PRINT: ISSN 0972-3757 ONLINE: ISSN 2456-6330
IJHG 1(1): 55-63 (2001)
DOI: 10.31901/24566330.2001/01.01.06

Why Do Complex Traits Resist DNA Analysis ?

F. Clerget-Darpoux, H Selinger-Leneman and M-C Babron

INSERM U.535., 80 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex, France
Fax: 33 1 49 59 53 31; E-mail: clerget@ccr.jussieu.fr

KEY WORDS Multifactorial disease; genome screen; candidate gene; linkage; association.

ABSTRACT The etiology of many human diseases is complex. The number of factors involved, the importance of their individual effect and the level of heterogeneity are unknown. To unravel the genetic etiology of these diseases, a popular strategy is to search for genetic risk factors by testing linkage systematically over the entire genome. The power of such an approach very much depends on the unknown characteristics of the genetic factors and the main difficulty is to establish a good trade-off between false positives and false negatives. Avoiding a high rate of false positives will lead to low power for detecting a genetic factor with a moderate effect. In addition, when a genetic factor is detected, the precise localisation of this factor is generally not possible under this method. To go further in the identification of factors involved in the disease process, one has to set up a candidate gene strategy. If the candidate gene polymorphism is not directly available, information may be obtained through closely linked markers. In such a case, we may expect, in addition to linkage, allelic association between the candidate gene and marker alleles. However, the choice of candidate genes as well as markers showing allelic association is not simple. Some have proposed to consider every gene as a candidate and to screen the whole genome using the Transmission Disequilibrium Test. However, the problems of multiple testing and of heterogeneity between populations may cripple this approach. Despite extraordinary advances in molecular and computer techniques, it is likely that for multifactorial diseases the only genetic risk factors that can be detected are those with fairly strong effect. Even in this case, it is important to design strategies that increase the power of detection.

INTRODUCTION

Advances in molecular and computer techniques have spurred the search for genes involved in human diseases. Most human diseases are multifactorial in the sense that they are due to several risk factors, both genetic and environmental. We face the difficulty of untangling genetic risk factors from familial environment and from cultural factors. The segregation of a trait in families due to cultural background may very well mimic a genetic transmission. This was nicely illustrated by McGuffin and Huckle (1990) showing that the familial segregation of "attending Medical School" very well fits a recessive gene transmission.

In many studies of multifactorial diseases, the importance of genetic factors is quantified by the increased risks for relatives of affected as compared to the general population (Risch 1987). As a matter of fact, these values are only measurements of familial aggregation. They do not differentiate among the respective contribution of genetic factors, familial or cultural environment. Moreover, based on some simplistic hypotheses such as multiplicative and equal effect of each genetic factor, some research studies even go as far as postulating the number of genetic factors to be detected. In reality, we unfortunately do not know how many genetic factors are involved, how important their individual effects are, how they interact together as well as with the environmental factors. The power to detect a gene involved in the disease (susceptibility disease gene) actually depends on the difference in the genotype distribution between affected and unaffected individuals in the studied population. This difference may vary from one population to another. Heterogeneity between populations is an additional difficulty in the study of multifactorial diseases.

In a given population, a genetic risk factor may be detected using genetic markers through correlation between the disease status and marker genotypes. Such a correlation exists either at the population level (tested by association studies) or at the family level (tested by linkage studies). The studies may be performed either systematically on the whole genome or by focusing on candidate genes. We will discuss in this paper the respective pros and cons of these two approaches. Besides, we will show that the difficulty to unravel the genetic component of a multifactorial disease does not reside, as some believe, in the choice

of a strategy but mainly in the uncertainty on what we are looking for.

SYSTEMATIC SCREENING OF THE GENOME BY LINKAGE TESTS

Linkage between a genetic marker and a disease means that the disease and marker transmissions in families are not independent. This implies the presence of a disease susceptibility gene in the marker region.

For a long time, the most widely used method has been the lod score method, proposed by Morton (1955). The method was meant to apply to traits with known mode of inheritance and it has been very successful to locate genes of diseases with Mendelian inheritance. These successes have generated great enthusiasm among genetic epidemiologists and promoted the naive idea that systematic screening of the genome by Morton's lod score method would allow determining the genetic basis of any human disease.

In Morton's lod score test, the key parameter is the recombination fraction which measures the proportion of recombined gametes from parents to their children. Estimation is thus possible when, for each family member, the genotype is known at each locus. If a disease is controlled by a gene at one of these loci (disease locus), the computation of a lod score for a given family requires to consider all possible genotypic configurations at the disease locus and to compute the probabilities of these configurations given the phenotypic information for the disease.

For a multifactorial disease, the underlying model is unknown. If its specification is incorrect, the recombination fraction, will not be correctly estimated (Clerget-Darpoux et al. 1986) and the true location of the risk factor may be wrongly excluded (Clerget-Darpoux and Bonaïti-Pellié 1993). To address the case of linkage studies for diseases with unknown mode of inheritance, we suggested (Clerget-Darpoux et al. 1986) extending the lod score function to a so-called mod score function. In the mod score, the variables are both the recombination fraction and the disease model parameters. Since it is asymptotically maximum for the true disease model, the power to detect linkage through mod score will be highest when the space of models where the maximisation is performed includes the true model. As the disease model may be very complex, it may require many parameters. On the other hand, one must avoid overparametrisation. This overparametrisation leads to many different parameter sets giving the same mod score. Consequently, the advantages of such an approach compared to other model free statistics are disputable (Clerget-Darpoux 2000).

An alternative and very popular strategy, which is applied to test linkage when the mode of inheritance is unknown, is the sib-pair Maximum Lod Score (MLS) test proposed by Risch (1990). In this approach, the variable of interest is no more the recombination fraction, as in Morton's lod score, but the distribution (z_2,z_1,z_0) measuring the proportion of times affected sibs share 2,1 or 0 marker alleles Identical By Descent (IBD). Under the null hypothesis of no linkage, the expected distribution is (0.25, 0.50, 0.25). Under the alternative hypotheses of linkage, it differs from (0.25, 0.50, 0.25) and complies with the triangle constraints (Suarez 1978; Holmans 1993) with $z_0 + z_1 + z_2 = 1$ and $2z_0 \leq z_1 = 0.5$. Table 1 compares the Morton's and Risch's lod score statistics.

Of course the statistical properties (type I error, power) also differ. Holmans (1993) and Eichenbaum-Voline et al. (1997) studied the properties of the MLS in the case of a single marker. The power to detect linkage depends not only on the characteristics of the risk factor but also on the informativity of the marker. The information may be increased by simultaneously considering linked markers (Kruglyak and Lander 1995a).

Table 1: Lod score statistics for testing linkage $Z(H1) = \log_{10} [L(H1) / L(H0)]$

	Variable	*H0*	*H1*	
Morton (1955)	Recombination fraction θ	1/2	$0 = \theta < 1/2$	Specification of model at disease locus
Risch (1990)	IBD vector $Z = (z_2,z_1,z_0)$	(1/4, 1/2, 1/4)	$Z \neq (¼,½,¼)$ + $2 z_0 \leq z_1 \leq 0.5$	"model free"

In a systematic screening approach, linkage is tested with many markers and the MLS threshold which corresponds to a type I error must be evaluated according to the set of tested markers. Lander and Kruglyak (1995) considered the situation of testing linkage at any point of the genome (full density and full polymorphism of the marker map). In this case, they showed that for a type I error of 5%, the MLS threshold is 4. For a less dense map, markers spaced 10cM, and for the same type I error, the threshold is 3.07 (Quesneville, personal communication). Consequently, the power to detect an existing risk factor depends on the underlying model for this risk factor, on the map characteristics (density and informativity of each marker) and on the sib pair sample size.

We have studied the properties of MLS distribution in four examples of susceptibility genes having a demonstrated involvement in multifactorial diseases (Table 2).

- M1 corresponds to the APOE gene in Alzheimer's disease (AD). The ApoE gene has three alleles ε2, ε3 and ε4, with respective frequency of 5%, 80% and 15%. The penetrance of the ε4ε4 genotype (risk for an ε4ε4 individual) and of the ε4ε3 genotype has been estimated to be 11.9 and 2.2 times the one of the ε3ε3 genotype respectively (Bickeböller et al. 1988). For such relative risks, the expected IBD distribution in affected sib pairs is z_2 =0.36, z_1 = 46,z_0 = 0.18.
- M2 corresponds to the insulin gene in Insulin Dependent Diabetes Mellitus (IDDM). The frequency of an allele of a flanking polymorphism of the insulin gene was shown to be increased in IDDM (0.85) as compared to the general population (0.70) (Bell et al. 1984). However, the IBD distribution observed on 95 affected sib pairs available in the Genetic Analysis Workshop 5 (Spielman et al. 1989) was z_2 = 0.26, z_1 = 0.50, z_0 = 0.24 and did not give evidence for linkage in this region. This is a very nice illustration that a susceptibility factor may be more detectable through association information than through linkage information (Cox and Spielman 1989) and, as an anecdote, gave rise to the Transmission Disequilibrium Test (TDT) (Spielman et al. 1993). This is particularly true for the effect of a very frequent allele with a dominant effect, as it is the case for the insulin gene in IDDM.
- M3 corresponds to the role of HLA in Multiple Sclerosis (MS). In this example, the implication of HLA is again better demonstrated by the strong association of the disease with the DR15 antigen than by linkage studies (Yaouancq et al. 1999). In 116 French sib pairs affected by MS and described in Reboul et al. (2000), the IBD distribution observed for HLA is z_2 = 0.34, z_1 = 0.48, z_0 = 0.18.
- M4 corresponds to the role of HLA in IDDM. In contrast with the above examples, the IBD distribution is very distorted from (0.25, 0.50, 0.25). Only 15 affected sib pairs were sufficient for Cudworth and Woodrow (1975) to conclude to linkage. In a very large sample pooled from literature, the IBD distribution for HLA in IDDM is z_2 = 0.58, z_1 = 0.36, z_0 = 0.06.

Table 2: Four examples of IBD distribution corresponding to a susceptibility factor in a multifactorial disease

	IBD distribution in affected sib pairs		
	Z2	*Z1*	*Z0*
M1 = APOE in Alzheimer's disease	0.36	0.46	0.18
M2 = Insulin in Type I diabetes	0.26	0.50	0.24
M3 = HLA in Multiple Sclerosis	0.34	0.48	0.18
M4 = HLA in Type I diabetes	0.58	0.36	0.06

In these four examples, we have evaluated the power of linkage detection by a systematic screening on 100 affected sib pairs and two different marker maps 10cM and 2cM between markers respectively. In both cases, the markers have 10 equifrequent alleles and the risk factor is strictly linked with one marker. The MLS distribution at each marker was obtained by simulation using GENSIM (Kruglyak et al. 1996) for the two marker maps. For the ten thousand replicates of 100 affected sib pairs, the maximum MLS (MaxMLS) and its position on the map were recorded.

Table 3 gives, for each example and for each map, the power to detect the risk factor (proportion of times the MaxMLS exceeds the threshold). For the first three examples, the power is

Table 3: Power of the MLS statistic in a sample of 100 affected sib pairs for a type I error of 5%

	10cM marker map *Prob (MaxMLS > 3.07)*	*2cM marker map* *Prob (MaxMLS > 4)*
M1 (APOE/AD)	18%	8%
M2 (INS/IDDM)	0%	0%
M3 (HLA/MS)	15%	8%
M4 (HLA/IDDM)	100%	100%

very low. It is even null for the detection of the insulin gene in IDDM. This is fully consistent with the result of Concannon et al. (1998) who obtained an MLS of 0.60 in the insulin region for 607 affected sib pairs. As seen in table 3, genetics is now faced with an amazing paradox. The denser the marker map, the lower the power to detect a susceptibility gene, for a given type I error ! This is because the threshold criterion for linkage conclusion increases with the number of tested markers. This illustrates the difficulty of the trade-off between false positives and false negatives in the study of a multifactorial disease.

Table 4: 95% interval for the MaxMLS values

	10cM marker map		*2cM marker map*	
M1 (APOE/AD)	[0.30	4.90]	[0.54	5.07]
M2 (INS/IDDM)	[0.00	2.02]	[0.00	2.05]
M3 (HLA/MS)	[0.35	4.62]	[0.40	5.03]
M4 (HLA/IDDM)	[7.53	19.70]	[10.0	23.30]

The 95% intervals of the MaxMLS are given in table 4. Note the huge variation of the MaxMLS from one replicate to another. In the example of APOE in AD and for a 10cM map, the probability for a MaxMLS to be greater than the threshold is the same as the probability to be smaller than 1. This may explain the weak power of replicating a positive result in an independent sample and the apparent discordance of results between samples.

Lastly, figure 1 shows the variation of the position of the MaxMLS retained for linkage (over the threshold) in the M3 and M4 examples. For M3 (HLA in MS), the position varies: only one third of MaxMLS are at the correct location and 14% of MaxMLS are obtained at more than 10cM from the true position. This indicates how limited the resolution of linkage analysis in the study of a multifactorial disease is. Hauser et al. (1996) proposed to use the affected sib pair information on two markers for a simultaneous estimation of the IBD distribution and of the recombination fraction. In fact, the likelihood surface for such a parametrisation may be very flat as indicated by our study and by the one of Kruglyak and Lander (1995b).

Contrasting with the first three examples, the power of detecting HLA in IDDM is 100% and although the variability of MaxMLS remains very large in that case, its value is always over the threshold. In addition, the MaxMLS is obtained 89% times at the true location. This shows that no general conclusion about linkage genome scans in multifactorial diseases can be given. For those susceptibility factors inducing a large IBD

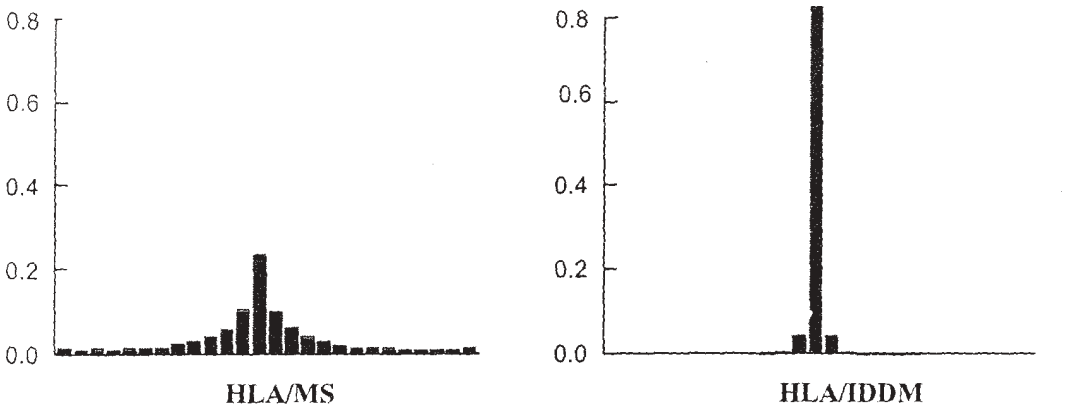

Fig. 1. Position of MaxMLS for the 2cM map when linkage is detected. The true position of the risk factor is in the middle of the map

deviation from (0.25,0.50,0.25), then the power of detection is good and the results are consistent from one study to another. Unfortunately when studying a multifactorial disease, it is impossible to know a priori whether some susceptibility factor can be detectable by the systematic linkage approach. Among the many susceptibility loci reported for IDDM, it is very difficult to distinguish those which may be false from the true ones. Except HLA (IDDM1), INS (IDDM2) and CTLA4 (IDDM12), all the others have been suggested through genome screens (see table 5, from Concannon et al. 1998). Except for IDDM15, the analysis of pooled samples performed by Concannon et al. did not confirm any of the suggested loci in spite of the very large sample sizes. Besides, it is possible that IDDM7, 12 and 13 – all located on chromosome 2q, 16cM apart – reflect the same susceptibility locus. The same may be said for IDDM5, 8 and 15 on chromosome 6q.

Other model-free statistics than MLS would lead to similar conclusions. In particular, the uncertainly on the location is always correlated to the uncertainty on the mode of inheritance.

It is sometimes argued that studying extended pedigrees may provide more information that affected sib pairs. Model free linkage statistics have been developed that use information on more distant relatives than sibs (Weeks and Lange 1988; Bishop and Williamson 1990; Kruglyak et al. 1996). In principle, one should gain power by considering extended genealogies rather than nuclear families. However, several drawbacks must be considered. First, the marker typing of several individuals of the genealogies are likely be missing. In this case, applying any linkage test requires the specification of marker allele frequencies. Error on these allele frequencies may dramatically increase the rate of false positives. It is a major problem for methods such as the APM method (Weeks and Lange 1988) where unaffected members are systematically untyped (Babron et al. 1993) or in the analysis of extended pedigrees where the first generations are missing (Ott 1992; Freimer et al. 1993). In addition, large genealogies with many affected members are rare, not representative of the disease segregation and it is tempting to pool them from different populations which can be heterogeneous in the marker allele frequencies (Margaritte-Jeannin et al. 1997). Besides, for diseases with large clinical spectrum, it is more difficult to select affected persons with a homogeneous phenotype, thus increasing even more the genetic heterogeneity in the studied sample. Lastly, pedigrees with many affected may correspond to a different genetic determinism, even sometimes to a monogenic subentity.

Table 5: Second generation screen of IDDM susceptibility loci by Concannon et al. (1998)

Susceptibility locus	*Location*	*#of pairs*	*MLS*
IDDM1 (HLA)	6p21	618	32.5
IDDM2 (INS)	11p25	607	0.60
IDDM3	15q26	506	0.03
IDDM4	11q13	778	0.43
IDDM5	6q25	852	1.46
IDDM6	18q21	302	0.00
IDDM7	2q31	653	0.72
IDDM8	6q27	730	1.14
IDDM9	3q21	543	0.23
IDDM10	10cen	609	0.40
IDDM11	14q24	433	0.28
IDDM12 (CTLA4)	2q33	585	0.84
IDDM13	2q34	418	0.36
IDDM15	6q21	772	3.51

In conclusion, only those factors with a fairly strong effect (inducing a large IBD deviation) will be detectable in a consistent way by linkage genome scans. For factors with moderate effect, we may obtain discordant results from different samples and other information have to be used to confirm the existence of a susceptibility disease gene. In any case, genome scans indicate - at best - a region in which a risk factor lies. To further identify the factors involved in the disease process, one has to set up a candidate gene strategy (Clerget-Darpoux 1998).

CANDIDATE GENE STRATEGY

Contrarily to systematic screening, one may focus on specific genes called “candidate genes”. These genes may be chosen as candidates because of their functionality. For example, the HLA and insulin genes are good candidates for IDDM. Candidates may also be chosen for their position, because they are located either in a region showing linkage or in a region homologous to one identified in another species. Here, the question addressed is not to localise a “disease gene”

but to demonstrate the role of the considered candidate gene in the disease and to calculate the risk for an individual to develop the disease according to the available information on this candidate gene.

When the functional polymorphism is available, the relative risk associated to each genotype may be estimated through random case and control samples as illustrated for ApoE in Alzheimer's disease (Bickeböller et al. 1988). The six ApoE genotypes correspond to different risks of developing Alzheimer's disease.

However the functional genotypes of the candidate gene are not always observable. The information on a genetic marker situated at or near the candidate gene locus can then be used. Two kinds of information may be considered:

- at the family level, linkage between the disease status and the marker genotype
- at the population level, allelic association between the marker and the disease loci (often improperly called linkage disequilibrium)

Most investigation methods on the role of a candidate gene use only one type of information (population or family). However, simultaneously taking into account the information on both the familial segregation and the population association of the marker with the disease increases the power of detecting the involvement of a candidate gene as a risk factor. This is the strength of the transmission/disequilibrium test (TDT) (Spielman et al. 1993).

This test applies to unrelated affected individuals who, as well as their two parents, have been typed for a biallelic marker (M1, M2). It compares the number of times the alleles M1 and M2 are transmitted and untransmitted from heterozygous parents M1M2 to their affected offspring. No difference is expected either in the absence of allelic association or in the absence of linkage. Consequently, a difference implies both linkage and association. The principle of this test has been extended to the case of multiallele marker loci (Sham and Curtis 1995; Bickeböller and Clerget-Darpoux 1995). Furthermore, the alleles not transmitted to the affected child(ren) create an internal family-based control group permitting estimation of relative risk (Falk and Rubinstein 1987; Terwilliger and Ott 1992; Knapp et al. 1993). A comprehensive review of these methods may be found in Schaid (1996).

The relative power of the TDT and IBD tests on sib pairs depends on both the underlying genetic model and the available family data. When allelic association is strong, the TDT can be more powerful than the IBD test (Clerget-Darpoux et al. 1995). This is well illustrated by showing the role of insulin in IDDM (Spielman et al. 1993) or of an HLA factor in multiple sclerosis (Yaouancq et al. 1997). In contrast, it is important to stress that the power of TDT is null in the absence of allelic association.

Modelling the effect of a gene, after its involvement has been shown, is a neglected step while it may considerably increase the power to detect other risk factors. In that case also, one may take advantage of the two types of information (allelic association and linkage) to evaluate the risk associated to each genotype. The Marker Association Segregation Chi-square (MASC) method (Clerget-Darpoux et al. 1988) was designed to achieve such a goal. It has been applied to model the role of HLA in IDDM (Clerget-Darpoux et al. 1991), in rheumatoid arthritis (Dizier et al. 1994; Génin et al. 1998) and in coeliac disease (Clerget-Darpoux et al. 1994; Bouguerra et al. 1999).

Candidate gene strategy appears thus to have several tangible benefits over systematic linkage tests on the whole genome. First, focusing on a small number of genes limits the number of tests. Second, information may be gained through allelic association, the ideal situation being complete allelic association i.e. working on the alleles of the functional factor itself. Unfortunately, selecting good candidate genes or markers showing allelic association is no simple matter.

The candidate genes may be chosen for their function. Such a choice requires a good knowledge of the disease physiopathology. However, this situation is rare since genetic studies on multifactorial diseases are mainly designed to better understand the disease process. Consequently, as proposed by Risch and Merinkangas (1997), it may be tempting to consider any gene as a candidate. We then face again the drastic problem of multiple testing. Only those factors

for which the genotype distribution is extremely different between patients and controls may be detected. If by chance, linkage studies or homology with other species or any other means suggest the presence of a risk factor in a given region, one may confine the search to the genes of this region. Even in that case, the number of genes to consider is extremely high since the length of the region to consider is large.

Choosing intragenic polymorphic markers is also a difficult task. The polymorphism of some genes is extremely high while it is very poor for others. Moreover, the allelic association between available polymorphisms and the alleles of interest (the functional ones) is never guaranteed even within the same gene. Allelic association results from the complex and unique history of the population under study as well as from stochastic events. It may thus differ from one population to another. Consequently, it may be difficult to replicate results in candidate gene studies. For a same candidate gene marker and a same sample size, the TDT could be significant in a given population but not in another one. This is the case, for example, for the studies of a CTLA4 exon 1 49 A/G polymorphism in auto-immune diseases (Table 6). In type I diabetes (see for review Nistico et al. 1996), the frequency of allele G is significantly higher in IDDM patients than in controls for Mediterranean Europeans and Mexican Americans. This is not found in Sardinians and in Caucasian Americans. For coeliac disease, the frequency of allele A is increased in the Scandinavian and French patients but not in the Italian and Tunisian ones (Clot et al. 2000). Given the above results, it becomes difficult to discriminate between different interpretations:

- false positive results ?
- heterogeneity in the genetic determinism of the disease ?
- heterogeneity in the allelic associations between the studied populations ?

Table 6: Associated allele of the CTLA4 exon 1 49 A/G polymorphism in different populations

Coeliac disease		*Type I diabetes*	
French	A	Mediterranean European	G
Scandinavian	A	Mexican American	G
Italian	ns	Caucasian American	ns
Tunisian	ns	Sardinian	ns

Going from demonstrating the role of a gene to understanding its functionality is the next step for unravelling the genetic component of a disease. While the degree of mapping resolution is of course finer in a candidate gene approach than in a random marker linkage approach, it is not sufficient to allow for the identification of a functional polymorphism. The susceptibility to a disease may result not only from a single variant in a gene but also from a complex interaction of several intragenic variants. Moreover, several good candidates may be clustered and the difficulty is then to discriminate between them. Allelic associations between two loci is created through mutation, population admixture, selection. Allelic associations and physical distances do not correlate significantly over small regions (Jorde et al. 1994). A prerequisite to progress in genetic epidemiology is therefore a good knowledge of population characteristics through population genetics studies.

PERSPECTIVES

Mapping disease loci that predispose to multifactorial diseases is the present challenge of geneticists. Most of the time, however, the euphoria of linkage findings is followed by numerous failures to confirm them in independent samples. The uncertainty on both the number of genetic factors involved and the importance of their marginal effect makes it impossible to predict success. The best possible approach is thus to design the most efficient strategies, keeping in mind that only factors with fairly strong effect, i.e. those for which there is a large difference in the patient and control genotype distribution, will be detected. Genetic risk factor for a multifactorial disease and rare morbid mutation for monogenic disease are entirely different concepts. Shifting from the monogenic disease paradigm is one of the most difficult steps.

The real difficulty is to increase the power of genetic risk factor detection without increasing the rate of false positive. This is not simple since with the increasing knowledge of the genome one is tempted to perform as many tests as feasible.

First, genetic heterogeneity in the data should

be minimised. The problem is that we do not know how genetic heterogeneity correlates to whatever is observable. Thus, the best is to limit heterogeneity on different criteria: clinical, population, familial recurrence of the disease. Large extended pedigrees with many affected, which are very informative when dealing with a monogenic disease, may be source of heterogeneity when studying a multifactorial disease. It is also important to take into account already known co-variables and risk factor(s) (genetic and/or environmental). This may be fundamental in case of interaction. For example conditioning on HLA may greatly increase the power when studying non HLA genetic risk factors in an auto-immune disease. The degree of exposure to the infectious agent is clearly necessary when studying the susceptibility to an infectious disease.

Focusing on a limited number of well chosen candidate genes limits the number of tests and, as such, increases detection power. Studies of the disease in animal species, studies of traits correlated with the disease but with a simple mode of inheritance, or physiopathological studies may help in this choice.

Working in particular population structures may also help. In particular, disease studies may be performed in isolated populations with recent founders and new methods (Bourgain et al. 2000) are currently developed which benefit from the particularities of these populations. In addition to a lower degree of heterogeneity, allelic association is often present between loci which are not necessarily very close. However, here again, one must escape from the paradigm of monogenic disease and must banish the concept of a unique ancestral mutation.

Unravelling the genetic component of multifactorial diseases is not a simple endeavour and requires developing the appropriate strategies. This implies tight collaborations among clinicians, physiopathologists, geneticists (molecular, population, epidemiologist) and biostatisticians.

ACKNOWLEDGMENTS

We are grateful to Christine Vaillant for technical assistance.

REFERENCES

Babron MC, Martinez M, Bonaïti-Pellié C, Clerget-Darpoux F 1993. Linkage detection by the affected-pedigree-member: What is really tested ? *Genet Epidemiol,* **10**: 389-394.

Bell GI, Horita S, Karam JH 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes,* **33**: 176-183.

Bickeböller H, Clerget-Darpoux F 1995. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. *Genet Epidemiol,* **12**: 865-870.

Bickeböller H, Campion D, Brice A, Amouyel P, Hannequin D, Didierjean O, Penet C, Martin C, Pérez-Tur J, Michon A, Dubois B, Ledoze F, Thomas-Antérion C, Pasquier F, Puel M, Demonet JF, Moreaud O, Babron MC, Meulien D, Guez D, Chartier-Harlin MC, Frebourg T, Agid Y, Martinez M, Clerget-Darpoux F 1997. Apolipoprotein E and Alzheimer disease : Genotype-specific risks by age and sex. *Am J Hum Genet,* **60**: 439-446.

Bishop DT, Williamson JA 1990. The power of identity-by-state methods for linkage analysis. *Am J Hum Genet,* **46**: 254-265.

Bouguerra F, Babron MC, Eliaou JF, Debbabi A, Clot J, Khaldi F, Clerget-Darpoux F 1997. Synergistic effect of two HLA heterodimers in the susceptibility to celiac disease in Tunisia. *Genet Epidemiol,* **14**: 413-422.

Bourgain C, Génin E, Quesneville H, Clerget-Darpoux F 2000. Search for multifactorial disease susceptibility genes in founder populations. *Ann Hum Genet,* **64**: 255-265.

Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J 1986. Effects of misspecifying genetic parameters in lod score analysis. *Biometrics,* **42**: 393-399.

Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J 1988. A new method to test genetic models in HLA associated diseases : the MASC method. *Ann Hum Genet,* **52**: 247-258.

Clerget-Darpoux F, Babron MC, Deschamps I, Hors J 1991. Complementation and maternal effect in insulin dependent diabetes. *Am J Hum Genet,* **49**: 42-49.

Clerget-Darpoux F, Bonaïti-Pellié C 1993. An exclusion map covering the whole genome: a new challenge for genetic epidemiologists ? *Am J Hum Genet,* **52**: 442-443.

Clerget-Darpoux F, Bouguerra F, Kastally R, Sémana G, Babron MC, Debbabi A, Bennaceur B, Eliaou JF 1994. High risk genotypes for celiac disease. *CR Acad Sci Paris,* **317**: 931-936.

Clerget-Darpoux F, Babron MC, Bickeböller H 1995. Comparing the power of linkage detection by the transmission disequilibrium test and the identity by descent test. *Genet Epidemiol,* **12**: 583-588.

Clerget-Darpoux F 1998. Overview of strategies for complex genetic diseases. *Kidney Int,* **53**: 1441-1445.

Clerget-Darpoux F 2000. Extension of the Lod Score: the Mod Score. In: DC Rao, M Province (Eds.): *Genetic Dissection of Complex Traits.* New York: Academic Press pp. 115-124.

Clot F, Babron MC 2000. Genetics of celiac disease. *Mol Genet Metabolism,* **71**: 76-80.

Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst

B, Morrison VA, Stirling b, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS 1998. A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nature Genet,* **19**: 292-296.

Cox N, Spielman R 1989. The insulin gene and susceptibility to IDDM. *Genet Epidemiol,* **6**: 65-69.

Cudworth AG, Woodrow JC 1975. Evidence for HLA-linked genes in 'juvenile' diabetes mellitus. *J Med Genet,* **2**: 8-11.

Dizier MH, Eliaou JF, Babron MC, Combe B, Sany J, Clot J, Clerget-Darpoux F 1993. Investigation of the HLA component involved in rheumatoid arthritis using the MASC method: rejection of the unifying shared epitope hypothesis. *Am J Hum Genet,* **5**:715-721.

Eichenbaum-Voline S, Génin E, Babron MC, Margaritte-Jeannin P, Prum B, Clerget-Darpoux F 1997. Caution in the interpretation of MLS. *Genet Epidemiol,* **14**: 1079-1084.

Falk CT, Rubinstein P 1987. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet,* **51**: 227-233.

Freimer NB, Sandkuijl LA, Blower SM 1993. Incorrect specification of marker allele frequencie : effect on linkage analysis. *Am J Hum Genet,* **56**: 1102-1110.

Génin E, Babron MC, McDermott MF, Mulcahy B, Waldron-Lynch F, Adams C, Clegg DO, Ward RH, Shanahan F, Molloy M, O'Gara F, Clerget-Darpoux F 1998. Modelling the major histocompatibility complex susceptibility to RA using the MASC method. *Genet Epidemiol,* **15**: 419-430.

Hauser ER, Boehnke M, Guo SW, Risch N 1996. Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol,* **13**: 117-137.

Holmans P 1993. Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet,* **52**: 362-374.

Jorde LB 1994. Linkage disequilibrium as a gene-mapping tool. *Am J Hum Genet,* **56**: 11-14.

Knapp M, Seuchter SA, Baur MP 1993. The haplotype-relative risk (HRR) method for analysis of association in nuclear families. *Am J Hum Genet,* **52**: 1085-1093.

Kruglyak L, Lander ES 1995a. Complete multipoint sib pair analysis of qualitative and quantitative traits. *Am J Hum Genet,* **57**: 439-454.

Kruglyak L, Lander ES 1995b. High resolution genetic mapping of complex traits. *Am J Hum Genet,* **56**: 1212-1223.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet,* **58**: 1347-1363.

Lander E, Kruglyak L 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet,* **11**: 241-247.

Margaritte-Jeannin P, Babron MC, Génin E, Eichenbaum-Voline S, Clerget-Darpoux F 1997. Heterogeneity of marker allele frequencies hinders interpretation of linkage analysis: Illustration on chromosome 18 markers. *Genet Epidemiol,* **14**: 669-674.

McGuffin P, Huckle P 1990. Simulation of Mendelism revisited: the recessive gene for attending Medical School. *Am J Hum Genet,* **46**: 994-999.

Morton NE 1955. Sequential tests for the detection of linkage. *Am J Hum Genet,* **7**: 277-318.

Nistico L, Buzzetti R, Pritchard LE et al 1996. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum Mol Genet,* **5**: 1075-1080.

Ott J 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. *Am J Hum Genet,* **51**: 283-290.

Reboul J, Mertens C, Levillayer F, Eichenbaum-Voline S, Vilkoren T, Cournu I, Babron MC, Lyon-Caen O, Clerget-Darpoux F, Fontaine B, Liblau R, The French Multiple Sclerosis Genetics Group 2000. Cytokines in genetic susceptibility to multiple sclerosis: a candidate gene approach. *J Neuroimmunol,* **102**: 107-112.

Risch N 1987. Assessing the role of HLA-linked and unlinked determinant of disease. *Am J Hum Genet,* **40**:1-14.

Risch N 1990. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet,* **46**: 242-253.

Risch N, Merinkangas K 1997. The future of genetic studies of complex human diseases. *Science,* **273**: 1516-1517.

Schaid DJ 1996. General score tests for associations of genetic markers with disease using cases and their parents. *Genet Epidemiol,* **12**: 117-137.

Sham PC, Curtis D 1995. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet,* **59**: 323-336.

Spielman R, Baur M, Clerget-Darpoux F 1989. Genetic Analysis of IDDM: Summary of GAW5. *Genet Epidemiol,* **6**: 43-58.

Spielman RS, McGinnins RE, Ewens WJ 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet,* **52**: 506-516.

Suarez BK 1978. The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens,* **12**: 87-93.

Terwilliger J, Ott J 1992. A haplotype-based "haplotype relative risk" approach to detecting allelic association. *Hum Hered,* **42**: 337-346.

Weeks DE, Lange K 1988. The Affected Pedigree Member method of linkage analysis. *Am J Hum Genet,* **42**: 315-326.

Yaouanq J, Semana G, Eichenbaum S, Quelvennec E, Roth MP, Clanet M, Edan G, Clerget-Darpoux F 1997. Evidence for linkage disequilibrium between HLA-DRB1 gene and multiple sclerosis. *Science,* **276**: 664-665.