Identifying modifier genes of monogenic disease: strategies and difficulties.

Emmanuelle Genin, Josué Feingold, Françoise Clerget-Darpoux

Identifying modifier genes of monogenic disease: strategies and difficulties

Emmanuelle Génin, Josué Feingold, Françoise Clerget-Darpoux

Inserm UMR-S535, Villejuif, F-94817, France

Université Paris Sud, Villejuif, F-94817, France

Key-words: modifier gene, Mendelian disorders, disease expression, linkage, association

Address for correspondence

Françoise Clerget-Darpoux

Inserm U535

BP 1000

94817 Villejuif Cedex, France

Email: clerget@vjf.inserm.fr

Tel: 01.45.59.53.63

The original publication is available at www.springerlink.com

Fax: 01.45.59.53.31

Abstract

Substantial clinical variability is observed in many Mendelian diseases, so that patients with the same mutation may develop a very severe form of disease, a mild form or show no symptoms at all. Among the factors that may explain these differences in disease expression are modifier genes. In this paper, we review the different strategies that can be used to identify modifier genes and explain their advantages and limitations. We focus mainly on the statistical aspects but illustrate our points with a variety of examples from the literature.

Introduction

Genetic factors can play extremely diverse roles in the etiology of human diseases. A single rare mutation can fully account for a monogenic Mendelian disease, while a set of numerous genetic and environmental factors must be present to cause a multifactorial disease. Huge clinical variability can be observed even for simply determined diseases, and this variability may itself involve genetic factors, the so-called modifier genes.

The modifier gene concept is not new, having been introduced in 1941 by Haldane (Haldane 1941). It may be useful to review its definition, which varies from one study to another. Here are some definitions found in literature:

1. "A gene that is recognized by its alteration of the phenotypic expression of genes at one or more other loci" (Futuyma 1998)
2. "A gene that alters the expression of a gene at another locus" (Hall and Horton 1997)
3. "A gene that affects the phenotypic expression of another gene" (Suzuki et al. 2004)
4. "A gene capable of modifying the manifestation of a mutant gene without having an obvious effect on the normal condition" (Grüneberg 1963).

The variety of interpretations to which these definitions lend themselves highlights the vagueness of the concept. Some studies also refer to these situations as digenic or oligogenic inheritance models, depending on the number of genes involved (Nadeau 2001; Slavotinek and Biesecker 2003). In our view, the difference between modifier genes and oligogenicity lies in the definition of the phenotype. A good example is coat color in mice (for a review see (Silvers 1979)): color is controlled by gene B but its intensity (full or diluted) depends on gene D. If the phenotype is defined in three classes, as white, gray, and black, it can be explained by a digenic inheritance

model. If instead we consider that the primary phenotype is white or black (unaffected/affected), we find differences in intensity among the mice with the black phenotype (full black or gray) and gene D, which controls intensity, is a modifier gene.

Searching for modifier genes is different from searching for the gene(s) responsible for the disease. The differences are in the phenotype to be explained and in the study population. When genes involved in the disease are sought, the phenotype of individuals is usually defined as affected or unaffected whereas for modifier genes, the phenotype of interest must be a measure of the clinical variability in the population of affected individuals (**the disease phenotype** versus **a clinical phenotype**). The difference may appear subtle, especially if the goal is to find the genes that explain the lack of penetrance of a given mutation, for then disease and clinical phenotypes may be the same. Even in this case, however, the population under study differs, for it is restricted to the population of individuals who carry mutations known to be involved in the disease.

Many arguments support the concept of modifier genes and numerous studies have identified such genes in mice. A traditional example is multiple intestinal neoplasia. In mice, this is due to a dominant mutation of the Apc gene, but the number of intestinal tumors depends on the Mom-1 (Modifier of Min-1) gene (Dietrich et al. 1993). In humans, modifier genes are often suggested to explain clinical variability in monogenic diseases (Feingold 2000; Nadeau 2001; Wolf 1997), but very few modifier genes have been identified so far and the mechanisms underlying clinical variability remain poorly understood, probably because of the involvement of complex mechanisms and multiple factors.

Identifying these genetic modifiers may be of great interest from the viewpoints of both treatment and genetic counselling (Lyonnet et al. 2003), but it remains very challenging, despite the

powerful genetic tools available today. In this paper, we describe possible strategies to identify modifier genes, strategies that appear very similar to those used to study multifactorial diseases, but with additional specificities and problems.

Choice of clinical phenotypes and study populations

The first and probably the most important steps in the study of modifier genes are to define the clinical phenotype for which one seeks modifier genes and to choose the study population.

The clinical phenotype may be **qualitative.** Examples include the presence or absence of meconial ileus in cystic fibrosis, the presence or absence of Hirschsprung's disease in Ondine's curse, the presence or absence of scoliosis in neurofibromatosis, and the four severity classes of spinal muscular atrophy. Alternatively it may be **quantitative,** such as age at onset in Friedreich's ataxia or Huntington's disease, survival time in hypertrophic cardiomyopathy, or forced expiratory volume value in one second (FEV_1) in cystic fibrosis. Choosing a relevant clinical phenotype may be difficult, and **adjustment for appropriate covariables** (such as age and sex) is often necessary. For example, in studying hypertrophic cardiomyopathy, one may be interested in maximal wall thickness (MWT), interventricular septum thickness (IVS) or, as proposed by Spirito and Marron (Spirito and Marron 1990), a score combining several measurements of left ventricular hypertrophy (LVH score). All these variables must be adjusted for appropriate covariates, such as age, sex and body surface area (Forissier et al. 2005). Success in finding genetic factors may depend on the choice of variable and on the inclusion of appropriate covariates in the model. The study by Milet et al. (Milet et al. 2007) of serum ferritin levels in hereditary hemochromatosis illustrates this point: the effect of one SNP in the bone morphogenetic protein 2 (BMP-2) gene was significant after adjustment for age and sex (nominal

p<0.0001 and p<0.0075 after correction for 75 tests) and was only borderline without any such adjustment (nominal p=0.0007 and p=0.053 after correction for 75 tests)(J. Milet, personal communication). Beaumont et al. (1976) (Beaumont et al. 1976) studied the incidence of ischemic disease in familial hypercholesterolemia and xanthomatosis and found that the difference observed between men and women disappeared after adjustment for smoking habits. The choice of phenotype is also difficult when looking for modifier genes that may be involved in lung disease severity in cystic fibrosis patients. Most studies use empirical Bayes estimates of FEV_1 (% predicted) at age 20, rather than crude FEV_1, because this age has been shown to be best for distinguishing between patients with mild and severe disease (Schluchter et al. 2006).

The **population of individuals to be studied** must also be defined**.** Depending on the disease, one might focus on individuals with a given mutation known to be involved in the monogenic disease. For example, the study of the modifier genes involved in different phenotypes associated with cystic fibrosis is usually performed in the subpopulation of individuals homozygous for the DeltaF508 mutation of the CFTR gene. This restriction on a given mutation is possible however only for monogenic disease where one major mutation accounts for a large proportion of the cases. In many instances, this condition cannot be met and investigators must consider the population of patients with any of the mutations known to be involved in the disease. As discussed below, it is then necessary to determine the proportion of variability in the trait that is explained by any difference in these primary mutations.

Different explanations of variability in disease expression

Several factors other than modifier genes might explain disease expression variability, and it is important to verify that these other factors do not explain all the variability in disease expression. In particular, **environmental factors** might be involved but in this review, we will focus on the

genetic causes of disease expression variability.

Genetic heterogeneity of the primary factor involved in the disease

Genetic heterogeneity may either be at the gene level, with different genes involved in the different sub-entities of the disease, or at the mutation level, with different mutations of the same gene leading to different phenotypic expression of the disease. One major example of genetic heterogeneity is hypertrophic cardiomyopathy, an autosomal dominant disease that can be due to more than 300 different mutations, most of them (~65%) located in two genes encoding the sarcomeric proteins MyBPC3 and Myh7. Survival time is longer for patients with an MyBPC3 mutation than for those with an Myh7 mutation (Charron et al. 2002; Richard et al. 2003). However, heterogeneity at the mutation level might also influence survival time, as shown by Watkins et al. (Watkins et al. 1992) for mutations in Myh7.

Breast and ovarian cancers furnish another example. Breast cancer can caused by mutations in the BRCA1 or the BRCA2 genes. BRCA2 mutation carriers are at greater risk of developing ovarian cancer, but this risk also depends on the position of the mutation in these genes: in the BRCA1 gene, the risk is smaller when the mutation is in exons 13 to 24 (Shattuck-Eidens et al. 1995) and in the BRCA2 gene, the risk is greater when the mutation is at the 3'end of the gene, compared with the 5'region (Gayther et al. 1997). Another example is cystic fibrosis, where pancreatic insufficiency is usually observed in patients homozygous for severe mutations of the CFTR gene (class I, II or III mutations).

The age of onset in Huntington's disease (HD) is also an interesting example. This disease is caused by an autosomal dominant expanded triplet (CAG) repeat, located in the huntingtin gene on chromosome 4p16.3. Alleles with more than 40 CAG repeats are considered fully penetrant, but some significant differences in the age of onset of motor symptoms depend on the specific CAG count: a higher number of repeats is associated with earlier onset.

Verifying that genetic heterogeneity does not explain all the phenotype variability of interest is an important step, because this issue may have implications for which patients should be studied and specifically, whether the study should be limited to patients who do or do not carry a given mutation. It can also help guide the choice of the clinical phenotype to be studied. To study a quantitative phenotype, for example, the age of onset of motor symptoms in HD, clinical phenotype might be defined by determining the proportion of the total variance in the trait that is explained by the mutation type (bearing in mind that variance is a measure of statistical dispersion, averaging the squared distance of its possible values from the expected value). Using data on 443 HD patients with CAG repeats ranging from 40 to 86, Wexler et al. (Wexler et al. 2004) found that the length of the repeat accounted for 72% of the variance in age of onset. Accordingly the residual age of onset after accounting for the contribution of the CAG repeat lengths is an appropriate phenotype to study.

Effect of another variant in the gene in cis or trans position

Phenotypic variability might also be due to the effect of another variant in the gene, in either the trans or cis position to the primary mutation. In some dominant diseases, clinical expression may depend on the normal allele (trans effect). Good examples are erythropoietic protoporphyria, where a low expressed allelic variant of the ferrochelatase gene located in trans from the mutation explains the variability in disease expression (Gouya et al. 1996; Gouya et al. 2006; Gouya et al. 1999) and hereditary elliptocytosis, where the α^{LELY} allele increases mutation expression when it is located in trans position to the mutation (Delaunay et al. 1995). Alternatively, cis effects might be suspected when the haplotype carrying the disease mutation varies with the clinical phenotype. A striking example is that of Creutzfeldt-Jakob and familial fatal insomnia: carriers of a single

mutation in codon 128 of the prion protein gene located on chromosome 20 develop one or the other disease depending on a polymorphism at codon 129 of the same gene that codes for two different amino acids — valine or methionine (Goldfarb et al. 1992).

Modifier genes of disease expression

Disease expression variability might also be explained by the effect of genes other than the primary one involved in the disease, and it is these that are usually referred to as modifier genes. Their effect on disease expression may vary from strong effects under a "monogenic-like" model to much milder effects under a "multifactorial-like" model.

Under the monogenic-like model, a single modifier gene exhibits rare fully or almost fully penetrant mutations that explain all or a very important part of the variability in disease expression. Examples of this type of modifier genes can be found among the genes involved in the splicing machinery (see the recent review of Wang and Cooper (Wang and Cooper 2007)).

Under the multifactorial-like model, disease expression depends on the effects of several genetic variants located in different modifier genes that, by themselves, only explain a small proportion of the variability but interact both with one another and with environmental factors. This is probably the most common situation and is the one on which we have chosen to focus. These genetic modifiers are in fact very similar to the genetic risk factors involved in complex diseases, and the same strategies are used to help uncover these genetic modifying factors. The main difference is that population sizes are much more limited.

Evidence for the role of genetic factors in the variability of disease expression

Before planning the search for genetic factors involved in any monogenic variability in disease expression, **the role of familial factors** must be shown by comparing the correlation of the phenotype of interest in related and unrelated patients. If genetic factors play a role, **inter-family variability** should be greater than **intra-family variability**. This is often difficult to study especially for rare monogenic diseases. Even for more common diseases such as cystic fibrosis (CF), the task is not necessarily easy since genetic counselling and prenatal diagnosis have considerably reduced the number of families with multiple affected individuals. A recently published article by Vanscoy et al. (Vanscoy et al. 2007) collected data about the severity of lung disease in 231 sibling pairs affected by CF. This was possible only through a huge project, the "CF Twin and Sibling Study," which involves 71 CF care centers throughout the United States.

If possible, collection of **data on twins** can be especially useful in showing that genetic factors are included among these familial factors and in obtaining estimates of the heritability associated with different phenotypes of lung disease severity. The study by Vanscoy et al. (Vanscoy et al. 2007) studied both monozygotic and dizygotic CF twins and concluded that modifier genes are likely to be involved in the variability of CF lung disease.

Another way of demonstrating the role of genetic factors in disease expression variability is to take advantage of **particular population contexts,** as in the study of the age of onset of HD in Venezuelan kindreds (Wexler et al. 2004), the largest and best characterized HD population in the world (Okun and Thommi 2004). Most affected individuals are descendents of one woman who lived in the early nineteenth century in a stilt village on Lake Maracaibo, died from HD and passed her abnormal allele through ten generations. Using data about 443 heterozygous members of the Venezuelan kindreds with CAG repeat lengths varying from 40 to 86 and applying a variance component approach to the residual age of onset phenotype (after accounting for the repeat length contribution), Wexler et al. showed that this phenotype has a 38% heritability and

therefore that genetic factors besides the CAG repeat lengths are probably involved in the determination of age of HD onset (Wexler et al. 2004).

Twin and sibling studies might also provide clues to the genetic model underlying disease expression variability. We would expect to see a complete or almost complete phenotypic correlation within families if this variability were explained by the existence of different mutations and, similarly, if cis-acting genetic variants were involved. If rare trans-acting genetic variants play a causal role, phenotype concordance would be expected in about half the siblings and complete concordance in monozygotic twins. For modifier genes in a monogenic-like model, a similar pattern of concordance would be expected with complete or almost complete concordance in monozygotic twins, while concordance rates for modifier genes in a multifactorial-like model would be much smaller.

Strategies to show the role of genetic modifiers

Strategies used to show the role of genetic factors in phenotypic expression are often classified into two categories **depending on the type of data available**: linkage studies and association studies. Another distinction often made is based on the approach, which can be either a **systematic approach** where the whole genome is scanned or a **more focussed approach,** where candidate genes or candidate pathways are selected.

Linkage analysis of family data

If data are available about affected siblings, one useful analysis may be a linkage screen, which compares the number of alleles shared identical by descent by affected sibs between phenotypically-concordant and discordant sibpairs. Three types of affected sibpairs can in fact be distinguished f**or qualitative (binary) phenotypes**: sibpairs concordant for the phenotype of

interest, sibpairs concordant but without the phenotype and sibpairs discordant for the phenotype. Note that concordant sibpairs without the phenotype of interest are often included in linkage studies to search for modifier genes. Depending on the genetic model and in particular on the prevalence of the phenotype of interest in the patient population and on its recurrence among siblings (Houlston and Tomlinson 1998), these concordant sibpairs without the phenotype might not be helpful in detecting linkage with this phenotype, but they can help to show genetic factors that protect against it. Such a linkage strategy was used to search for genetic factors involved in meconium ileus in CF patients. Zielenski et al. (Zielenski et al. 1999) studied a sample of 152 CF affected DeltaF508 homozygous carrier sibpairs. The distribution of affected sibpairs sharing 2, 1 and 0 alleles identical by descent observed in the 19q13 region differed according to whether both sibs had meconium ileus. A strong departure from the expected proportions of ¼, ½ and ¼ was observed, especially in the sibpairs discordant for meconium ileus (see Table 1). On average they shared fewer alleles identical by descent than expected. Because most of the evidence for linkage came from this category of patients, discordant sibpairs might in fact have been used to search for modifier genes. Computing power under various genetic models, Houlston and Tomlinson (Houlston and Tomlinson 1998) showed that the power to detect linkage could be considerably increased by using phenotypically discordant sibpairs, compared with concordant pairs. The range of models investigated was limited, however, because they considered only the effect of a dominant or a recessive gene with high penetrance. Under scenarios with smaller penetrances, the required sample sizes would be very large and impossible to collect. Discussions of the relative interest of the different types of affected sibpairs can also be found in the literature on using linkage analysis to search for genetic risk factors involved in multifactorial disease (see for example (Rogus and Krolewski 1996)). A more recent genome-wide linkage study aimed at identifying genetic factors involved in the occurrence of meconium ileus in CF did not confirm

the linkage with the 19q13 region but found regions of suggestive linkage on chromosomes 4q35, 8p23 and 11q25 (Blackman et al. 2006).

Linkage analyses **for quantitative phenotypes** can also be performed with several different methods, including variance components. One example is the study of the age of onset of motor symptoms in HD by the HD-MAPS group (Li et al. 2003). A genome-wide scan of 629 affected sibpairs from 295 families used a variance component approach to age at onset, adjusted for the number of CAG repeats. It found evidence suggestive of linkage (LOD ≥ 2.19) with two regions of chromosome 6 (6p21-23 and 6q24-26). A more recent study by the same group of 102 additional sibpairs confirmed the linkage with the 6q region (Li et al. 2006) with a LOD of 4.94 on the combined sample of more than 700 sibpairs. Methods other than variance component that can also be used to search for linkage with quantitative phenotypes include Haseman-Elston regression and Bayesian methods (for a review, see (Feingold 2001)).

It is also possible to dichotomize the quantitative phenotype rather than analyzing it, and then use the same methods as for qualitative traits. The choice of the optimal cutoff point for dichotomization is not necessarily easy, however, methods to optimize this point should be considered. The ordered subset analysis (OSA) method proposed by Hauser et al. (Hauser et al. 2004) might be of particular interest in the search for genetic modifier. The idea of the method is to use the information on a quantitative covariate to rank affected sibpairs in subsets of increasing size where linkage is tested with the disease. It was presented as a method to test for linkage in the presence of genetic heterogeneity but it can also be useful for determining the role of modifier genes in Mendelian diseases.

Although linkage analyses can successfully identify modifier genes, they also have limitations, which are summarized in Table 2. In particular, they lack power to detect the effect of frequent

alleles. Because such alleles are likely to enter genealogies more than once, tracking them can be difficult.

Association studies of case-control data

An alternative strategy to linkage consists in testing for associations in case samples. This is the most widely used strategy in the search for modifier genes involved in CF-associated lung disease, probably because it requires sampling patients only, rather than collecting family data, as in linkage studies. **For qualitative phenotypes**, the distribution of marker genotypes is compared in patients with and without the phenotype of interest to detect markers that show differences. These may be involved in phenotype expression or associated (in linkage disequilibrium) with loci involved in phenotype expression. **For quantitative phenotypes**, the average value of the phenotype for the different genotypes can be compared with ANOVA or t-tests. Depending on the phenotype and the model underlying the effect of the modifier gene on the trait, power may be increased with the first approach after dichotomization of the trait (Fardo et al. 2007). As in linkage studies, however, the choice of the appropriate cutoff point might not be easy, and methods that rank cases based on their quantitative traits, with measures similar to the OSA method, should be considered (Macgregor et al. 2006).

The major limitation of population-based association tests is that once an association is detected, it is often difficult to determine if it is due to the direct effect of the polymorphism of interest on the trait, to the effect of another variant in linkage disequilibrium with the markers being studied, or to another unknown confounding factor that might not even be genetic. To illustrate the latter possibility, consider a multicenter study of CF-associated lung disease where care practices differ substantially between centers. Suppose that, in center A, patients show a milder form of the disease because of a better care practice and that moreover in the region where center A is

located, there are also important allele frequency differences at one marker M, compared with the population in other regions. An association between marker M and lung disease severity might then be found, not because marker M is indeed involved but because the subgroup of patients with the mild forms will have an excess number of individuals from center A where allele frequencies at marker M are different. This is the well-known issue of population admixture, encountered in population-based association studies. The problem is probably even more acute in the search for modifier genes, which require multicenter studies to collect large enough patient populations and which may consider phenotypes more subtle than affected versus unaffected. Several methods to help prevent false conclusions due to population admixture have been proposed, ranging from careful matching to the use of random markers to help detect hidden sub-populations (Cardon and Palmer 2003). These strategies appear to be used relatively rarely in the search for genetic modifiers, and population stratification is often neglected. The recent availability of large samples of individuals genotyped for hundreds of thousands of markers has demonstrated that allele frequency differences between populations are a concern even within populations previously considered relatively homogeneous (Choudhry et al. 2006; Steffens et al. 2006).

New methods are currently being developed to account and correct for population stratification in association studies (Epstein et al. 2007; Li and Yu 2008; Luca et al. 2008; Price et al. 2006). Another strategy consists in using family-based association tests with case-parent trio designs and the transmission disequilibrium test (TDT) (Spielman et al. 1993). The advantage of this approach is that it tests for both linkage and association and thus ensures that any significant results are not due to population admixture. The basic idea of these tests is to compare the alleles that parents do and do not transmit to their affected children. Thus, the search for modifier genes must examine whether or not there is a difference in parental transmissions according to the

phenotypic categories of the affected children. The sampling of case-parent trios might not be difficult when studying diseases such as CF, which occur early in life, but this strategy has never been used to identify genetic modifiers.

When the phenotype of interest is quantitative, different methods have been proposed to test for associations with case-parent trio data. One is the quantitative TDT (QTDT) method, which relies on a variance component approach (Abecasis et al. 2000). Alternatively, one might consider ordered TDT (OTDT), which, like the OSA method for linkage analysis, is based on the ordering of patients as a function of their quantitative phenotypic measures (Perdry H et al. 2007). The method's aim is to find a critical value of the phenotypic measure that separates the trios into two groups with significantly different transmission rates. No assumption about the distribution of the phenotype in the population is made, contrary to the QTDT method, which requires normal distributions. Perdy et al. found the OTDT to be more powerful than the QTDT in a large variety of models(Perdry et al. 2007).

Blind search: Systematic genome-wide screens

The systematic screen, which consists in searching for the genetic factors involved in the phenotype of interest over the whole genome, was initially only possible for linkage testing. Maps of ~400 microsatellite markers spaced an average of 10 centimorgans (cM) apart over the whole genome were available and could be used to perform linkage tests. This intermarker distance was enough to ensure good coverage of the genome when information about the co-segregation of phenotype and markers in families was used. A much denser map is needed, however, to test for associations. Only very recently have such maps become available, through international projects such as HapMap (www.HapMap.org). Maps of single nucleotide polymorphisms (SNPs) are now available that provide good coverage of all common variations

(those with a minor allele frequency of more than 5%) of the human genome. These maps include more than several hundred thousand markers that are available in chips (the new Affymetrix genome-wide SNP-array 6.0 includes 1.8 million genetic markers randomly chosen to cover the genome, and its competitor, the Illumina Human1M BeadChip, allows genotyping of 1 072 820 markers, an important proportion of them located in genes). In recent years, these chips have been used to perform genome-wide association (GWA) studies to search for the genetic risk factors involved in complex diseases such as Type-2 diabetes, obesity, age-related macular degeneration, Crohn disease, etc.

In age-related macular degeneration, an association was detected after a genome-wide screen with the early 100K chips on 96 cases and 50 controls with an intronic SNP located in the complement factor H gene. After re-sequencing the region, a nonsynonymous Y402H polymorphism (due to a T to C base change at SNP rs1061170) in linkage disequilibrium with the primary associated SNP was suggested to be polymorphism involved in disease susceptibility (Klein et al. 2005). Subsequent studies have confirmed this association, and a meta-analysis estimates that the risk is increased by 6 and 2.5 for, respectively, the homozygous and heterozygous carriers of the C allele at rs1061170 (Thakkinstian et al. 2006). This study on age-related macular degeneration is often used as a proof-of-principle that GWA studies might work even with small samples; however, this is only one example among many others where genetic factors are found to be associated but confer very small risks (odds ratio of less than 1.5). A good example is the GWA studies performed on type-2 diabetes. Six GWA studies published in 2007 provided evidence for 6 new gene regions. Added to the five genes already known to be associated with type-2 diabetes, this makes a total of 11 gene regions associated with this disease with risk allele frequencies ranging from 0.31 to 0.87 and allelic odds ratio from 1.10 to 1.37 (Frayling 2007). To obtain the statistical evidence needed to detect such factors, very large

samples of several thousand cases and controls are needed. Indeed, the large number of markers tested makes it necessary to correct for multiple testing and thus to use very stringent criteria to conclude in a significant association (see BOX 1 for the multiple testing issue). Investigators will find this still more difficult when searching for modifier genes, because they will need to consider patient sub-samples and will rarely have available sufficient large samples. Figure 1 reports the required sample sizes to detect the association with a 80% power, on the assumption of a genetic risk factor with an effect similar to that reported by Drumm et al. (Drumm et al. 2005) for the C509T polymorphism of the TGFβ1 gene in CF-associated lung disease. Sample sizes are reported for both population-based and case-parent trio samples, and a Bonferroni correction for multiple testing is performed to ensure that the global type-one error rate is less than 5%. If 500,000 markers are tested to cover the whole genome, 1905 individuals with the phenotype of interest and 1905 without it will be required to ensure 80% power to detect the association. For modifier genes, this means 3810 patients with mutations involved in the Mendelian disease, a number completely unrealistic even for the most frequent Mendelian disorders. A case-parent trio design would require 1914 patients and their parents. Although this sample size is also impossible to collect, it offers a non-negligible economy in terms of number of patients and shows that the trio design deserves more attention in modifier gene studies. Also note that when new associations are described, replication studies of independent samples are essential. Guidelines for replicating genotype-phenotype associations have been proposed (Chanock et al. 2007), but they are difficult to follow when looking for genetic modifiers since independent samples of sufficient sizes with the same phenotypic information as the initial sample may simply not exist.

Biology-driven approach: Candidate gene tests

Instead of blind searches for modifier genes over the whole genome, it may be wiser to focus on a more limited number of carefully chosen genes, the so-called candidate genes. The difficulty here is choosing these candidates. Different approaches can be used to select them. One might look first at the genes involved in the same pathway as the primary mutation in the disease. For example, for familial hypercholesterolemia due to a mutation in the LDL receptor gene, genes of the lipoprotein pathway are good candidates. Alternatively, one might decide to focus on genes located in another pathway and involved in somewhat more indirect disease consequences. In CF, for example, candidate gene studies have considered genes involved in the inflammatory process. In hereditary hemochromatosis, a recent study showed that genes in the BMP pathway and involved in the expression of hepcidin, a peptide hormone produced by the liver that controls plasma iron concentration, might be promising candidates to explain the penetrance variability of the HFE p.C282Y mutation in homozygote carriers (Milet et al. 2007). Interestingly, the authors focused on an indirect measure of disease penetrance, the serum ferritin levels of C281Y homozygotes. This is the first association detected between common variants in genes of the BMP pathway and iron burden. Further studies will need to determine if this is specific to p.C282Y carriers, by testing for the effect of these variants on serum ferritin levels in the general population. Another example involves dilated cardiomyopathy, where candidate genes in different pathways are being studied, in particular the beta-adrenergic pathway and the renin-angiotensin-aldosterone system. New approaches have also been used based on animal models, which allow a better control of the environment (Komajda and Charron 2004).

Discussion

When the clinical expression of a monogenic disease varies considerably between patients, it is tempting to try to explain it by the effect of some other genetic factors that modify the expression of the primary mutation, i.e., modifier genes. However, before launching expensive and time-consuming genetic studies to identify these genetic modifiers, it is important to make sure that they really exist and that environmental factors or other mechanisms, such as genetic heterogeneity, do not suffice to explain this clinical variability. Investigators should also keep in mind that the effect of modifier genes might be very complex, as it is for the genetic risk factors involved in common diseases. Several genetic variants might be involved and may interact to modulate the effect of the primary mutation. It is even possible that phenotypic variability may be explained not by the patients' modifier genotypes but by their mothers' genotypes, as recently reported for the maternal Apo E genotype in Smith-Lemli-Opitz syndrome (Witsch-Baumgartner et al. 2007).

Structural genomic variations, such as copy number variants (CNVs), might also be involved in the variability of penetrance and phenotypic expression in Mendelian diseases (for a review see (Beckmann et al. 2007)). Some but not all of these CNVs can be detected because of their linkage disequilibrium with common SNPs, and alternative strategies will need to be developed to test for associations with both CNVs and SNPs. Recent reports concerning the functional structure of the human genome show that other mechanisms, such as differences in transcription, may also explain disease expression variability and that the transcription domain of genes might extend very far beyond the usual regulatory sequences. These findings open up new perspectives for the search of cis-acting alleles (Encode-Project-Consortium 2007).

An important issue in the search for modifier genes is the choice of study phenotype. It is often possible to use different variables to characterize disease expression, and studies of the

heritability associated with them might help to choose the appropriate phenotype. The studies on CF and Huntington's disease illustrate this well. Nonetheless, it may be an impossible task for many Mendelian disorders because of the rarity of cases. The candidate genes may also dictate the choice of phenotype to look at, as in the example of hereditary hemochromatosis and the BMP pathway.

It is very tempting to follow the trend from restricted candidate gene studies to extensive genome-wide scans, as in the genetic studies of common diseases. There is not, however, a single best strategy that will work in all scenarios. The strategy depends, rather, on the model underlying the genotype-phenotype relation and since this model is not known, it is impossible to predict what will be found. Recent genome-wide association studies show that this strategy could work but it might also be disappointing, especially in modifier gene studies where only limited samples are available and replication on independent samples difficult. With the decreasing cost of genome-wide SNP arrays, it might be worthwhile to do whole genome chips and analyse only the candidate gene regions rather than to design special assays to study the candidate genes. This is true for candidate genes that are well covered in HapMap. However, recent studies show that a relatively high proportion of the common polymorphisms (minor allele frequency above 5%) detected in 500 genes through re-sequencing are not tagged by SNPs from HapMap, ranging from 50% to 20% depending on the population (Xu et al. 2007). The same study estimates that only approximately 30% of the nonsynonymous SNPs are in high LD with any HapMap SNP. This evaluation shows that candidate gene studies with re-sequencing of the selected genes in a subset of individuals remains a strategy to consider. It is also important to keep in mind that linkage information may be useful in understanding the role of a genetic variant in the phenotypic variation. Association and linkage provide complementary information and efforts to collect family data need to be continued (Bourgain et al. 2007; Clerget-Darpoux and Elston 2007). This

is particularly true when studying modifier genes for which familial information might be more easily available.

The search for modifier genes is difficult but worth being pursued — not only for the direct possibilities it might offer patients affected by the disease but also for the better knowledge of biological pathways that will flow indirectly from this quest. Although interest has shifted gradually from monogenic to more common multifactorial diseases, it is important to keep in mind that monogenic diseases represent a simpler model of diseases that teach us many things about the genetic basis of more complex diseases (Antonarakis and Beckmann 2006). The study of Mendelian disorders may also lead to the discovery of novel drug targets (Brinkman et al. 2006).

References

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66: 279-92

Antonarakis S, Beckmann J (2006) Mendelian disorders deserve more attention. Nat Rev Genet 7: 277-82

Beaumont V, Jacotot B, Beaumont JL (1976) Ischaemic disease in men and women with familial hypercholesterolaemia and xanthomatosis A comparative study of genetic and environmental factors in 274 heterozygous cases. Atherosclerosis 24: 441-450

Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet 8: 639-46

Blackman S, Deering-Brose R, McWilliams R, Naughton K, Coleman B, Lai T, Algire M, Beck S, Hoover-Fong J, Hamosh A, Fallin M, West K, Arking D, Chakravarti A, Cutler D, Cutting G (2006) Relative contribution of genetic and nongenetic modifiers to intestinal obstruction in cystic fibrosis. Gastroenterology 131: 1030-9

Bourgain C, Genin E, Cox N, Clerget-Darpoux F (2007) Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? Eur J Hum Genet 15: 260-3

Brinkman R, Dubé M, Rouleau G, Orr A, Samuels M (2006) Human monogenic disorders - a source of novel drug targets. Nat Rev Genet 7: 249-60

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361: 598-604

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF, Jr., Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS (2007) Replicating genotype-phenotype associations. Nature 447: 655-60

Charron P, Héron D, Gargiulo M, Richard P, Dubourg O, Desnos M, Bouhour J, Feingold J, Carrier L, Hainque B, Schwartz K, Komajda M (2002) Genetic testing and genetic counselling in hypertrophic cardiomyopathy: the French experience. J Med Genet 39: 741-6

Choudhry S, Coyle NE, Tang H, Salari K, Lind D, Clark SL, Tsai HJ, Naqvi M, Phong A, Ung N, Matallana H, Avila PC, Casal J, Torres A, Nazario S, Castro R, Battle NC, Perez-Stable EJ, Kwok PY, Sheppard D, Shriver MD, Rodriguez-Cintron W, Risch N, Ziv E, Burchard EG (2006) Population stratification confounds genetic association studies among Latinos. Hum Genet 118: 652-64

Clerget-Darpoux F, Elston RC (2007) Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. Hum Hered 64: 91-6

Delaunay J, Wilmotte R, Alloisio N, Marechal J (1995) The quiet yet dangerous alpha[LELY] allele of red cell spectrin. M.S. Med. sci. 11: 752-754

Dietrich W, Lander E, Smith J, Moser A, Gould K, Luongo C, Borenstein N, Dove W (1993) Genetic identification of Mom-1, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. Cell 75: 631-9

Drumm ML, Konstan MW, Schluchter MD, Handler A, Pace R, Zou F, Zariwala M, Fargo D, Xu A, Dunn JM, Darrah RJ, Dorfman R, Sandford AJ, Corey M, Zielenski J, Durie P,

Goddard K, Yankaskas JR, Wright FA, Knowles MR (2005) Genetic modifiers of lung disease in cystic fibrosis. N Engl J Med 353: 1443-53

Encode-Project-Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816

Epstein MP, Allen AS, Satten GA (2007) A simple and improved correction for population stratification in case-control studies. Am J Hum Genet 80: 921-30

Fardo D, Celedon JC, Raby BA, Weiss ST, Lange C (2007) On dichotomizing phenotypes in family-based association tests: quantitative phenotypes are not always the optimal choice. Genet Epidemiol 31: 376-82

Feingold E (2001) Methods for linkage analysis of quantitative trait loci in humans. Theor Popul Biol 60: 167-80

Feingold J (2000) Les gènes modificateurs dans les maladies héréditaires. Médecine Sciences 16: I-V

Forissier J, Charron P, Tezenas du Montcel S, Hagège A, Isnard R, Carrier L, Richard P, Desnos M, Bouhour J, Schwartz K, Komajda M, Dubourg O (2005) Diagnostic accuracy of a 2D left ventricle hypertrophy score for familial hypertrophic cardiomyopathy. Eur Heart J 26: 1882-6

Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet 8: 657-62

Futuyma D, J (1998) Evolutionary biology, 3rd edn. Sinauer, MA

Gayther S, Mangion J, Russell P, Seal S, Barfoot R, Ponder B, Stratton M, Easton D (1997) Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. Nat Genet 15: 103-5

Goldfarb L, Petersen R, Tabaton M, Brown P, LeBlanc A, Montagna P, Cortelli P, Julien J, Vital C, Pendelbury W (1992) Fatal familial insomnia and familial Creutzfeldt-Jakob disease: disease phenotype determined by a DNA polymorphism. Science 258: 806-8

Gouya L, Deybach J, Lamoril J, Da Silva V, Beaumont C, Grandchamp B, Nordmann Y (1996) Modulation of the phenotype in dominant erythropoietic protoporphyria by a low expression of the normal ferrochelatase allele. Am J Hum Genet 58: 292-9

Gouya L, Martin-Schmitt C, Robreau A, Austerlitz F, Da Silva V, Brun P, Simonin S, Lyoumi S, Grandchamp B, Beaumont C, Puy H, Deybach J (2006) Contribution of a common single-nucleotide polymorphism to the genetic predisposition for erythropoietic protoporphyria. Am J Hum Genet 78: 2-14

Gouya L, Puy H, Lamoril J, Da Silva V, Grandchamp B, Nordmann Y, Deybach J (1999) Inheritance in erythropoietic protoporphyria: a common wild-type ferrochelatase allelic variant with low expression accounts for clinical manifestation. Blood 93: 2105-10

Grüneberg H (1963) The Pathology of Development; a Study of Inherited Skeletal Disorders in Animals. Wiley, New York

Haldane J (1941) The relative importance of principal and modifying genes in determining some human diseases. Journal of Genetics 41: 149-157

Hall J, Horton W (1997) Genetics Glossary. Growth, Genetics and Hormones

Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M (2004) Ordered subset analysis in genetic linkage mapping of complex traits. Genet Epidemiol 27: 53-63

Houlston RS, Tomlinson IP (1998) Modifier genes in humans: strategies for identification. Eur J Hum Genet 6: 80-8

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005)

Complement factor H polymorphism in age-related macular degeneration. Science 308: 385-9

Komajda M, Charron P (2004) A new approach for the identification of modifier genes in heart failure. Pharmacogenomics J 4: 221-3

Li J-L, Hayden M, Almqvist E, Brinkman R, Durr A, Dodé C, Morrison P, Suchowersky O, Ross C, Margolis R, Rosenblatt A, GÃ³mez-Tortosa E, Cabrero D, Novelletto A, Frontali M, Nance M, Trent R, McCusker E, Jones R, Paulsen J, Harrison M, Zanko A, Abramson R, Russ A, Knowlton B, DjoussÃ© L, Mysore J, Tariot S, Gusella M, Wheeler V, Atwood L, Cupples L, Saint-Hilaire M, Cha J, Hersch S, Koroshetz W, Gusella J, MacDonald M, Myers R (2003) A genome scan for modifiers of age at onset in Huntington disease: The HD MAPS study. Am J Hum Genet 73: 682-7

Li J-L, Hayden M, Warby S, Durr A, Morrison P, Nance M, Ross C, Margolis R, Rosenblatt A, Squitieri F, Frati L, Gomez-Tortosa E, Garcia C, Suchowersky O, Klimek M, Trent R, McCusker E, Novelletto A, Frontali M, Paulsen J, Jones R, Ashizawa T, Lazzarini A, Wheeler V, Prakash R, Xu G, Djousse L, Mysore J, Gillis T, Hakky M, Cupples LA, Saint-Hilaire M, Cha J-H, Hersch S, Penney J, Harrison M, Perlman S, Zanko A, Abramson R, Lechich A, Duckett A, Marder K, Conneally PM, Gusella J, MacDonald M, Myers R (2006) Genome-wide significance for a modifier of age at neurological onset in Huntington's Disease at 6q23-24: the HD MAPS study. BMC Medical Genetics 7: 71

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. 95: 221-227

Li Q, Yu K (2008) Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genet Epidemiol 32: 215-26

Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann HE, Schreiber S, Krawczak M, Lu Y, Styche A, Devlin B, Roeder K, Trucco M (2008) On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. Am J Hum Genet 82: 453-63

Lyonnet S, Feingold J, Frezal J (2003) Genotype-phenotype relationships. In: DN C (ed) Nature encyclopedia of human genome. Nature Publishing Group, London, pp 56-63

Macgregor S, Craddock N, Holmans PA (2006) Use of phenotypic covariates in association analysis by sequential addition of cases. Eur J Hum Genet 14: 529-34

Milet J, Dehais V, Bourgain C, Jouanolle AM, Mosser A, Perrin M, Morcet J, Brissot P, David V, Deugnier Y, Mosser J (2007) Common variants in the BMP2, BMP4, and HJV genes of the hepcidin regulation pathway modulate HFE hemochromatosis penetrance. Am J Hum Genet 81: 799-807

Nadeau JH (2001) Modifier genes in mice and humans. Nat Rev Genet 2: 165-174

Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74: 765-9

Okun MS, Thommi N (2004) Americo Negrette (1924 to 2003): Diagnosing Huntington disease in Venezuela. Neurology 63: 340-343

Perdry H, Maher BS, Babron MC, McHenry T, Clerget-Darpoux F, ML M (2007) An ordered subset approach to including covariates in transmission disequilibrium. BMC Proceedings In press

Perdry H, Babron M-C, Clerget-Darpoux F (2007) The Ordered Transmission Disequilibrium Test: detection of modifier genes. Submitted

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat

Genet 38: 904-9
Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits
10.1093/bioinformatics/19.1.149. Bioinformatics 19: 149-150
Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, Benaiche A, Isnard R, Dubourg O, Burban M, Gueffet J, Millaire A, Desnos M, Schwartz K, Hainque B, Komajda M (2003) Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation 107: 2227-32
Rogus J, Krolewski A (1996) Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. Am J Hum Genet 59: 1376-81
Schluchter MD, Konstan MW, Drumm ML, Yankaskas JR, Knowles MR (2006) Classifying severity of cystic fibrosis lung disease using longitudinal pulmonary function data. Am J Respir Crit Care Med 174: 780-6
Shattuck-Eidens D, McClure M, Simard J, Labrie F, Narod S, Couch F, Hoskins K, Weber B, Castilla L, Erdos M (1995) A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. JAMA 273: 535-41
Silvers WK (1979) The Coat Colors of Mice: A Model for Mammalian Gene Action and Interaction. Springer Verlag, New York.
Slavotinek A, Biesecker LG (2003) Genetic modifiers in human development and malformation syndromes, including chaperone proteins. Human Molecular Genetics 12: R45-R50
Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506-16
Spirito P, Marron BJ (1990) Relation between extent of left ventricular hypertrophy and occurrence of sudden cardiac death in hypertrophic cardiomyopathy. J Am Coll cardiol 15: 1020-1027
Steffens M, Lamina C, Illig T, Bettecken T, Vogler R, Entz P, Suk EK, Toliat MR, Klopp N, Caliebe A, Konig IR, Kohler K, Ludemann J, Diaz Lacava A, Fimmers R, Lichtner P, Ziegler A, Wolf A, Krawczak M, Nurnberg P, Hampe J, Schreiber S, Meitinger T, Wichmann HE, Roeder K, Wienker TF, Baur MP (2006) SNP-based analysis of genetic substructure in the German population. Hum Hered 62: 20-9
Suzuki T, Kashiwagi A, Mori K, Urabe I, Yomo T (2004) History dependent effects on phenotypic expression of a newly emerged gene. Biosystems 77: 137-141
Thakkinstian A, Han P, McEvoy M, Smith W, Hoh J, Magnusson K, Zhang K, Attia J (2006) Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration. Hum Mol Genet 15: 2784-90
Vanscoy L, Blackman S, Collaco J, Bowers A, Lai T, Naughton K, Algire M, McWilliams R, Beck S, Hoover-Fong J, Hamosh A, Cutler D, Cutting G (2007) Heritability of lung disease severity in cystic fibrosis. Am J Respir Crit Care Med 175: 1036-43
Wang G-S, Cooper TA (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. 8: 749-761
Watkins H, Rosenzweig A, Hwang D, Levi T, McKenna W, Seidman C, Seidman J (1992) Characteristics and prognostic implications of myosin missense mutations in familial hypertrophic cardiomyopathy. N Engl J Med 326: 1108-14

Wexler N, Lorimer J, Porter J, Gomez F, Moskowitz C, Shackell E, Marder K, Penchaszadeh G, Roberts S, GayÃ¡n J, Brocklebank D, Cherny S, Cardon L, Gray J, Dlouhy S, Wiktorski S, Hodes M, Conneally P, Penney J, Gusella J, Cha J, Irizarry M, Rosas D, Hersch S, Hollingsworth Z, MacDonald M, Young A, Andresen J, Housman D, De Young M, Bonilla E, Stillings T, Negrette A, Snodgrass S, Martinez-Jaurrieta M, Ramos-Arroyo M, Bickham J, Ramos J, Marshall F, Shoulson I, Rey G, Feigin A, Arnheim N, Acevedo-Cruz A, Acosta L, Alvir J, Fischbeck K, Thompson L, Young A, Dure L, O'Brien C, Paulsen J, Brickman A, Krch D, Peery S, Hogarth P, Higgins DJ, Landwehrmeyer B (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proc Natl Acad Sci U S A 101: 3498-503

Witsch-Baumgartner M, Schwentner I, Gruber M, Benlian P, Bertranpetit J, Bieth E, Chevy F, Clusellas N, Estivill X, Gasparini P, Giros M, Kelley RI, Krajewska-Walasek M, Menzel J, Miettinen TA, Ogorelkova M, Rossi M, Scala I, Schinzel A, Schmidt K, Schonitzer D, Seemanova E, Sperling K, Syrrou M, Talmud P, Wollnik B, Krawczak M, Labuda D, Utermann G (2007) Age and origin of major Smith-Lemli-Opitz Syndrome (SLOS) mutations in European populations. J Med Genet

Wolf U (1997) Identical mutations and phenotypic variation. Hum Genet 100: 305-21

Xu Z, Kaplan NL, Taylor JA (2007) Tag SNP selection for candidate gene association studies using HapMap and gene resequencing data. Eur J Hum Genet 15: 1063-70

Zielenski J, Corey M, Rozmahel R, Markiewicz D, Aznarez I, Casals T, Larriba S, Mercier B, Cutting GR, Krebsova A, Macek M, Jr., Langfelder-Schwind E, Marshall BC, DeCelie-Germana J, Claustres M, Palacio A, Bal J, Nowakowska A, Ferec C, Estivill X, Durie P, Tsui LC (1999) Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13. Nat Genet 22: 128-9

Figures and Tables

Table 1: Occurrence of meconium ileus (MI) in cystic fibrosis: linkage analysis with the 19q13 region (Zielenski et al. 1999)

Sibpairs[a]	**IBD=0**	**IBD=1**	**IBD=2**	**Total**	**Conformity test chi^2 (p-value)**[b]
both MI	0 (0%)	3 (42.8%)	4 (57.2%)	7	4.7 (0.09)
neither MI	22 (19.6%)	59 (52.7%)	31 (27.7%)	112	4.8 (0.41)
discordant	20 (60.6%)	11 (33.3%)	2 (6.1%)	33	23.3 (8.7 10^{-6})
Total	42 (27.6%)	73 (48.1%)	37 (24.3%)	152	0.6 (0.75)

[a] 152 sibpairs homozygous for the DeltaF508 mutation of the CFTR gene were sampled and grouped together depending on whether both sibs had a MI, none had a MI or were discordant for MI. The number of sibpairs sharing 0, 1 or 2 alleles identical by descent (IBD=0, 1 or 2) are reported with the respective proportions in parentheses.

[b] Results of the test of conformity of the observed IBD distribution with the expected proportions under the null hypothesis of no linkage (1/4, 1/2, 1/4)

Table 2: Advantages and limitations of the linkage strategy in searching for modifier genes

Advantages	**Limitations**
+ Robustness - to allelic heterogeneity - to population stratification + Coverage of the genome with a limited number of markers - 400 microsatellite markers - 6,000 to 10,000 SNPs + Heritability estimates	+ Power decreases with increasing risk allele frequency: limited power for allele frequency greater than 0.3 + Candidate regions identified through linkage studies are very large, spanning several megabases, and include several genes + Difficulty to obtain large samples of related patients with the phenotype of interest

Box 1 Multiple testing issues in association studies

When the number of markers tested increases, it is necessary to take into account the fact that multiple tests are performed. That is, if one defines as significant any tests with a p-value below 5%, and only one test is performed, the probability of incorrectly rejecting the null hypothesis and concluding in an association is 5%. If N tests are performed, this probability is increased proportionally: when N=100 there will be on average 5 false-positive results and if N=100,000 this number will be 5000. To limit the proportion of false positives, corrections can be made for multiple testing. One of the most commonly is the Bonferroni correction. To ensure a global type-one error of 5% for N tests, it considers significant only tests with a p-value of less than 0.05/N. This correction is conservative when tests are not independent. Other less conservative corrections have been proposed, which take into account the correlation that may exist between markers through linkage disequilibrium (Li and Ji 2005; Nyholt 2004). Even after accounting for linkage disequilibrium, the significance level for ensuring a genome-wide type 1 error of 5% remains on the order of 10^{-7}.

Figure Legends

Figure 1 Sample sizes required to reach 80% power to detect an association when different number of markers are tested.

The total number of patients (white bars) or case-parent trios (black bars) required to reach a power of 80% are shown. A genetic factor is assumed, with an effect similar to that of the C509T polymorphism of the TGFβ1 gene in CF-associated lung disease: a 0.34 allele frequency acting recessively with an odds ratio of 2 (Drumm et al. 2005). For the patient samples, it is assumed that the phenotype of interest is present in 50% of the individuals and absent in the remaining 50%. For the trio samples, all patients are assumed to have the phenotype. The program Genetic Power Calculator (Purcell et al. 2003) was used with a Bonferroni correction for multiple testing.

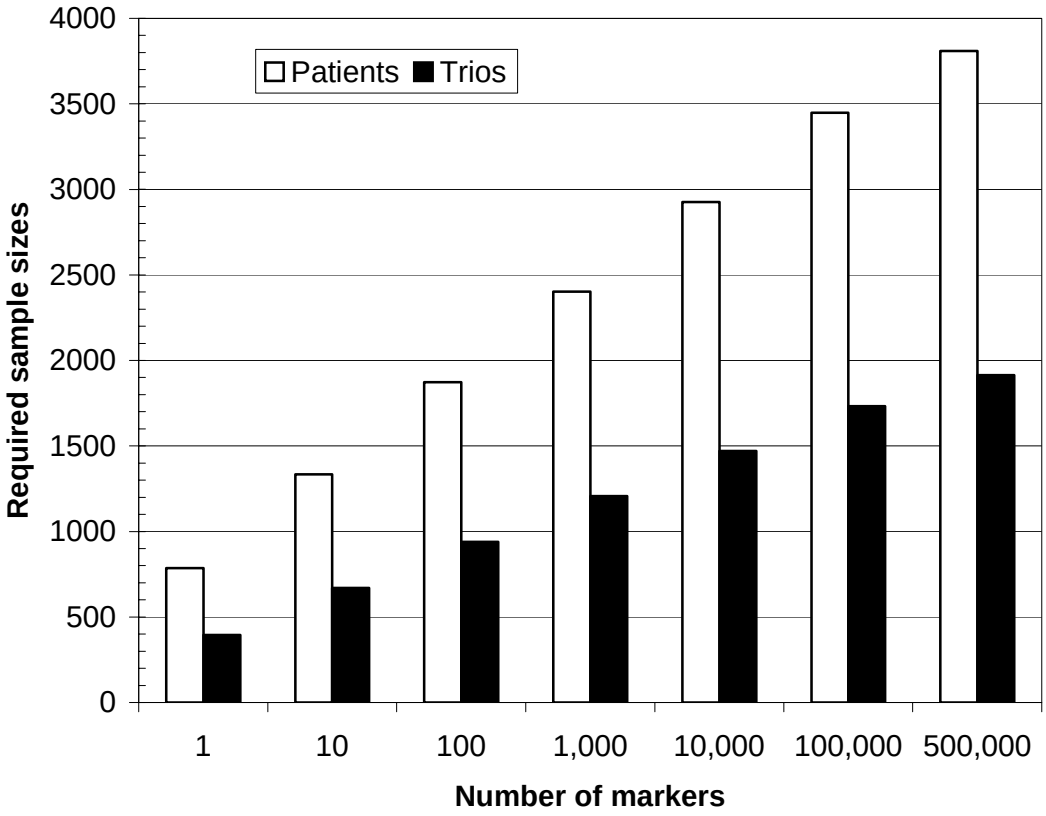

Figure 1