Strategies for mapping susceptibility genes in age-related maculopathy

ALAN F. WRIGHT

Abstract

The prospects for mapping genes conferring susceptibility to common disorders such as age-related maculopathy have increased as a result of advances in high-throughput genotyping using single nucleotide polymorphisms. This should not obscure the magnitude of the task of identifying what are potentially a diverse group of genetic variants with small or intermediate effects on disease susceptibility. Strategies for increasing the odds of success in such an endeavour are discussed.

Age-related maculopathy (ARM) is the commonest blinding disease in the industrial world, with a prevalence that increases with age reaching almost 1 in 3 over the age of 70 years.[1,2] It is a genetically complex disorder with an approximately three-fold increased risk to the siblings of affected individuals compared with the general population (reviewed by Yates and Moore[3] and Gorin[4]). This provides indirect evidence for genetic factors in disease susceptibility, although the genetic risk may differ according to disease severity[5] and shared familial environment cannot be discounted as a contributor to familial clustering. Twin studies provide further support for a genetic component in disease susceptibility, since the concordance rate for disease in identical twins is greater than for non-identical twins.[6–8] When a shared environmental effect is likely to contribute to familial clustering, a comparison of risks in adopted and full sibs, full and half-sibs and twins reared together or apart, can be helpful, but these have not been reported in ARM or in age-related macular degeneration (AMD). The environmental component has to be considered a potentially significant contributor to familial clustering in ARM/AMD since at least one environmental variable, smoking, is a well-established risk factor,[9,10] and others have been proposed.[3] Despite these caveats, the extent of familial clustering and twin concordance data provide reasonable but indirect evidence of genetic factors in susceptibility to ARM and AMD.

Most traits, whether in plants or humans, show inheritance patterns that are more consistent with blending than discrete inheritance. Indeed, Gregor Mendel was lucky to discover the discrete nature of inheritance, since only a minority of traits in peas show monogenic inheritance.[11] What genetic models are available to explain the more complex patterns of multigenic inheritance? Simple additive models of interaction between alleles at a small number of genes are readily shown to produce a continuously varying trait distribution and an inheritance pattern consistent with many complex traits.[12] For discrete traits such as disease, a 'liability model' is helpful, in which a continuous distribution of liability is assumed, which is determined by genetic, environmental and interaction influences.[13] Disease supervenes with increasing likelihood once a threshold is reached. In this model, disease risks are non-additive even though the interactions between genes are additive.[12] This is consistent with what is observed in many complex disorders, where disease risk declines exponentially with decreasing degree of relationship to the proband (compared with a linear decline under a monogenic model). This implies non-additive interactions between the genetic and/or environmental components of risk.

Liability to common diseases, such as ARM, is therefore presumed to be determined by a large number of interacting risk factors, some influenced by environment, others by genetic variation. Many of the major risk factors in ARM remain to be identified or confirmed,[3] but one could speculate that they will include a diverse set of influences such as age, antioxidant capacity, smoking history, family history, cardiovascular factors, diet, ethnic group, basement membrane and extracellular matrix factors, light exposure, lipofuscin turnover and ApoE genotype. The 'iceberg model' is a simple way of envisaging the disease process, four-fifths of which lies 'below the water' and is clinically invisible, and represents the cumulative interactions of causal risk factors. Clinical disease is a small and late stage of the disease process, but is all that can be

A.F. Wright ✉
MRC Human Genetics Unit
Western General Hospital
Edinburgh EH4 2XU, UK

Table 1. *Approximate prevalence of common disorders, including age-related maculopathy (ARM), in the United States of America*

Disorder	Approximate prevalence in USA
Cardiovascular disease	58 million
Diabetes	16 million
Cataract	17 million
Osteoporosis	15–20 million
Cancer	8 million
ARM	2–15 million
Glaucoma	3 million
Alzheimer disease	4 million
Parkinson disease	1.5 million
Retinitis pigmentosa	0.07 million

Source: http://www.cdc.gov/health; http://www.nei.nich/gov

seen 'above the water'. Different individuals with ARM/AMD are likely to have different combinations of risk factors, each operating over a lifetime, making the study of patient groups with ARM or AMD, rather than the subclinical components of the disease process, a complex and difficult task.

How many genes influence a complex trait such as ARM or AMD? Evidence from experimental organisms suggests that variants at a large number of genes have small marginal effects on the trait (i.e. average effects exerted by all variants at a genetic locus), while a small number of genes exert large marginal effects, including those causing rare monogenic forms of the disease.[14] The latter tend to be caricatures rather than models of the 'garden variety' of disease, but their identification is a very important means of breaking down the genetic components of risk. Genes with small marginal effects tend to be strongly context-dependent, while those exerting large marginal effects tend to be largely uninfluenced by the genetic and environmental context in which they act.[14]

One way of estimating the number of disease genes in ARM is to extrapolate from what is known about genetically simpler retinal disorders such as retinitis pigmentosa (RP). Sampling calculations can be used to estimate the number of genes causing the autosomal recessive form of RP, which shows that over 100 genes are predicted to cause this subgroup alone (A.D. Carothers and A.F. Wright, unpublished). The genetic diversity in RP is indeed staggering, although in two genetic subsets of autosomal recessive RP, Bardet Biedel syndrome and Usher syndrome, the number of causal genes is reduced to between six and eight.[15] These considerations suggest that there may be a rough correlation between the number of genetic loci and physiological complexity. However, it is not clear where ARM lies in the continuum of physiological complexity. It is probably high in the scale, bearing in mind the diversity of known and potential risk factors referred to above. In short, there may a hundred or more genes contributing to risk in diseases such as ARM or AMD. The goals are to identify those with large marginal effects on disease risk, or to identify variants with smaller marginal effects but which are common in the population, and so contribute disproportionately to the

population attributable risk (i.e. the extent to which disease prevalence declines if these are hypothetically eliminated from the population).[16] Both goals offer a means of identifying novel mechanisms or pathways which could lead to new therapeutic approaches.

The approximate population prevalence of various common disorders in the United States population is shown in Table 1, where it can be seen that ARM affects almost two orders of magnitude more people than the monogenic disorder RP. Why is it so common? The most obvious answer is that it is a disorder of ageing, since disease rates increase exponentially with age. It follows from this that ARM is genetically influenced, to the extent that ageing reflects our innate ability to combat the entropic effects of age, a genetic property of each species. This does not imply that genetic differences between individuals account for disease variability in the population. One study of ageing in Bruch's membrane suggests that if we were all to live to age 120 years, we would all have ARM.[17] The ageing process is genetically influenced[18] but there are striking differences in the rates of ageing between tissues and it is not clear to what extent retinal ageing involves distinct processes of repair and damage limitation. There are examples of single genes which influence both generalised and tissue-restricted ageing processes[19] and it is possible that similar genes will be found to influence ARM/AMD. Evolutionary considerations suggest that generalised ageing represents the loss of selective constraints at many loci of small individual effect.[20] Tissue-specific ageing, and perhaps ARM is a good example, may represent a narrower spectrum of genetic influences.

What other genetic models might explain the high population prevalence of ARM? Two contrasting views are, first, that common or polymorphic genetic variants (i.e. allele frequency > 1% in the population) at a limited number of genes are the major source of genetic risk.[21] If this is true, their identification is important since they will account for a large proportion of the population attributable risk. Examples of such common variants affecting disease risk are shown in Table 2, but it is unclear whether these just represent the few cherries that happen to have fallen from the tree, while the majority remain well out of reach.[21] The allele frequencies of such variants tend to vary in a gradual fashion across ethnic groups, which may go some way towards explaining ethnic differences in disease prevalence.

Table 2. *Polymorphic genetic variants influencing susceptibility to common disorders (see Online Mendelian Inheritance in Man[15])*

Disorder	Common susceptibility variants
Alzheimer disease	ApoE alleles (ε4, ε2)
AIDS	CCR5 allele (Δ32-BP)
Type 1 diabetes mellitus	DQB1*0302-DRB1*0401
	DQB1*0201-DRB1*03
Type 2 diabetes mellitus	PPARG allele (Pro12Ala)
	CLPN10-g.4852G/A
Venous thrombosis	F5 allele (R506Q)

An alternative and more pessimistic scenario is that there are a very large number of variants, in a large number of genes, influencing disease susceptibility. The number of variants at the rhodopsin, *ABCA4*, *RDS* and *RPGR* genes, each of which causes retinal disease, numbers between 60 and 140, the majority of them rare. There are over 450 rare variants in the *BRCA2* gene with major effects on breast cancer susceptibility, and only six that are polymorphic, one of which increases risk by 30%.[22] Genetic heterogeneity due to different *variants* within the same gene (allelic heterogeneity) is almost invariably high, even in late-onset genetic disorders with high reproductive fitness. In addition, the number of different *genes* causing disease is also high (non-allelic heterogeneity) – for example, over 125 genes causing monogenic forms of retinal degeneration are known,[23] and these account for only a minority, perhaps about one-third, of patients. One reason for the large number of rare variants influencing late-onset disorders is the 1000-fold expansion of the human population which has taken place within comparatively recent times (< 10 000 years), resulting in a large reservoir of mostly rare variants, many of which are also deleterious.[21] In addition, it takes a very long period of time for a new, selectively neutral variant to reach a high frequency in the population: conversely, it is much more likely to be lost or to remain at low frequency.[24] The number of susceptibility genes in ARM/AMD, which can be regarded as a subset of the retinal degeneration phenotype, is plausibly less than the > 125 loci causing this phenotype, but it is unlikely to be a small number. There may therefore be many genes *and* variants capable of influencing a disease such as ARM. The importance of the distinction between a small number of common variants influencing disease and a large number of rare variants at many genes is important in devising a mapping strategy to detect genes of small or intermediate effect.

Gene mapping strategies

Genetic mapping exploits a variety of strategies to enrich for disease susceptibility genes in a patient or disease-risk group (for simplicity, this will be referred to as the 'affected' group) versus the control group. The sample size required to detect a gene depends on its frequency within the study population and its enrichment within affecteds relative to controls. For example, the enrichment in a monogenic disorder is measured by comparing the frequency of a risk allele (or linked marker) in affected and unaffected members in a population of families, each with a high likelihood of carrying (often rare) variants at a particular susceptibility gene. However, if this were done in a group of dominant RP kindreds, it would be unlikely to work, since the number of causal genes (non-allelic heterogeneity) is too high. In the case of a genetically complex disease such as ARM/AMD, the situation is even more difficult, for several reasons. Firstly, the number of 'affected' families which result from non-genetic factors or chance is not trivial. Secondly, non-allelic heterogeneity is likely to be

very high, as discussed above. Thirdly, genes with small or intermediate rather than large marginal effects must be assumed, which provide only modest enrichment for risk alleles in the affected group. The enrichment depends on two factors: firstly, how common the gene is in the study versus the control population, and secondly, the 'penetrance' of the risk allele (extent to which the genotype results in the disease phenotype).[21] The penetrance component is, by definition, reduced in ARM/AMD compared with a monogenic disorder, and if the gene frequency is also low (i.e. in the absence of a polymorphic risk allele), then large sample sizes will be required.

There are several strategies which aim to increase the frequency and enrichment for a risk allele in a complex disease population. The most obvious is to choose probands (i) with a strong family history of disease, or of disease determinants (the major strategy used to enrich for rare monogenic disease genes). This increases the chance that probands have a common disease risk allele rather than non-genetic phenocopies of the disease. Other strategies relevant to ARM/AMD include choosing probands (ii) with early onset of disease (e.g. onset < 65 years), (iii) with severe disease (e.g. AMD rather than ARM), (iv) from a geographic or ethnic cluster (may indicate a locally high frequency of a risk allele), (v) from a disease subgroup (e.g. 'wet' AMD or a monogenic model of ARM), (vi) from a population with low environmental exposure (e.g. non-smokers), (vii) from an inbred population (gene dosage effects may result in more affected individuals as a result of homozygosity) or (viii) from a population isolate (less genetic heterogeneity).[21]

Family versus population mapping

There are two complementary approaches to mapping disease susceptibility genes in genetically complex disorders. The most commonly used approach is a family-based linkage analysis. The problem of extreme variability in disease alleles (allelic heterogeneity) is avoided because each family is an independent unit and genetic markers can be used to track susceptibility alleles within each family, even if the alleles differ between families. The results are readily pooled over all families. The method is therefore robust provided that the majority of families share the same disease susceptibility gene (limited non-allelic heterogeneity). The problem with this approach, which is the mainstay of Mendelian gene mapping, is that it can lack power to detect genes of small effect. The sample size, even with a few hundred affected sib pairs, may be insufficient to detect genes with the magnitude of effects expected for many common diseases.[25] However, if the goal is to map one of the limited number of genes contributing intermediate or large effects to disease susceptibility, then this is a powerful strategy. The logistic and cost implications of conducting linkage studies using large affected sib pair

collections are significant, so that the limited success to date has been achieved mainly by the biotechnology industry.

The second method of mapping genes of small or intermediate effect is by 'genetic association' mapping within whole populations. The population is here considered to be an extension of the family, and assumes one or a small number of common ancestors, hundreds or thousands of generations ago. The aim is to use a dense array of genetic markers to track the ancestral susceptibility gene through to the present generation of affected individuals and to show enrichment for the associated marker in affected relative to control members of the same population. The assumption of a shared founder susceptibility allele, present in a significant fraction of the affected population, is quite consistent with the common (polymorphic) variant hypothesis, but runs counter to the multiple risk allele hypothesis. Two HLA class II haplotypes (i.e. combination of risk alleles on a single chromosomal background) confer a high risk of type 1 diabetes mellitus (t1DM) (Table 2) and are present in about 2% of Europeans, about one-tenth of whom develop diabetes.[26] This situation is not typical, since these haplotypes are almost certainly common in the population because of strong selection for disease resistance in the past, which can dramatically increase the population frequency of a favourable haplotype over relatively short periods of time (e.g. < 2000 years). The same may apply to the common CCR5 variant which influences susceptibility to HIV-1 infection (Table 2). There are many common variants in the human population, most of which tend to be ancient. The majority of genetic variants with frequencies > 10% in the present population are predicted to pre-date the emergence of anatomically modern humans from Africa, 100 000 years ago.[27] Population genetic theory, however, suggests that there is a much larger number of rare variants, which are equally capable of influencing disease and which, collectively, may provide a larger reservoir of potential disease susceptibility alleles. In the latter case, association mapping would be unlikely to work, since there would be no common founder allele with a predominant effect on disease susceptibility. For this reason, researchers are tending to focus on smaller founder populations where genetic diversity is reduced and recent founder alleles are more likely to be detectable. For example, in a recent study of asthma in an isolated North American Hutterite sect, about 700 individuals were studied, all of whom were descended from a pool of about 65 founders within the last 400 years.[28] This led to the mapping of several potential asthma susceptibility genes. In a comparable study of Hirschsprung's disease in a Mennonite sect, a low penetrance gene (EDNRB) was directly identified for this complex disorder.[29]

SNPs and association mapping

Single nucleotide polymorphisms (SNPs) are single nucleotide base substitutions which occur at specific sites in the genome, most of which are thought to have arisen as unique events in the distant past. The least common SNP allele is, by definition, present in most populations with a frequency at least > 1%, and generally > 10%. Public and private consortia have undertaken to identify and provide information on 300 000 SNP markers, spanning the entire human genome, within the next few years.[30] Already, the first high-density SNP maps of human chromosomes have been published.[31,32] SNPs are essential for high-density mapping of ancient disease susceptibility genes in human populations. Conventional microsatellite marker maps are insufficiently dense to track such genes over hundreds or thousands of generations, since they rapidly become separated from the risk locus by ancestral recombination events. SNPs are also available at sufficient density to allow the 'tagging' of most candidate genes and to search for disease associations within population samples.

An ideal strategy for mapping a complex trait such as ARM or AMD would be to carry out an initial low-density genome scan using 400 conventional (microsatellite) markers in a large sample of affected sib pairs by linkage analysis. This should detect any intermediate or large effect genes influencing disease susceptibility. Replication of these results in a second sample, recruited using an identical strategy from the same population, would help to distinguish true- from false-positive linkages. A higher-resolution scan, using SNPs in the region of a replicated linkage, would then be carried out to fine map and, if possible, detect any founder disease alleles. This could be carried out either within the affected sib pair sample or in an independent case–control cohort from the same population.

Case–control studies and genetic admixture

Case–control or cohort studies are a substantially less expensive means of investigating the role of candidate genes in a disease such as ARM. The rather stringent assumption of a limited number of founder susceptibility alleles is a major prerequisite, but the power to detect such variants can be substantial, even if the marginal effects of the gene are small. A major problem with this approach results from the widespread genetic admixture of modern human populations within relatively recent times (< 1000 years) as a result of migration due to political and social upheavals. Admixture is the norm for most urban populations and it can cause problems in association studies if there are substantial differences in the allele frequencies of the parent populations. It can take long periods of time for these differences to re-equilibrate in the admixed population, depending on the extent of integration. In many cases, population stratification occurs as a result of linguistic, cultural or geographic barriers between the parental populations. If the affected group is poorly matched with the controls in

a case–control association study, spurious and unreplicable associations can occur between SNPs within candidate genes and disease. Low-frequency variants at a candidate disease gene may be a particular problem since they are quite likely to differ between the parent populations and to cause inadvertent differences in frequency between disease and control groups, unless these are well matched. The use of non-transmitted parental alleles as the 'pseudo-control' group in affected sib pairs or parent–sib trios is designed to control for this problem and is the gold standard for such studies.[33] Alternatively, if stratification is suspected, it can be estimated and corrected for by genetic typing of unlinked markers.[34]

Association mapping in case–control or cohort studies can, in principle, be used to map genes using high-density SNP maps spanning the entire genome. This has not yet been achieved but retrospective analyses suggest that it should be possible.[35] The major problems are the expense and the number of false-positives that would arise from typing as many as 100 000 or more SNP markers. If a Bonferroni-type correction and false-positive rate (genome-wide p value) of 5% is used, a sample size of at least several thousand individuals would be required to achieve significance.[36]

Candidate genes

For the reasons discussed above, there is an argument for careful consideration of candidate gene analysis as a means of identifying complex disease susceptibility genes in disorders such as ARM/AMD. The Human Genome Mapping Project will soon provide a profile of all human genes. Expression arrays showing which genes are expressed in which tissues or stages of development are starting to become available and will help to select and highlight candidate genes, even in the absence of information on their function. Increased understanding of the pathophysiology of ARM/AMD and of the traits, such as neovascularisation, lipofuscin, drusen or sub-RPE (retinal pigment epithelium) deposit formation, that appear to underlie disease susceptibility, will also facilitate the identification of candidate genes. New information from proteomics, which not only identifies proteins expressed in a tissue, such as RPE, but also identifies protein families and functionally related domains, will extend the choice of candidate genes. Finally, disease models may be the single most effective means of identifying candidate genes that are relevant to ARM/AMD. Animal models have led to the direct identification of several monogenic disorders of the retina. Human models of maculopathy, including diseases such as Stargardt disease (*ABCA4*), Best disease (*VMD2*), Doyne's honeycomb macular dystrophy (*EFEMP1*), pattern dystrophies (*RDS*), Sorsby fundus dystrophy (*TIMP3*), late-onset retinal degeneration, central areolar choroidal dystrophy, North Carolina macular dystrophy and other autosomal dominant forms of macular dystrophy, all provide insights into ARM/AMD.

Summary

In summary, the complexity of ARM and AMD is such that genetic studies of disease-associated traits such as lipofuscin deposition, macular antioxidants or sub-RPE deposits, will become increasingly important. ARM is likely to result from a combination of generalised factors, such as ageing-related oxidative damage, and local factors which influence macular function, each under varying degrees of independent genetic control. AMD probably requires additional risk factors, influencing extracellular matrix and neovascular components of disease. Genetic mapping of such variants is likely to require minimising the genetic diversity of the study population and maximising the enrichment of risk alleles within the 'affected' population. The majority of the one thousand monogenic disease genes identified in the course of the Human Genome Mapping Project have been identified by a combination of genetic mapping and candidate gene approaches. In the case of complex diseases, the combination of genetic mapping, which has no preconceptions as to which genes are involved, and the increasingly informed selection of candidate genes from the region of search, is likely to underpin future success in this important but difficult undertaking.

References

1. VanNewkirk MR, et al. The prevalence of age-related maculopathy. Ophthalmology 2000;107:1593–600.
2. Klein R, Klein BEK, Linton KL. Prevalence of age-related maculopathy. The Beaver Dam Eye Study. Ophthalmology 1992;99:933–43.
3. Yates JRW, Moore AT. Genetic susceptibility to age-related macular degeneration. J Med Genet 2000;37:83–7.
4. Gorin MB. Genetics of age-related maculopathy. In: Traboulsi EI, editor. Genetic diseases of the eye. New York: Oxford University Press, 1998:407–34.
5. Seddon JM, Ajani UA, Mitchell BD. Familial aggregation of age-related maculopathy. Am J Ophthalmol 1997;123:199–206.
6. Klein ML, Mauldin WM, Stoumbos VD. Heredity and age-related macular degeneration: observations in monozygotic twins. Arch Ophthalmol 1994;112:932–7.
7. Meyers SM, Greene T, Gutman FA. A twin study of age-related macular degeneration. Am J Ophthalmol 1995;120:757–66.
8. Gottfredsdottir MS, Sverrisson T, Musch DC, Stefansson E. Age related macular degeneration in monozygotic twins and their spouses in Iceland. Acta Ophthalmol Scand 1999;77:422–5.
9. Seddon JM, Willett WC, Speizer FE, Hankinson SE. A prospective study of cigarette smoking and age-related macular degeneration in women. JAMA 1996;276:1141–6.
10. Smith W, Mitchell P, Leeder SR, Wang JJ. Plasma fibrinogen levels, other cardiovascular risk factors, and age-related maculopathy: the Blue Mountains Eye Study. Arch Ophthalmol 1998;116:583–7.
11. Marantz Henig R. A monk and two peas. London: Weidenfeld and Nicolson, 2000.
12. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.

13. Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th edn. Harlow: Longman, 1996.

14. Sing CF, Haviland MB, Reilly SL. Genetic architecture of common multifactorial diseases. In: Variation in the human genome. Ciba Foundation Symposium 197. Chichester: Wiley, 1996:211–29.

15. Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/omim

16. Schildkraut JM. Examining complex genetic interactions. In: Haines JL, Pericak-Vance MA, editors. Approaches to gene mapping in complex human diseases. New York: Wiley-Liss, 1998:379–410.

17. Marshall J, Hussain AA, Starita C, Moore DJ, Patmore AL. Aging and Bruch's membrane. In: Marmor MF, Wolfensberger TJ, editors. The retinal pigment epithelium: function and disease. New York: Oxford University Press, 2000:669–92.

18. Martin GM, Austad SN, Johnson TE. Genetic analysis of ageing, role of oxidative damage and environmental stresses. Nat Genet 1996;13:25–34.

19. Martin GM. Genetic syndromes in man with potential relevance to the pathobiology of aging. Birth Defects 1978;14:5–39.

20. Kardia SLR, Stengard J, Templeton A. An evolutionary perspective on the genetic architecture of susceptibility to cardiovascular disease. In: Stearns SC, editor. Evolution in health and disease. Oxford: Oxford University Press, 1999:231–45.

21. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. Nat Genet 1999;23:397–404.

22. BIC database. http://www.nhgri.nih.gov/Intramural_research/Lab_transfer/BIC

23. Retinal Information Network. http://www.sph.uth.tmc.edu/RetNet

24. Hartl DL, Clark AG. Principles of population genetics. Sunderland, MA: Sinauer Associates, 1997.

25. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.

26. Todd JA. From genome to aetiology in a multifactorial disease, type 1 diabetes. BioEssays 1999;21:164–74.

27. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999;22:139–44.

28. Ober C, Tsalenko A, Parry R, Cox NJ. A second generation genome wide screen for asthma-susceptibility alleles in a founder population. Am J Hum Genet 2000;67:1154–62.

29. Puffenberger EG, et al. A missense mutation of the endothelin-B receptor gene in multigenic Hirschsprung's disease. Cell 1994;79:1257–66.

30. Collins FS, Guyer MS, Chakravarti A. Variations on a theme cataloguing DNA sequence variation. Science 1997;278:1580–1.

31. Altshuler D, et al. An SNP map of the human genome generated by reduced representation shotgun sequence. Nature 2000;407:513–6.

32. Mullikin JC, et al. An SNP map of human chromosome 22. Nature 2000;407:516–20.

33. Schaid DJ. Transmission disequilibrium, family controls, and great expectations. Am J Hum Genet 1998;63:935–41.

34. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65:220–8.

35. Lai E, Riley J, Purvis I, Roses A. A 4-Mb high density single nucleotide polymorphisms-based map around human APOE. Genomics 1998;54:31–8.

36. McCarthy JJ, Hilfiker R. The use of single-nucleotide polymorphism maps in pharmacogenomics. Nature Biotechnol 2000;18:505–8.