in the genome of any two unrelated individuals are identical. Of the 0.1 per cent difference, SNPs account for over 80 per cent of the sequence variations in the human genome. Although they are distributed unevenly in the genome, on average they can be found at intervals of 500 to 1,000 basepairs.[39] The simple bi-allelic nature of SNPs allows a great potential for automated genotyping at different levels of throughput.[59–61] Indeed, the availability of many different SNP genotyping methods (details not mentioned here) facilitates and accelerates the gene mapping process of complex eye diseases.

Variations occurring in the coding sequences (exons) of a gene are termed coding SNPs. DNA sequence variations that are not expressed at the protein level are called non-coding SNPs. Non-coding SNPs can be found in introns (non-coding sequences between exons of a gene), the promoter region, 5′ untranslated region (5′UTR), 3′ untranslated regions (3′UTR) of a gene and the large amount of DNA sequences between any two genes. Some coding SNPs may have functional consequences and hence phenotypic effects, if they change the amino acid sequence of the protein encoded by the gene and they are called 'non-synonymous'. For example, if a codon sequence GGC is changed to GAC, the encoded amino acid will be changed from glycine to aspartic acid and this is a non-synonymous change. Some SNPs in exons do not change encoded amino acid. For example, if a codon GGC is changed to GGA, the same amino acid (glycine) is encoded and this is called a 'synonymous' change. In addition, not all of the functional changes lie within exons.

Myopia locus[†]	Authors (years)	Inheritance/ QTL[‡]	Location	Ethnicity of subjects	Types of families	Linkage analysis (PL or NPL)[§]	Affected status[¶]	Max LOD[††]
MYP1	Schwartz *et al* (1990)[41]	XR	Xq28	Danish	Large pedigrees	PL	—	4.80
MYP2	Young *et al* (1998)[42]	AD	18p11.31	American and Chinese	Moderate to large multigenerational families	PL	≤ -6.00 D SE	9.59
MYP2	Lam *et al* (2002)[43]	AD	18p11.31	Hong Kong Chinese	Moderate pedigrees	PL	≤ -6.00 D	2.10
MYP3	Young *et al* (1998)[44]	AD	12q21-23	German/Italian	A large pedigree	PL	≤ -6.00 D SE	3.85
MYP3	Farbrother *et al* (2004)[45]	AD	12q21-23	UK population	Nuclear families	PL, NPL	≤ -6.00 D in the least negative meridian of both eyes	2.54
MYP4	Naiglin *et al* (2002)[46]	AD	7q36	French and Algerian	Large to moderate pedigrees	PL, NPL	≤ -6.00 D both eyes	2.81
MYP5	Paluru *et al* (2003)[47]	AD	17q21-22	English/Canadian	A large pedigree	PL	≤ -6.00 D SE	3.17
MYP6	Stambolian *et al* (2004)[48]	AD	22q12	American families of Ashkenazi Jewish descent	Large pedigrees	PL, NPL	≤ -1.00 D in each meridian for both eyes	3.54
MYP6	Stambolian *et al* (2006)[49]	AD	22q12	Additional Jewish descent	Pedigrees	PL, NPL	≤ -1.00 D in each meridian for both eyes	4.73
MYP6	Klein *et al* (2007)[50]	QTL	22q12	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38 D	*P* value = 0.00330
MYP7	Hammond *et al* (2004)[51]	QTL	11p13	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	6.10
MYP8	Hammond *et al* (2004)[51]	QTL	3q26	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	3.70
MYP9	Hammond *et al* (2004)[51]	QTL	4q12	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	3.30
MYP10	Hammond *et al* (2004)[51]	QTL	8p23	UK population	Dizygotic twin pairs	NPL	Mean SE < 0 D	4.10
MYP10	Stambolian *et al* (2005)[52]	AD	8p23	Old Order Amish	Families with affected sibs	PL, NPL	≤ -1.00 D in each meridian for both eyes	2.03
MYP11	Zhang *et al* (2005)[53]	AD	4q22-27	Han Chinese in a small village of central China	A large pedigree	PL	Range: -5.00 to -20.00 D	3.11
MYP12	Paluru *et al* (2005)[54]	AD	2q37.1	US family of northern Europe	A large pedigree	PL	≤ -6.00 D SE; range: -7.25 to -27.00 D	4.75
MYP13	Zhang *et al* (2006)[55]	XR	Xq23-25	Chinese	A large pedigree	PL	≤ -6.00 D SE; range: -6.00 to -20.00 D	2.75
MYP13	Zhang *et al* (2007)[56]	XR	Xq23-27.2	Chinese	A large pedigree	PL	≤ -6.00 D SE; range: -7.00 to -16.00 D	2.79
MYP14	Wojciechowski *et al* (2006)[57]	QTL	1q36	Ashkenazi Jewish	Moderate to large multigenerational families	PL	Mean -3.46 D SE; ≤ -1.00 D in each meridian for both eyes	9.54
MYP15	Klein *et al* (2007)[50]	QTL	1q41	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38 D	*P* value = 0.00019
MYP16	Klein *et al* (2007)[50]	QTL	7p21	Americans of Northern European and/or German ancestry	Sib-pairs	NPL	Mean +0.44 D SE; range: -12.12 to +8.38 D	*P* value = 0.0023
MYP17	Nallasamy *et al* (2007)[58]	AD	10q21.2	Hutterite population from South Dakota	A large pedigree	PL	Mean -7.04 D; range -3.75 to -13.25 D	3.22

[†] Myopia loci are shown in standard gene symbols approved by Human Gene Nomenclature Committee but *MYP15*, *MYP16* and *MYP17* are newly identified and the gene symbols are still tentative
[‡] Inheritance is indicated as X-linked recessive (XR), autosomal dominant (AD). QTL represents quantitative trait locus.
[§] PL represents parametric linkage analyses whereas NPL represents non-parametric linkage analyses, usually affected sibpair analysis
[¶] SE represents spherical equivalent in dioptres (D)
[††] Max. LOD stands for maximum 'logarithm of the odds' score. Only *P* values are given for *MPY6*, *MYP15* and *MYP16* in Klein's study.

Table 1. Myopia loci mapped by linkage analysis

Non-coding SNPs may have functional consequences, if they affect the regulation of gene transcription, mRNA splicing or mRNA stability. Promoters and regulatory elements affect gene transcription and are usually located at 5′UTR and the upstream region of a gene. Some regulatory elements are found at 3′UTR or in introns where they may be close to the splice sites.[62] Therefore, efforts to identify functional SNPs should include the proximal and distal regulatory sequences in the 5′ and the 3′ ends of the gene.

SNPs are present in healthy normal individuals, who can be recruited easily for any study. SNP screening within candidate genes (small-scale) or throughout the whole genome (large-scale) can be conducted before the collection of patient samples is completed. Thus, SNPs have become the most popular genetic tools for gene mapping of complex diseases. The NCBI maintains the dbSNP database as the major public repository for SNP information, although many other SNP databases (with or without charges) are also available in both public and private sectors. More than five million SNPs with minor allele frequency greater than 10 per cent are expected to exist in the human genome.[63] In fact, about four million SNPs across the human genome in four different ethnic populations (Nigerian, American, Japanese and Chinese) have now been genotyped in the International Haplotype Map (HapMap) Project (see below). SNP allele frequencies and their linkage disequilibrium (LD; see below) patterns can be easily accessed in the HapMap database.

Comparison of microsatellites and SNPs

Microsatellites are more informative when compared with SNP markers. It is because, on average, many microsatellites are heterozygous in about 70 per cent of human samples tested.[64] They are often located between genes rather than within genes. Classically, genetic studies of linkage analysis use microsatellites to locate a disease gene to a coarse chromosomal region. SNPs within and flanking genes in the candidate region can then be used as genetic markers for further studies. Although a single SNP marker is less informative than a microsatellite marker, analysis of multiple SNPs, and hence haplotypes, enriches the information content and improves the efficiency for gene mapping. Haplotype is the combination of alleles of two or more loci closely linked on the same chromosome. SNPs are also more stable, especially for those that are not involved in amino acid change because they are less likely to be subjected to natural selection and thus, do not change much from generation to generation.[65] This makes them easier to follow in population studies.

PHENOTYPE DEFINITION

A phenotype refers to a specific feature of an individual that can be observed, categorised or measured by some means. From the perspective of gene mapping, phenotypes can be classified into two broad types: qualitative (or discontinuous) and quantitative (or continuous) traits. A qualitative trait is one with two or more distinct categories and each category can be easily separated from all other categories (for example, the ABO blood group of an individual). In most medical conditions, 'diseased' traits are recognised by their state of being significant deviations from the norm. Signs and symptoms indicate whether a disease state is present or absent. On the other hand, a quantitative trait exhibits a wide range of possible 'categories' that show a continuous spectrum of variation (for example, body weight and blood pressure). It does not follow typical Mendelian inheritance. Very likely, it is controlled by many genes with small effects that are influenced by the environment and their interaction with other loci.

Myopia is a refractive error that is measured in a continuous scale and in the unit of dioptre (D). The refractive error of an eye is the result of the total refractive power of ocular components, which include axial length, anterior chamber depth, corneal power and lens thickness, which are also continuous traits. Traits with well-defined Mendelian patterns of phenotypic expression are optimal for linkage analysis (Figure 3). Therefore, continuous traits are usually converted into discontinuous or dichotomous traits by defining a 'threshold' for the disease of interest for genetic studies. Different definitions of myopia have been adopted in different clinical and genetic studies and this leads to difficulties in comparing the results among studies. The most common definition of myopia is a refractive error with equivalent spherical power less than -0.50 D. In most studies of linkage analysis, low myopia is defined as a refractive error between -0.50 D and -6.00 D, whereas high myopia refers to a refractive error equal to or over -6.00 D. A few genetic studies used a more minus refractive power as the cut-off threshold for high myopia to increase the statistical power to detect modest and small gene effects.[41,66,67]

Most myopic eyes have a longer axial length.[26,34,67] While heredity has a strong impact on axial length, the role of heredity on anterior chamber depth and corneal power is still controversial[26,29–33,70] but profiles of these ocular components vary among myopes. Strict inclusion criteria and careful phenotyping of myopic subjects are important for the success of genetic studies of myopia. Axial length can be taken into account in myopic subject recruitment to reduce the complexity of the genetic factors potentially involved. As large amounts of astigmatism will have an effect on the final equivalent sphere, myopic subjects with high astigmatism (-2.00 D or worse is suggested but there is no consensus)[68,69] should be excluded. This is because the genetic and environmental influences of astigmatism may be different from those of myopia. To illustrate this potential classification error, consider an eye with -4.00 D of spherical error and -4.00 D of astigmatism. The final equivalent sphere will be -6.00 D, qualifying it to be classified as an eye with high myopia.

APPROACHES TO FINDING GENES INVOLVED IN MYOPIA

Although there are two major classical approaches to finding genes involved in human traits and diseases, one of these approaches known as *genetic mapping* is

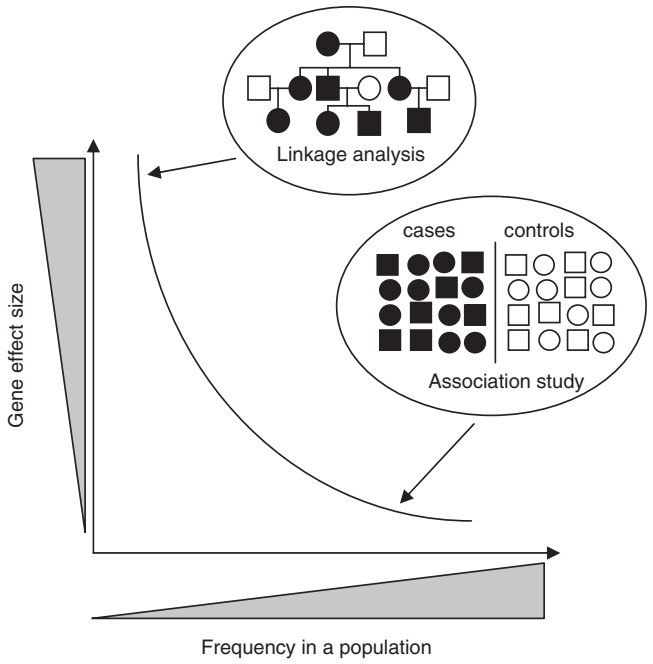

Figure 3. Gene mapping strategies for diseases different in population frequency and in gene effect size. Genes involved in rare Mendelian diseases show large effects and are best mapped by parametric linkage analysis of large families. Genes involved in common complex diseases exhibit small effects and are best identified by genetic association studies like case-controlled studies and family-based association studies. Non-parametric linkage analysis (not shown) falls between these two approaches.

more appropriate for identifying genes for myopia. This genetics-based approach relies only on the behaviour of genes during transmission from generation to generation in a family. It can be applied even though the underlying molecular mechanism of the disease of interest—myopia—is not known. In fact, the only requirement for applying this approach is that the disease or trait can be clearly defined. Indeed, myopia can be defined objectively and unambiguously by the amount of refractive error.

With genetic mapping, there are two strategies: linkage analysis and association study (Figure 3). For each strategy, there are different methods that deal with different scenarios and issues. Linkage analysis can be parametric or non-parametric. Parametric linkage analysis requires the assumption of a correct genetic model, while non-parametric linkage analysis does not assume any genetic model. On the other hand, association studies can be population-based or family-based. Population-based association studies work on unrelated individuals with or without the disease of interest, while family-based association studies work on affected children from small nuclear families.

LINKAGE ANALYSIS

Parametric linkage analysis

If a disease gene and a genetic marker locus are located on different chromosomes, the alleles of the disease gene are transmitted to the next generation independently of the alleles of the marker locus. In other words, these two sets of alleles are free to form new combinations during transmission to the next generation—a situation as dictated by Mendel's second law (the principle of independent assortment). If a disease gene and a genetic marker locus are located far apart on the same chromosome, the chance of exchanging genetic material and hence the alleles (that is, recombination) between these two loci is very likely to occur as a result of crossing-over between a pair of homologous chromosomes during meiosis. In other words, the transmission of the alleles of these two loci still obeys Mendel's second law. Thus, the transmission of alleles at each locus from generation to generation is random and independent of the other locus under these two conditions. These loci are said to be unlinked. In this case, the ratios of alleles of the two loci observed in each generation will be close to those predicted by the Mendelian principle of independent assortment and the rate of recombination is 50 per cent.

If the marker locus is close to the disease gene on the same chromosome, they are more likely to segregate together (without crossing-over) rather than independently during meiosis. Linkage between these two loci occurs and the recombination rate is less than 50 per cent.[71] In other words, the recombination rate becomes smaller when two loci are much closer to each other. By taking advantage of these properties of meiosis, measurement of genetic linkage can take place in family studies for disease gene mapping.

Linkage analysis is a method for determining if there is significant evidence for co-segregation of alleles at a marker locus and alleles at a hypothetical disease locus. It involves explaining the inheritance patterns of genotypes observed in a pedigree affected by a disease. This analysis is useful only when the disease model is correctly specified and usually is applied to simple Mendelian diseases or traits, which show strong effects of the mutations involved. Thus, this method is called model-based

(or parametric) linkage analysis. If a disease is caused by a dominant mutation (M) at gene G on a chromosome, all affected members in a pedigree should inherit this chromosomal region around the mutant allele (Figure 4). Assume that two genetic markers are close to each other (linked) on the same chromosome. When the disease mutation is located between these two linked genetic markers, affected members in the pedigree are expected to inherit two particular alleles of these two markers (A and b) together with the mutant allele (M) of the disease gene from an affected parent. The known position of the linked markers indicates the approximate chromosomal location of the disease-causing gene.

The linkage of markers to disease mutation is usually quantified by the 'logarithm of the odds' (LOD) score, which provides a statistical evaluation of the co-segregation of the marker and the mutation. Under a specific mode of inheritance (for example, autosomal dominant), the likelihood of getting specific marker alleles and a disease (that is, the disease mutation) in a large family under two hypotheses are compared. The hypothesis that the marker is linked to the disease locus at a particular recombination rate (θ) is compared to the null hypothesis that there is no linkage between the marker and the disease locus (that is, $\theta = 0.5$). Therefore, a likelihood ratio can be calculated as the ratio of the likelihood of the data, if the loci are linked at a particular θ to the likelihood of the data, if the loci are unlinked ($\theta = 0.5$). The $\log_{10}$ of this ratio is the LOD score. The LOD score serves to indicate whether there is linkage at a particular value of θ between the marker and the disease locus. A LOD score of 3 (an odds ratio of 1000 : 1) is traditionally agreed as evidence of linkage.[35]

The LOD score approach has the advantage that linkage studies can be compared and analysed together. It is because LOD scores can be simply added together (that is, odds ratios multiplied together) when the linkage for different families with the same genetic markers is investigated. For example, if the LOD scores of families A and B are 1.5 and 2.5, respectively, at $\theta = 0.01$, neither shows convincing evidence of linkage on its own. If two LOD scores are added together, the data become significant with a LOD score of 4.0 but the same diagnostic criteria and genetic model must be used for both families. Moreover, a large pedigree with several generations is more informative than several small families each with fewer generations and hence more powerful to detect linkage.

To increase the power of detecting linkage, several marker loci can be simultaneously tested and analysed for linkage with the disease: 'multipoint linkage analysis'. The combination of alleles of these genetic markers that inherit together in a family with the disease is called the linked haplotype. For example, the disease mutation (allele M) and the markers (alleles A and b) are inherited together on the same chromosome, as a haplotype unit (A-M-b) in two affected offspring (Figure 4). The recombinant haplotype A-M-B is transmitted to one affected offspring (Figure 4).

In past decades, many eye disease or eye-related loci were successfully mapped

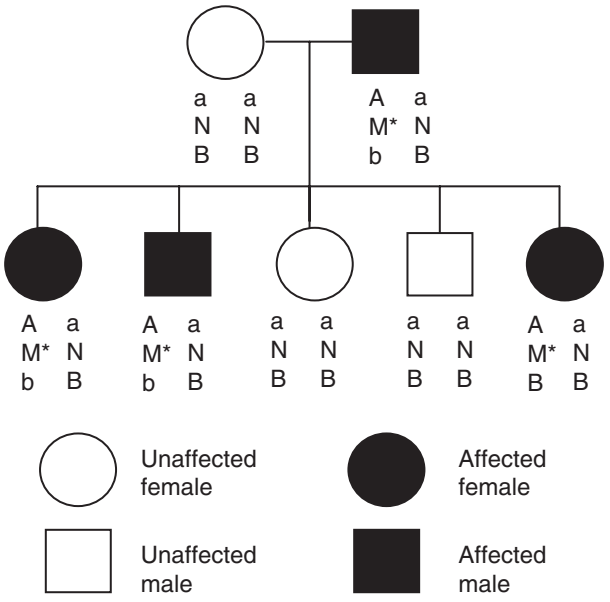

Figure 4. Linkage and recombination. Allele M is the disease allele caused by a mutation in the gene G, whereas allele N is the normal allele at the same locus. Note that allele M is marked by an asterisk for easy recognition. The affected father (black square in the first generation) carries the disease allele M on a chromosome together with marker alleles A and b. Two affected offspring (shown on the left in the second generation) also inherit the alleles A and b from the father because these two alleles are linked to the disease-causing allele M on the same chromosome. The youngest affected daughter (shown on the right) inherits the marker allele B (instead of allele b) from the father as a result of recombination between the disease locus (M/N) and the marker locus (B/b) during meiosis. The disease locus is linked to these two marker loci and is closer to marker locus A/a than to marker locus B/b.

	Parametric linkage analysis	Non-parametric linkage analysis
Family samples	Large pedigrees with many diseased subjects	At least two affected members for each family Best when parents available
Mode of inheritance	Assumption of mode of inheritance (AD, AR and X-linked)†	Model-free
Statistical power	High for Mendelian diseases Low for complex diseases	Moderate for Mendelian diseases and large genetic effects of complex diseases
Advantages	Highest power for Mendelian diseases More efficient using genome scan approach	Model-independent High power to detect moderate and large genetic effects of complex diseases Genome scan possible
Disadvantages	Need to ascertain pedigrees that are difficult to find Specific genetic model required Limited by heterogeneity if present	Problems of genotyping errors, non-paternity and sample mix-up detectable only when parents and/or siblings available Not sensitive to detect small genetic effects

† AD stands for autosomal dominant and AR autosomal recessive

Table 2. Comparison between parametric and non-parametric linkage analyses

by the study of large pedigrees with linkage analysis (Figure 3). These eye diseases were usually monogenic, polygenic or syndromic (syndromes associated with ocular manifestations), such as Stickler syndrome[72] and Marfan syndrome.[73] They follow typical modes of inheritance: autosomal dominant (AD), autosomal recessive or X-linked. On the other hand, gene mapping for complex eye disorders including myopia is still in its infancy. Linkage analysis is usually used as the first step in the study of complex eye diseases, as there is very little knowledge of the disease mechanism. For large-scale studies, a whole genome scan is carried out with several hundred (400 to 800) microsatellites spaced evenly throughout the genome. This means a systematic search for genetic effects at different locations along all human chromosomes but it is very expensive and time-consuming.

Table 1 summarises the myopia loci mapped to date. Updated information can be accessed easily at Online Mendelian Inheritance in Man (OMIM). The first myopia locus *MYP1* was mapped to Xq28 in 1990 and was responsible for X-linked high myopia (pathological myopia) associated with amblyopia and deuteranopia.[41] More loci for high myopia were found subsequently, using linkage analysis with the assumption of AD models in most cases. These non-syndromic high myopia loci were located at chromosome 18p11,[42,43] 12q21-23,[44,45] 7q36,[46] 17q21-22,[47] 4q22-27,[53] 2q37,[54] Xq23-25[55,56] and 10q21.2.[58] The symbols 'p' and 'q' represent the short and the long arms of a chromosome, respectively, and the numbers before and after 'p' or 'q' indicate the chromosome number and the sub-region in the corresponding arm respectively. The loci for low myopia were mapped at chromosome 22q12 and 1q36 by Stambolian's group and Wojciechowski's group, respectively.[48,49,57] Different myopia loci mapped for different ethnic groups support the presence of genetic heterogeneity.

Complex eye disorders clearly do not follow Mendelian inheritance patterns. Standard linkage analysis has limited power in detecting such small genetic effects in these disorders. If a wrong model is specified, it is likely to miss the true linkage. Moreover, it is usually very difficult to recruit large pedigrees. Mapping genes for complex eye disorders was amenable to 'non-parametric' linkage analysis and association study. Both parametric and non-parametric analyses were adopted for the same set of family samples in a few studies on myopia to increase the chances of identifying the myopia genes.[45,46,48,49,52]

Non-parametric linkage analysis: allele-sharing methods

The allele-sharing method, also known as affected pedigree member method, is an alternative to the parametric linkage analysis for identifying disease locus and has some advantages over classical linkage analysis (Table 2). First, no assumption about the mode of inheritance is needed. Second, small nuclear families instead of large pedigrees are used and are easier to recruit. Third, the fraction of a genetically heterogeneous phenotype that is determined by a given locus can be estimated. This method is more robust than parametric linkage analysis but less powerful than the linkage analysis with a correctly specified mode of inheritance. The allele-sharing method is still very useful when the disease models are unclear, especially for complex diseases.

The allele-sharing method aims at determining if affected relatives in a pedigree inherit specific copies of a chromosomal region more often than expected by chance of random segrega-

tion (Figure 5).[35] In other words, it studies how often a particular allele at a marker locus is shared identical-by-descent (IBD) within the pedigree. IBD means the inheritance of the same allele from a common ancestor. If there is a disease-causing mutation in a specific chromosomal region in a high proportion of families, two affected individuals from the same family will share an allele of the marker locus more often than expected by chance (50 per cent for sibling pairs).

Affected sibpair (ASP) analysis is the simplest type of allele-sharing method and is commonly used for gene mapping of complex eye diseases or traits. ASP analysis ignores the affection status of the sibs' parents and requires affected sibling pairs. Thus, this approach is very useful if parental genotype information is missing. In ASP analysis, the mean proportion of marker alleles shared by two affected siblings in a nuclear family is compared with a 50 per cent shared proportion expected by random Mendelian segregation. Under random segregation, the expected distribution of sharing of 0, 1 or 2 alleles IBD is 25, 50 and 25 per cent, respectively, between the sibling pairs of a nuclear family. If enough sibling pairs from many nuclear families are compared, excess sharing of alleles IBD is measured by simple chi-square (χ^2) test. Significant χ^2 indicates large deviation from the expected distribution and implies the linkage between the disease and the marker.[35] ASP analysis maps genes to a coarse chromosomal region and is not very powerful in detecting small genetic effects of complex diseases. Further investigation of positive loci should be followed by the alternative genetic approach: association studies.

As a huge number of genetic markers and advanced genotyping technologies is now available, allele sharing methods combined with the whole-genome scan approach have been used frequently to identify coarse chromosomal regions of many multifactorial eye diseases. Hammond and colleagues[51] mapped new myopia loci to chromosomes 11p13, 3q26, 4q12 and 8p23 by using ASP with twin pairs (Table 1). The locus on chromosome 8p23 was confirmed later in the Old Order Amish by Stambolian's group using the ASP method.[52] Recently, using sibpair analysis, Klein and associates[50] also identified novel regions of suggestive linkage to ocular refraction on 1q41 and 7q21 (Table 1). The same group also confirmed the *MYP6* locus at 22q12 previously mapped by Stambolian's group.[48–50] Nowadays, many software packages for ASP analysis are available, such as MAPMAKER/SIBS within GENEHUNTER. Many on-line resources provide the software for genetic linkage analysis, such as the Laboratory of Statistical Genetics at Rockefeller University.

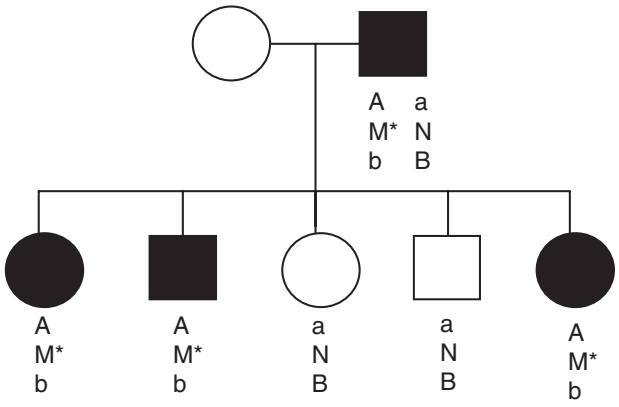

Figure 5. Allele sharing method. The disease locus has a disease allele M and a normal allele N. Note that allele M is marked by an asterisk for easy recognition. The disease locus is flanked by two linked marker loci carrying alleles A/a and B/b. The mutation M and also the haplotype A-M-b transmits to the three affected offspring from the affected father (the ancestor), and thus are shared identical-by-descent (IBD) between the affected offspring. Similarly, the normal N and hence the haplotype a-N-B are shared IBD between the two unaffected offspring. Note that only one of the two haplotypes is shown for each offspring. The meanings of the symbols are as shown in Figure 4.

ASSOCIATION STUDY

Association studies represent an alternative approach to identifying susceptibility genes for complex diseases. It is more powerful than linkage analysis in detecting genes with small effects in complex diseases (Figure 3).[36] This method is based on measuring linkage disequilibrium (LD) or allelic association between a marker and a causative genetic variation. When the marker and the causative genetic variation for a disease are physically close together on the same chromosome with tight linkage, they are likely to be inherited together as a particular combination of alleles from generation to generation in a population. This situation is known as LD and the set of associated SNP alleles in a region of the same chromosome is called a haplotype. Allelic association refers to a significantly increased or decreased frequency of a marker allele in a disease and represents deviations from the random occurrence of the alleles with respect to disease phenotype. Allelic association can be due to LD maintained by tight linkage. The HapMap Project of the human genome provides a tremendous amount of information on the patterns of variation in the genome of four different

ethnic populations.[74] It includes the chromosomal regions with sets of strongly associated SNPs, the haplotypes in those regions and the SNPs (called tag SNPs) that are representative of those haplotypes. Thus, one of the applications of the HapMap database is to provide tag SNPs that act as representatives for many other surrounding sequence variations. The researchers can genotype a much smaller number of tag SNPs to determine the collection of haplotypes present in each subject. This can help in the quest of complex disease genes by reducing the number of SNPs to be tested in association studies.

Population-based association study

Population-based association study (or case-controlled study) is the most widely applied strategy of association studies for complex diseases/traits. It compares a particular genetic marker or a set of markers between a group of affected or diseased individuals (cases) and a group of unaffected or normal individuals (controls) within a population. Both case and control groups should be unrelated to each other (that is, without any blood relation with each other). A greater percentage of a marker allele (or genotype) in affected individuals is considered as evidence of association between the disease phenotype and the marker allele (or genotype). The chi-square (χ^2) test is used for measuring the statistical significance of such differences between the allele (or genotype) frequency distributions in cases and controls.

In case-controlled association studies, unrelated individual samples rather than family samples are required. Subject recruitment for this method is easier than for linkage analysis or family-based association studies (see next section). Case-controlled association study becomes a better approach for gene mapping, if the eye diseases are late-onset or parents of the cases are not available for study, such as age-related macular degeneration and adult-onset glaucoma. A few myopia susceptibility genes, including transforming growth factor β-induced factor (*TGIF*),[75] transforming growth factor beta 1 (*TGFB1*)[76] and lumican (*LUM*)[77] have been identified by this approach in Chinese populations.

Positive association can arise in three situations.[35] First, it is due to the direct effect of a genetic variant on the risk of suffering from an eye disorder. In this case, the same positive association is expected in all populations tested. Second, positive association can occur if the marker SNP being tested is in LD with the causative genetic variant (that is, a certain allele of the marker SNP tends to occur on the same chromosome that carries the causative genetic variant). Different alleles of the marker may be associated with the disease in different populations because the pattern of LD among different SNPs can vary according to the history of the populations under study.

The third scenario is worrying because positive association can be falsely obtained if the population under study is not homogeneous. In other words, this false positive result is an artifact of population admixture, which refers to a population with multiple subgroups that have different allele frequencies. This can be due to either recent admixture of different populations or to inappropriate matching of patients and controls.[35] For illustration, consider a population that is a mix of two subgroups. Assume that subgroup 2 has a higher prevalence of a disease under study, a higher allele frequency of a SNP than subgroup 1 and subjects with the disease (that is, cases) are preferentially recruited from subgroup 2 in an unknowingly biased manner. This will give a higher allele frequency of the SNP in the cases than in the controls and, in turn, produce a spurious association between the SNP and the disease. Even if no differences in disease prevalence between different subgroups are present, spurious association may still arise if the probability of cases or controls being selected into the study is not independent of the original population subgroups (selection bias). The extent to which population admixture or stratification creates problems for association studies has been extensively discussed.[78–81] That many genetic association studies cannot be replicated may be due to this problem. Therefore, carefully matched cases and controls to avoid population stratification are essential for a high-quality association study.

Family-based association study

As discussed above, failure to match cases and controls in population-based association studies can lead to spurious association due to population stratification. This problem can be overcome if the family-based association study approach is adopted. Instead of unrelated individuals, nuclear families have to be used and each family consists of affected offspring and their parents. These family-based studies are based on the idea that non-transmitted alleles from the parents of the affected offspring act as internal controls, and transmitted alleles are the cases (Figure 6). Therefore, parents can provide both the cases and the controls, and the issue of ethnic mismatches can be avoided. The price to pay for this advantage is that the amount of genotyping work is increased by 50 per cent when compared with case-controlled association studies. Recent research also suggested that involvement of unaffected siblings would increase the power of studies of mapping loci for quantitative traits.[82–84]

The transmission disequilibrium test (TDT) is the most widely used version of family-based association tests. In essence, it tests for distortion (or disequilibrium) in the transmission of alleles from a heterozygous parent to an affected child by comparing the frequency of the allele transmitted to the affected child to the frequency of the allele not transmitted.[85] As an illustration, consider four heterozygous (AB) parents (Figure 6). If there is no association between the marker and the disease, alleles A and B are likely to be transmitted equally to the affected children. If allele A increases the risk of contracting the disease, it will be preferentially transmitted (three times out of four) to the affected children. In TDT analysis, McNemar χ^2 test is used to test the null hypothesis that each marker allele is transmitted from heterozygous parents to an affected child with a probability of 50 per cent expected by random Mendelian seg-

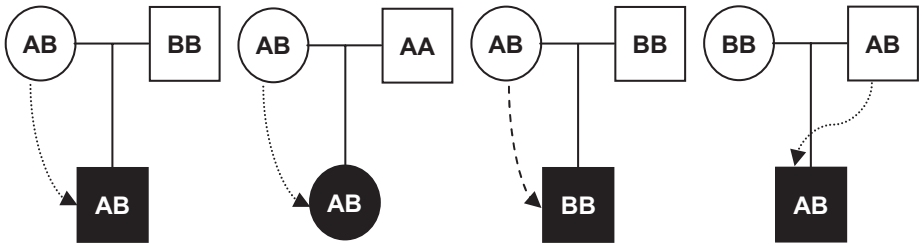

Figure 6. Transmission disequilibrium test (TDT). Four trio families are shown, each with an affected child. Each family has a heterozygous parent with genotype AB, who can transmit either allele A or B to the affected child. In three families out of four, allele A is transmitted to affected children from their heterozygous parents and allele B can act as an internal control. Apparently, allele A is transmitted to affected children more frequently than expected. In one family (second from the right), allele B is transmitted to the affected child from the heterozygous parent and allele A can act as an internal control. Homozygous parents (AA or BB) are not informative. For TDT study, heterozygous parents are needed and the frequencies of transmitted and non-transmitted alleles to affected offspring are compared. The McNemar chi-square test is used to test the statistical significance of the observed difference. If there are m affected children receiving allele A from their heterozygous parents and n affected children receiving allele B from their heterozygous parents, then the McNemar chi-square statistic is simply $(m - n)^2/(m + n)$ with one degree of freedom.

regation. The TDT can also be extended to families with more than one affected child (myopic offspring) because, under the null hypothesis of no linkage/association, every child in a family has an independent 50 per cent probability of inheriting each of the two alleles from a heterozygous parent.[86] Recently, the TDT approach has been used successfully to identify two myopia susceptibility genes in Chinese populations: the hepatocyte growth factor (*HGF*) and the myocilin (*MYOC*) genes.[67,87]

The TDT is less useful when the disease is due to the contribution of rare alleles. It is because of the small proportion of heterozygotes carrying the rare alleles within the population. Another disadvantage is that it is much more difficult to recruit families than unrelated individuals. The comparison of case-controlled association tests and family-based association tests is summarised in Table 3.

SCALE OF GENE MAPPING EFFORTS

No matter which of the above-mentioned genetic approaches is used, the scale of gene mapping efforts can vary tremendously. It can range from systematically investigating a large number of genetic markers throughout the human genome (genome-wide scan) to strategically testing only a few markers within a small candidate region (candidate-gene approach). Obviously, the time, cost and labour involved vary accordingly.

Genome-wide approaches

One of the common approaches to mapping genes implicated in complex diseases is genetic linkage analysis using a whole genome scan.[35] It entails a systematic search in large families for genetic effects at different locations along all human chromosomes even though the mechanisms underlying a disease are not known. Although the time and cost of this approach are really tremendous, this approach is useful for mapping a gene of interest to a large region of chromosomal segment extending about a few thousand kb in length. Genome-wide linkage scans usually employ several hundred microsatellite markers (in the range of 400 to 800) spaced at regular intervals in the human genome. Many myopia loci were identified using this approach in the recent decade (Table 1).[42,44,46–58] As there is an explosion of high-throughput and advanced SNP genotyping technologies,[61,88–90] efficient genome-wide linkage scans can now be achieved easily with 5,000 to 10,000 SNPs within a relatively short period.

High-throughput genotyping platforms have also recently made genome-wide association studies a reality.[91–94] We need to do association tests on all alleles for an extremely large number of markers (in the range of 0.1 to 1.0 million), an enterprise resulting in over one million data points for just a single individual. Correction for many different statistical comparisons is needed. It can be imagined that an astronomical number of samples is required to detect associations that are statistically significant even for moderate or small gene effects.

Candidate-gene approaches

According to the estimation of the Human Genome Project, there are about 30,000 genes in the human genome. It is very challenging, perhaps impossible for small-scale research studies, to test all the genes with association methods. Therefore, a candidate-gene approach instead of a genome-wide approach can be used for identifying susceptibility genes of complex eye diseases. This is also widely applied to other complex diseases such as diabetes and hypertension. This means that some candidate genes are first selected and DNA sequence variations (as genetic markers) within and flanking the candidate genes are then characterised for subsequent use in association studies.

Candidate genes are usually selected on the basis of their biology and function.[95] For studies of myopia, the genes are expected to be expressed in the eye. We can also choose some candidate genes that are expressed in the specific ocular tissue related to the process of myopic development. Examples include genes encoding the components of scleral tissues or enzymes related to scleral thinning or remodelling in myopic eyes;[96,97] genes underlying genetic ocular or systemic diseases usually with high myopia as one of

	Case-controlled study	Family-based association study
Types of samples	Unrelated affected (cases) and unaffected (controls) individuals	Small nuclear families including patients and their parents TDT requires heterozygous parents†
Statistical power	Low for Mendelian diseases High for detecting small genetic effects in complex diseases	Low for Mendelian diseases Lower than case-controlled study for alleles with small genetic effects on the basis of genotyping work
Advantages	Provides information about the effects of genotype (homozygous and heterozygous) on the phenotype or severity of the diseases Easy for sample collection and statistical analysis	Presence of internal control to avoid the problem of population stratification Reducing genotyping errors of affected offspring by tracing genotypes of their parents
Disadvantages	False positive association due to ethnically mismatched cases and controls (population stratification)	Need to ascertain large numbers of nuclear families; recruitment more difficult than case-controlled studies, especially for late-onset diseases

† TDT stands for transmission disequilibrium test

Table 3. Comparison between case-controlled study and family-based association study

the clinical features[95] (for example, Stickler syndrome,[72] Marfan syndrome[73] and juvenile-onset open angle glaucoma[98]) and genes encoding enzymes, binding proteins or receptors involved in the metabolism of retinoic acid, which is an important signalling molecule in the eye and a potential mediator between refractive error and compensatory eye growth.[99–101]

Little is known about the complete functional spectrum in the process of myopia development. A more practical approach is to start with genome-wide linkage scans (parametric and/or non-parametric) for families in which the disease of interest is segregating. Genes that are expressed in the eye and are located within or close to the chromosomal regions showing positive linkage to high myopia can be selected and then investigated by association studies.

Markers and candidate genes in the *MYP2*, *MYP3* and *MYP5* loci have been followed by association studies subsequent to genome-wide linkage scans.[75,77,102–105] The results are contradictory in some instances; an example is the putative association between the *TGIF* gene (a *MYP2* candidate gene) and high myopia.[75,102] In addition to association studies, other types of studies can be carried out to explore candidate chromosomal regions previously identified by genome-wide linkage analysis. Examples of such studies include additional linkage analysis using different families for candidate regions of *MYP2*, *MYP3* and *MYP5* loci,[43,45,106,107] DNA sequence analysis of candidate genes in *MYP2* and *MYP3* regions[108–113] and gene expression studies of *MYP2* candidate genes.[108–110] Overall, these studies provide additional evidence for zooming in a more refined candidate region,[103] supporting/confirming previous findings,[43,45,104,113] questioning/refuting previous finding,[45,106,107] supporting putative candidate genes[75,77] or excluding putative candidate genes.[111,105,108–110,112]

DNA pooling

Candidate-gene approach tends to be piecemeal and less efficient, while genome-wide association studies are very expensive in terms of instrumentation and reagent usage. One alternative efficient strategy is to perform initial genotyping and analysis not on individual DNA samples but on DNA pools constructed by mixing DNA from many individuals.[114] The advantage of working on DNA pools is obvious. In the simplest situation, the relative allele frequencies of SNPs in a set of, say, 400 cases and 400 controls can be estimated in two DNA pools rather than 800 individual samples. This represents a 400-fold increase in efficiency. It acts as an initial screening process and can be used to screen a large number of SNPs from many candidate genes within a relatively short time at an affordable cost. If significant differences are found in the relative allele frequencies between the case pool and the control pool, sample-by-sample genotyping can follow to confirm the initial positive findings. These might result in higher rates of false positive or false negative results because of the inherent measurement errors in the estimation of relative allele frequencies in DNA pools.

There are many studies reporting the statistical analysis of data derived from such approaches.[115–117] Many new modified designs and new statistics, such as for family-based association tests[116] and haplotype estimation,[118,119] have been developed to improve the accuracy and reproducibility of DNA pooling methods. One very exciting development is to use DNA pools in initial genome-wide association scans based on microsatellite or SNP markers; this is very efficient and cost-effective.[120–123] Therefore, DNA pooling is potentially one of the efficient tools for large-scale association studies of complex traits including myopia.

ISSUES FOR CONSIDERATION IN STUDIES OF MYOPIA GENETICS

Animal models: mapping quantitative trait loci

Human genetic studies are inherently very difficult to control in terms of genotype and environmental exposure. Genetic studies on animals by manipulative interbreeding are much easier in planning and execution. The sample size problem can be overcome and replication of studies is also much easier. Many on-line resources of different animal genomes, such as rat and mouse genome resources, are parts of the NCBI databases and are free for the public use. These genome resource centres not only combine the data from different genome sequence projects but also provide a lot of information about the SNP markers and genetic variations within the mouse and the rat genomes and many useful hyperlinked websites. Therefore, gene mapping of complex (quantitative) eye traits in animal models can provide invaluable information on the mechanisms of myopia development. Two loci (*Eye1* and *Eye2*) controlling eyeball size and weight of the mouse were successfully mapped by linkage analysis.[124] Sequence similarity (or homology) between different parts of the genomes of humans, the mouse and the rat is well established. Genes in these candidate regions and expressed in eyes or related to eye growth can be selected as myopia candidate genes for genetic analysis. Recently, the human homologous counterparts of the *Eye1* and *Eye2* candidate genes were found to be associated with high myopia by association studies: the *HGF* gene at chromosome 7q21.1 as human homologue of *Eye1* and the *TGFB1* gene at chromosome 19q13.1 as human homologue of *Eye2*.[67,76] Thus, animal models of myopia can provide important information for mapping myopia genes in humans.

Statistical power

Association studies are more powerful than linkage analysis in detecting genes with small effects, a common situation in complex diseases or traits.[36] False positive association results can be obtained because of confounding by population stratification or as chance findings resulting from failure to handle problems due to multiple comparisons.[125] Therefore, it is very important to replicate positive association results with independent sets of samples, from either unrelated individuals or preferably nuclear families, from the same ethnic groups and from different ethnic groups. Positive results in association tests very often cannot be replicated.[125] One of the possible causes of failure to replicate positive findings in subsequent studies may be the low statistical power to detect moderate and small gene effects.[114] Statistical power and hence sample size are crucial determinants of quality in genetic association studies. In general, statistical power increases with larger sample size.

By definition, the power of a statistical test is defined as the probability that the test will correctly reject the null hypothesis when it is false. In other words, it refers to the ability of a genetic association study to demonstrate an association between a genetic marker and the disease under study if one exists. In addition to the threshold level of significance required (classically set at 0.05), the power of an association study depends on the following factors: the effect size of the functional genetic variation of the disease gene (or simply disease allele), the frequency of the disease allele (usually unknown at the start of a study), the frequency of the marker allele (to be tested in a study) and the LD between the disease allele and the marker allele.[126,127] For a given sample size, the power of an association study is higher if:

1. the effect size of the disease allele is higher
2. the difference between the disease allele frequency and the marker allele frequency is smaller
3. the disease allele and the marker allele are on the same chromosome (instead of on different homologous chromosomes) and/or
4. the LD between the disease allele and the marker allele is higher.

On the contrary, for a given magnitude of power, the sample size required for an association study is smaller under similar conditions.

Typically, the effect size of the disease allele is small for genes involved in complex diseases. The size of the genetic effect is usually indicated by the odds ratio (OR) for case-controlled association studies and the genotype relative risk (GRR) for family-based association studies. OR provides a valid estimate of the relative risk in case-controlled studies. It is the ratio of the odds of exposure to the susceptible genetic variant (allele or genotype) among the cases (disease group) to that among the controls.[35,126] GRR describes the increased chance of having the disease for an individual with one genotype (predisposing genotype) over the risk if he or she carries the other allele.[82] An effect size (OR or GRR) greater than 1 indicates an increased risk, while an effect size between 0 and 1 implies a protective effect.

An illustrative example showing the relationship between genetic effect size and sample size is the study of the association between angiotension-converting enzyme insertion/deletion polymorphisms and myocardial infarction. Initial studies had small to moderate sample sizes and gave ORs ranging from less than 1 to about 3.[128] The largest single study to date gave a very small OR of 1.10.[129] Note that such a small effect size was found statistically significant only with a very large sample size: nearly 5,000 cases and 6,000 controls. Attempts to replicate positive findings should estimate the effect size and the required corresponding sample size for following studies to achieve the statistical power for detecting an association. The same principle of sample size requirement applies to both OR for case-controlled studies and GRR for family-based association studies. In reality, it is very challenging to recruit a large sample size for high myopia, especially nuclear families, with stringent refractive criteria. Collaborations of multiple eye centres and institutes should facilitate the collection of a huge number of subjects for association studies.

Heterogeneity

Different results were found among different studies of linkage analysis

(Table 1) or association studies for myopia.[41–58,66,67,75–77,87,102–113] Some genetic variations are specific to certain ethnic groups and this may partially account for the heterogeneity of complex disorders. The population-specific SNPs are particularly important in association studies and relevant in context that myopia is much more prevalent in Chinese than in Caucasians.[130] The importance of ethnic-specific SNPs can best be illustrated by a functional coding SNP (Q506R) found in the Factor V gene. This coding SNP, with an allele frequency of three to seven per cent in Caucasians, accounts for a large proportion of thrombosis in these populations[131] but this coding SNP is not detected in Oriental populations, in which thrombotic disorders are less common. A recent study shows that common genetic variations contribute very much to differences observed in different ethnic groups in the gene expression levels of many genes and suggests that allele frequency differences at regulatory polymorphisms may also account for differences in prevalence of complex diseases in some populations.[132] Thus, replication of association studies in Caucasian populations may help to explain the differences in myopia prevalence with Chinese, if the SNP alleles are much more frequently found in Chinese. If the association of myopia with certain genetic variants can be reliably replicated, it is worth further investigating the functional effects of the genetic variants on the gene, such as the level of gene expression.[133] Stringent entry criteria may make subject recruitment very difficult in Caucasian populations because of lower prevalence of myopia and the probable lower allele frequencies of predisposing genetic variants.

Gene-gene and gene-environment interactions

Neither genes nor environmental factors work alone: they interact with one another to give a final outcome—myopia in this case. The responses to environmental factors are expected to vary as a function of the underlying genetic make-up, a phenomenon known as gene-environment interaction.[134] The effects of a particular genotype at a certain locus may also vary with the genotype at another locus, a situation termed gene-gene interaction. Environmental risk factors associated with or contributing to myopic development have been studied extensively.[135] Genetic factors contributing to myopia are also areas of active research. Myopia is a complex eye trait that is affected by both genetic and environmental factors, and their interactions. How genetic factors interact with each other and with environmental factors to produce myopia is not known. To realise the benefits of personalised medicine for prevention and treatment of myopia, it is both vital and challenging to investigate and understand such interactions, which can either enhance or reduce the risk of developing myopia. Quantification of such interactions requires stratified and multivariate analyses of studies with very large sample size. In recent years, many statistical and bioinformatic tools have been developed to address the issues of these gene-gene and gene-environment interactions.[136–139]

CONCLUSIONS

Current genotyping techniques and statistics in genetics enable researchers to identify myopia susceptibility genes even without any prior knowledge of the pathophysiological mechanisms responsible for myopia. Many myopia susceptibility loci have been mapped successfully by linkage analysis. The databases of the NCBI and the HapMap Project provide rich information of genetic variations, which act as genetic markers for linkage analysis and association studies and expedite the mapping of myopia loci. The complex nature of myopia challenges the measurement and analysis of multiple genetic and environmental components and their interactions. Attention should also be paid to the interpretation of multiple and conflicting findings of association studies. The replication of myopia genetic studies is essential for confirmation of results and before further functional investigations. Development of analytical methods based on more sophisticated models of complex disease is required for gene-gene and gene-environment interactions that should become a trend in the near future. The researcher can focus on specific candidate genes that may be associated with myopia, or even look across the entire genome to find chromosomal regions that may be associated with this complex eye trait/disease.

ACKNOWLEDGEMENT

We thank Dr Andrew Lam for commenting on the manuscript from a non-geneticist's viewpoint.

GRANTS AND FINANCIAL SUPPORT

Our work on the genetics of myopia is supported by grants from The Hong Kong Polytechnic University (PolyU), Hang Seng Bank Jubilee Endowment Fund and Research Grants Council (Competitive Earmarked Research Grant). Grants from PolyU include support from the Area of Strategic Development (ASD) for Centre for Myopia Research, the Dean's Reserve (Faculty of Health and Social Sciences) and a Central Research Grant.

INFORMATION ON ELECTRONIC DATABASES AND WEBPAGES

The following databases and webpages are mentioned and/or discussed in the article.

Comparative Maps at the NCBI: http://www.ncbi.nlm.nih.gov/projects/Homology/

Human Genome Project: http://www.ornl.gov/TechResources/Human_Genome/home.html

HUGO Gene Nomenclature Committee: http://www.gene.ucl.ac.uk/nomenclature/

International HapMap Project: http://www.hapmap.org/index.html.en

Mouse Genome Resources at the NCBI: http://www.ncbi.nlm.nih.gov/genome/guide/mouse/

OMIM entry for *MYP2* locus (with hyperlinks to other myopia loci): http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=160700

Online Mendelian Inheritance in Man (OMIM) database at the NCBI: http://

www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Rat Genome Resources at the NCBI: http://www.ncbi.nlm.nih.gov/genome/guide/rat/index.html

Single Nucleotide Polymorphism database (dbSNP) at the NCBI: http://www.ncbi.nlm.nih.gov/SNP/

Web resources of genetic linkage analysis at Rockefeller University: http://linkage.rockefeller.edu/

REFERENCES

1. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991; 7: 403–422.
2. Wang Q, Klein BE, Klein R, Moss SE. Refractive status in the Beaver Dam Eye Study. *Invest Ophthalmol Vis Sci* 1994; 35: 4344–4347.
3. Katz J, Tielsch JM, Sommer A. Prevalence and risk factors for refractive errors in an adult inner city population. *Invest Ophthalmol Vis Sci* 1997; 38: 334–340.
4. Attebo K, Ivers RQ, Mitchell P. Refractive errors in an older population: the Blue Mountains Eye Study. *Ophthalmology* 1999; 106: 1066–1072.
5. Wensor M, McCarty CA, Taylor HR. Prevalence and risk factors of myopia in Victoria, Australia. *Arch Ophthalmol* 1999; 117: 658–663.
6. Munoz B, West SK, Rodriguez J, Sanchez R, Broman AT, Snyder R, Klein R. Blindness, visual impairment and the problem of uncorrected refractive error in a Mexican-American population: Proyecto VER. *Invest Ophthalmol Vis Sci* 2002; 43: 608–614.
7. Kempen JH, Mitchell P, Lee KE, Tielsch JM, Broman AT, Taylor HR, Ikram MK, Congdon NG, O'Colmain BJ. The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. *Arch Ophthalmol* 2004; 122: 495–505.
8. Goh WS, Lam CS. Changes in refractive trends and optical components of Hong Kong Chinese aged 19–39 years. *Ophthalmic Physiol Opt* 1994; 14: 378–382.
9. Lin LL, Shih YF, Lee YC, Hung PT, Hou PK. Changes in ocular refraction and its components among medical students: a 5-year longitudinal study. *Optom Vis Sci* 1996; 73: 495–498.
10. Lam CS, Goldschmidt E, Edwards MH. Prevalence of myopia in local and international schools in Hong Kong. *Optom Vis Sci* 2004; 81: 317–322.
11. Woo WW, Lim KA, Yang H, Lim XY, Liew F, Lee YS, Saw SM. Refractive errors in medical students in Singapore. *Singapore Med J* 2004; 45: 470–474.
12. Wildsoet CF. Structural correlates of myopia. In: Rosenfield M, Gillmartin B, eds. Myopia & Nearwork. Oxford: Butterworth Heinemann, 1998. p 31–56.
13. Wallman J. Nature and nurture of myopia. *Nature* 1994; 371: 201–202.
14. Mutti DO, Zadnik K, Adams AJ. Myopia. The nature versus nurture debate goes on. *Invest Ophthalmol Vis Sci* 1996; 37: 952–957.
15. Curtin BJ. The Myopias: Basic Science and Clinical Management. New York: Harper & Row, 1985.
16. Goss DA, Hampton MJ, Wickham MG. Selected review on genetic factors in myopia. *J Am Optom Assoc* 1988; 59: 875–884.
17. Yap M, Wu M, Liu ZM, Lee FL, Wang SH. Role of heredity in the genesis of myopia. *Ophthalmic Physiol Opt* 1993; 13: 316–319.
18. Zadnik K, Satariano WA, Mutti DO, Sholtz RI, Adams AJ. The effect of parental history of myopia on children's eye size. *JAMA* 1994; 271: 1323–1327.
19. Zadnik K, Mutti DO. Myopia—a challenge to optometric research. *J Am Optom Assoc* 1995; 66: 145–146.
20. The Framingham Offspring Eye Study Group. Familial aggregation and prevalence of myopia in the Framingham Offspring Eye Study. *Arch Ophthalmol* 1996; 114: 326–332.
21. Hu DN. Twin study on myopia. *Chin Med J (Engl)* 1981; 94: 51–55.
22. Teikari JM. Myopia and stature. *Acta Ophthalmol Scand* 1987; 65: 673–676.
23. Lin LL, Chen CJ. A twin study on myopia in Chinese school children. *Acta Ophthalmol Scand Suppl* 1988; 185: 51–53.
24. Hammond CJ, Snieder H, Gilbert CE, Spector TD. Genes and environment in refractive error: the twin eye study. *Invest Ophthalmol Vis Sci* 2001; 42: 1232–1236.
25. Lyhne N, Sjolie AK, Kyvik KO, Green A. The importance of genes and environment for ocular refraction and its determiners: a population based study among 20–45-year-old twins. *Br J Ophthalmol* 2001; 85: 1470–1476.
26. Dirani M, Chamberlain M, Shekar SN, Islam AF, Garoufalis P, Chen CY, Guymer RH, Baird PN. Heritability of refractive error and ocular biometrics: the Genes in Myopia (GEM) twin study. *Invest Ophthalmol Vis Sci* 2006; 47: 4756–4761.
27. Goss DA, Jackson TW. Clinical findings before the onset of myopia in youth: 4. Parental history of myopia. *Optom Vis Sci* 1996; 73: 279–282.
28. Liang CL, Yen E, Su JY, Liu C, Chang TY, Park N, Wu MJ, Lee S, Flynn JT, Juo SH. Impact of family history of high myopia on level and onset of myopia. *Invest Ophthalmol Vis Sci* 2004; 45: 3446–3452.
29. Alsbirk PH. Anterior chamber of the eye. A genetic and anthropological study in Greenland Eskimos. *Hum Hered* 1975; 25: 418–427.
30. Alsbirk PH. Variation and heritability of ocular dimensions. A population study among adult Greenland Eskimos. *Acta Ophthalmol Scand* 1977; 55: 443–456.
31. Alsbirk PH. Refraction in adult West Greenland Eskimos. A population study of spherical refractive errors, including oculometric and familial correlations. *Acta Ophthalmol Scand* 1979; 57: 84–95.
32. Biino G, Palmas MA, Corona C, Prodi D, Fanciulli M, Sulis R, Serra A, Fossarello M, Pirastu M. Ocular refraction: heritability and genome-wide search for eye morphometry traits in an isolated Sardinian population. *Hum Genet* 2005; 116: 152–159.
33. Chen CY, Scurrah KJ, Stankovich J, Garoufalis P, Dirani M, Pertile KK, Richardson AJ, Mitchell P, Baird PN. Heritability and shared environment estimates for myopia and associated ocular biometric traits: the Genes in Myopia (GEM) family study. *Hum Genet* 2007; 121: 511–520.
34. Sorsby A, Leary GA. A longitudinal study of refraction and its components during growth. *Spec Rep Ser Med Res Counc (GB)* 1969; 309: 1–41.
35. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; 265: 2037–2048.
36. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273: 1516–1517.
37. Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997; 278: 1580–1581.
38. Chakravarti A. It's raining SNPs, hallelujah? *Nat Genet* 1998; 19: 216–217.
39. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES, *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; 280: 1077–1082.
40. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman

B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409: 928–933.

41. Schwartz M, Haim M, Skarsholm D. X-linked myopia: Bornholm eye disease. Linkage to DNA markers on the distal part of Xq. *Clin Genet* 1990; 38: 281–286.
42. Young TL, Ronan SM, Drahozal LA, Wildenberg SC, Alvear AB, Oetting WS, Atwood LD, Wilkin DJ, King RA. Evidence that a locus for familial high myopia maps to chromosome 18p. *Am J Hum Genet* 1998; 63: 109–119.
43. Lam DS, Tam PO, Fan DS, Baum L, Leung YF, Pang CP. Familial high myopia linkage to chromosome 18p. *Ophthalmologica* 2003; 217: 115–118.
44. Young TL, Ronan SM, Alvear AB, Wildenberg SC, Oetting WS, Atwood LD, Wilkin DJ, King RA. A second locus for familial high myopia maps to chromosome 12q. *Am J Hum Genet* 1998; 63: 1419–1424.
45. Farbrother JE, Kirov G, Owen MJ, Pong-Wong R, Haley CS, Guggenheim JA. Linkage analysis of the genetic loci for high myopia on 18p, 12q, and 17q in 51 U.K. Families. *Invest Ophthalmol Vis Sci* 2004; 45: 2879–2885.
46. Naiglin L, Gazagne C, Dallongeville F, Thalamas C, Idder A, Rascol O, Malecaze F, Calvas P. A genome wide scan for familial high myopia suggests a novel locus on chromosome 7q36. *J Med Genet* 2002; 39: 118–124.
47. Paluru P, Ronan SM, Heon E, Devoto M, Wildenberg SC, Scavello G, Holleschau A, Makitie O, Cole WG, King RA, Young TL. New locus for autosomal dominant high myopia maps to the long arm of chromosome 17. *Invest Ophthalmol Vis Sci* 2003; 44: 1830–1836.
48. Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M, Holmes T, Ciner E, Bailey-Wilson JE. Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. *Am J Hum Genet* 2004; 75: 448–459.
49. Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M, Holmes TN, Ciner E, Bailey-Wilson JE. Genome-wide scan of additional Jewish families confirms linkage of a myopia susceptibility locus to chromosome 22q12. *Mol Vis* 2006; 12: 1499–1505.
50. Klein AP, Duggal P, Lee KE, Klein R, Bailey-Wilson JE, Klein BE. Confirmation of linkage to ocular refraction on chromosome 22q and identification of a novel linkage region on 1q. *Arch Ophthalmol* 2007; 125: 80–85.
51. Hammond CJ, Andrew T, Tat Mak Y, Spector TD. A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: a genomewide scan of dizygotic twins. *Am J Hum Genet* 2004; 75: 294–304.
52. Stambolian D, Ciner EB, Reider LC, Moy C, Dana D, Owens R, Schlifka M, Holmes T, Ibay G, Bailey-Wilson JE. Genome-wide scan for myopia in the Old Order Amish. *Am J Ophthalmol* 2005; 140: 469–476.
53. Zhang Q, Guo X, Xiao X, Jia X, Li S, Hejtmancik JF. A new locus for autosomal dominant high myopia maps to 4q22-q27 between D4S1578 and D4S1612. *Mol Vis* 2005; 11: 554–560.
54. Paluru PC, Nallasamy S, Devoto M, Rappaport EF, Young TL. Identification of a novel locus on 2q for autosomal dominant high-grade myopia. *Invest Ophthalmol Vis Sci* 2005; 46: 2300–2307.
55. Zhang Q, Guo X, Xiao X, Jia X, Li S, Hejtmancik JF. Novel locus for X linked recessive high myopia maps to Xq23-q25 but outside MYP1. *J Med Genet* 2006; 43: e20.
56. Zhang Q, Li S, Xiao X, Jia X, Guo X. Confirmation of a genetic locus for X-linked recessive high myopia outside MYP1. *J Hum Genet* 2007; 52: 469–472.
57. Wojciechowski R, Moy C, Ciner E, Ibay G, Reider L, Bailey-Wilson JE, Stambolian D. Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36. *Hum Genet* 2006; 119: 389–399.
58. Nallasamy S, Paluru PC, Devoto M, Wasserman NF, Zhou J, Young TL. Genetic linkage study of high-grade myopia in a Hutterite population from South Dakota. *Mol Vis* 2007;13: 229–236.
59. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001; 2: 930–942.
60. Syvanen AC, Taylor GR. Approaches for analyzing human mutations and nucleotide sequence variation: a report from the Seventh International Mutation Detection meeting, 2003. *Hum Mutat* 2004; 23: 401–405.
61. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* 2003; 3: 77–96.
62. Majewski J, Ott J. Distribution and characterization of regulatory elements in the human genome. *Genome Res* 2002; 12: 1827–1836.
63. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001; 27: 234–236.
64. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; 380: 152–154.
65. Shastry BS. SNP alleles in human disease and evolution. *J Hum Genet* 2002; 47: 561–566.
66. Leung YF, Tam PO, Baum L, Lam DS, Pang CC. TIGR/MYOC proximal promoter GT-repeat polymorphism is not associated with myopia. *Hum Mutat* 2000; 16: 533.
67. Han W, Yap MK, Wang J, Yip SP. Family-based association analysis of hepatocyte growth factor (HGF) gene polymorphisms in high myopia. *Invest Ophthalmol Vis Sci* 2006; 47: 2291–2299.
68. Hashemi H, Hatef E, Fotouhi A, Mohammad K. Astigmatism and its determinants in the Tehran population: the Tehran eye study. *Ophthalmic Epidemiol* 2005; 12: 373–381.
69. Harvey EM, Dobson V, Miller JM. Prevalence of high astigmatism, eyeglass wear, and poor visual acuity among Native American grade school children. *Optom Vis Sci* 2006; 83: 206–212.
70. Saw SM, Carkeet A, Chia KS, Stone RA, Tan DT. Component dependent risk factors for ocular parameters in Singapore Chinese children. *Ophthalmology* 2002; 109: 2065–2071.
71. Yates JR, Connor JM. Genetic linkage. *Br J Hosp Med* 1986; 36: 133–136.
72. O'Brien DA, Phillips AJ. Stickler syndrome. *Clin Exp Optom* 2000; 83: 330–332.
73. Byers PH. Determination of the molecular basis of Marfan syndrome: a growth industry. *J Clin Invest* 2004; 114: 161–163.
74. The international HapMap Consortium. The International HapMap Project. *Nature* 2003; 426: 789–796.
75. Lam DS, Lee WS, Leung YF, Tam PO, Fan DS, Fan BJ, Pang CP. TGFbeta-induced factor: a candidate gene for high myopia. *Invest Ophthalmol Vis Sci* 2003; 44: 1012–1015.
76. Lin HJ, Wan L, Tsai Y, Tsai YY, Fan SS, Tsai CH, Tsai FJ. The TGFbeta 1 gene codon 10 polymorphism contributes to the genetic predisposition to high myopia. *Mol Vis* 2006; 12: 698–703.
77. Wang IJ, Chiang TH, Shih YF, Hsiao CK, Lu SC, Hou YC, Lin LL. The association of single nucleotide polymorphisms in the 5'-regulatory region of the lumican gene with susceptibility to high myopia in Taiwan. *Mol Vis* 2006; 12: 852–857.
78. Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci U S A* 1998; 95: 11389–11393.

79. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002; 3: 299–309.
80. Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury M, Lau J, O'Brien TR, Rothman N, Stroup D, Taioli E, Thomas D, Vainio H, Wacholder S, Weinberg C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. *Am J Epidemiol* 2002; 156: 300–310.
81. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev* 2002; 11: 505–512.
82. Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 1995; 268: 1584–1589.
83. Zhang H, Risch N. Mapping quantitative-trait loci in humans by use of extreme concordant sib pairs: selected sampling by parental phenotypes. *Am J Hum Genet* 1996; 59: 951–957.
84. Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet* 2001; 68: 1463–1474.
85. Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996; 59: 983–989.
86. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; 52: 506–516.
87. Tang WC, Yip SP, Lo KK, Ng PW, Choi PS, Lee SY, Yap MK. Linkage and association of myocilin (MYOC) polymorphisms with high myopia in a Chinese population. *Mol Vis* 2007; 13: 534–544.
88. Syvanen AC. Toward genome-wide SNP genotyping. *Nat Genet* 2005; 37 Suppl: S5–S10.
89. Craig DW, Stephan DA. Applications of whole-genome high-density SNP genotyping. *Expert Rev Mol Diagn* 2005; 5: 159–170.
90. Gunderson KL, Kuhn KM, Steemers FJ, Ng P, Murray SS, Shen R. Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics* 2006; 7: 641–648.
91. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; 6: 95–108.
92. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; 6: 109–118.
93. Farrall M, Morris AP. Gearing up for genome-wide gene-association studies. *Hum Mol Genet* 2005; 14 Spec No. 2: R157–R162.
94. Christensen K, Murray JC. What genome-wide association studies can do for medicine. *N Engl J Med* 2007; 356: 1094–1097.
95. Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet* 2005; 366: 1315–1323.
96. McBrien NA, Gentle A. Role of the sclera in the development and pathological complications of myopia. *Prog Retin Eye Res* 2003; 22: 307–338.
97. Rada JA, Shelton S, Norton TT. The sclera and myopia. *Exp Eye Res* 2006; 82: 185–200.
98. Grodum K, Heijl A, Bengtsson B. Refractive error and glaucoma. *Acta Ophthalmol Scand* 2001; 79: 560–566.
99. Seko Y, Shimizu M, Tokoro T. Retinoic acid increases in the retina of the chick with form deprivation myopia. *Ophthalmic Res* 1998; 30: 361–367.
100. Bitzer M, Feldkaemper M, Schaeffel F. Visually induced changes in components of the retinoic acid system in fundal layers of the chick. *Exp Eye Res* 2000; 70: 97–106.
101. Mertz JR, Wallman J. Choroidal retinoic acid synthesis: a possible mediator between refractive error and compensatory eye growth. *Exp Eye Res* 2000; 70: 519–527.
102. Hasumi Y, Inoko H, Mano S, Ota M, Okada E, Kulski JK, Nishizaki R, Mok J, Oka A, Kumagai N, Nishida T, Ohno S, Mizuki N. Analysis of single nucleotide polymorphisms at 13 loci within the transforming growth factor-induced factor gene shows no association with high myopia in Japanese subjects. *Immunogenetics* 2006; 58: 947–953.
103. Dewan AT, Alvear AB, Peterson J, Holleschau A, King RA. Further refinement of the MYP2 locus for autosomal dominant high myopia by linkage disequilibrium analysis. *Ophthalmic Genet* 2001; 22: 69–75.
104. Heath S, Robledo R, Beggs W, Feola G, Parodo C, Rinaldi A, Contu L, Dana D, Stambolian D, Siniscalco M. A novel approach to search for identity by descent in small samples of patients and controls from the same Mendelian breeding unit: a pilot study on myopia. *Hum Hered* 2001; 52: 183–190.
105. Liang CL, Hung KS, Tsai YY, Chang W, Wang HS, Juo SH. Systematic assessment of the tagging polymorphisms of the COL1A1 gene for high myopia. *J Hum Genet* 2007; 52: 374–377.
106. Mutti DO, Semina E, Marazita M, Cooper M, Murray JC, Zadnik K. Genetic loci for pathological myopia are not associated with juvenile myopia. *Am J Med Genet* 2002; 112: 355–360.
107. Ibay G, Doan B, Reider L, Dana D, Schlifka M, Hu H, Holmes T, O'Neill J, Owens R, Ciner E, Bailey-Wilson JE, Stambolian D. Candidate high myopia loci on chromosomes 18p and 12q do not play a major role in susceptibility to common myopia. *BMC Med Genet* 2004; 5: 20.
108. Scavello GS, Paluru PC, Ganter WR, Young TL. Sequence variants in the transforming growth beta-induced factor (TGIF) gene are not associated with high myopia. *Invest Ophthalmol Vis Sci* 2004; 45: 2091–2097.
109. Zhou J, Young TL. Evaluation of Lipin 2 as a candidate gene for autosomal dominant 1 high-grade myopia. *Gene* 2005; 352: 10–19.
110. Scavello GS Jr, Paluru PC, Zhou J, White PS, Rappaport EF, Young TL. Genomic structure and organization of the high grade Myopia-2 locus (MYP2) critical region: mutation screening of 9 positional candidate genes. *Mol Vis* 2005; 11: 97–110.
111. Clarke LA, Rogers KM, Wagner SH, Del-Bono EA, Haines JL, Gwiazda JE, Wiggs JL. Evaluation of small leucine-rich proteoglycans located on chromosome 12q21 in families with juvenile-onset myopia. *Invest Ophthalmol Vis Sci* 2004; 45: E-Abstract 1245.
112. Paluru PC, Scavello GS, Ganter WR, Young TL. Exclusion of lumican and fibromodulin as candidate genes in MYP3 linked high grade myopia. *Mol Vis* 2004; 10: 917–922.
113. Majava M, Bishop PN, Hagg P, Scott PG, Rice A, Inglehearn C, Hammond CJ, Spector TD, Ala-Kokko L, Mannikko M. Novel mutations in the small leucine-rich repeat protein/proteoglycan (SLRP) genes in high myopia. *Hum Mutat* 2007; 28: 336–344.
114. Sham P, Bader JS, Craig I, O'Donovan M, Owen M. DNA Pooling: a tool for large-scale association studies. *Nat Rev Genet* 2002; 3: 862–871.
115. Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am J Hum Genet* 1997; 61: 734–747.
116. Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998; 8: 1273–1288.
117. Teng J, Risch N. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. *Genome Res* 1999; 9: 234–241.
118. Ito T, Chiku S, Inoue E, Tomita M, Morisaki T, Morisaki H, Kamatani N. Esti-

mation of haplotype frequencies, linkage-disequilibrium measures, and combination of haplotype copies in each pool by use of pooled DNA data. *Am J Hum Genet* 2003; 72: 384–398.
119. Wang S, Kidd KK, Zhao H. On the use of DNA pooling to estimate haplotype frequencies. *Genet Epidemiol* 2003; 24: 74–82.
120. Sawcer S, Maranian M, Setakis E, Curwen V, Akesson E, Hensiek A, Coraddu F, Roxburgh R, Sawcer D, Gray J, Deans J, Goodfellow PN, Walker N, Clayton D, Compston A. A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain* 2002; 125: 1337–1347.
121. Butcher LM, Meaburn E, Liu L, Fernandes C, Hill L, Al-Chalabi A, Plomin R, Schalkwyk L, Craig IW. Genotyping pooled DNA on microarrays: a systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. *Behav Genet* 2004; 34: 549–555.
122. Meaburn E, Butcher LM, Schalkwyk LC, Plomin R. Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans. *Nucleic Acids Res* 2006; 34: e27.
123. Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR. Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. *Genes Immun* 2007; 8: 57–68.
124. Zhou G, Williams RW. Eye1 and Eye2: gene loci that modulate eye size, lens weight, and retinal area in the mouse. *Invest Ophthalmol Vis Sci* 1999; 40: 817–825.
125. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; 361: 865–872.
126. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004; 5: 89–100.
127. Gordon D, Finch SJ. Factors affecting statistical power in the detection of genetic association. *J Clin Invest* 2005; 115: 1408–1418.
128. Agerholm-Larsen B, Nordestgaard BG, Tybjaerg-Hansen A. ACE gene polymorphism in cardiovascular disease: meta-analyses of small and large studies in whites. *Arterioscler Thromb Vasc Biol* 2000; 20: 484–492.
129. Keavney B, McKenzie C, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R. Large-scale test of hypothesised associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. *Lancet* 2000; 355: 434–442.
130. Wilson A, Woo G. A review of the prevalence and causes of myopia. *Singapore Med J* 1989; 30: 479–484.
131. Rees DC, Cox M, Clegg JB. World distribution of factor V Leiden. *Lancet* 1995; 346: 1133–1134.
132. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007; 39: 226–231.
133. Knight JC. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond)* 2003; 104: 493–501.
134. Grigorenko EL. The inherent complexities of gene-environment interactions. *J Gerontol B Psychol Sci Soc Sci* 2005; 60 Spec No 1: 53–64.
135. Saw SM. A synopsis of the prevalence rates and environmental risk factors for myopia. *Clin Exp Optom* 2003; 86: 289–294.
136. Norris JM, Selinger-Leneman H, Genin E. Investigation of a candidate gene, environment, and G x E interaction using case-control and case-parent study designs. *Genet Epidemiol* 2001; 21 Suppl 1: S843–S848.
137. Liu X, Fallin MD, Kao WH. Genetic dissection methods: designs used for tests of gene-environment interaction. *Curr Opin Genet Dev* 2004; 14: 241–245.
138. Ritchie MD. Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease. *Neurosurg Focus* 2005; 19: E2.
139. Wallace HM. A model of gene-gene and gene-environment interactions and its implications for targeting environmental interventions by genotype. *Theor Biol Med Model* 2006; 3: 35.

Corresponding author:
Professor Shea Ping Yip
Department of Health Technology and Informatics
Hung Hom, Kowloon
Hong Kong SAR
CHINA
E-mail: shea.ping.yip@polyu.edu.hk