Am. J. Hum. Genet. 66:1945–1957, 2000

Parametric and Nonparametric Multipoint Linkage Analysis with Imprinting and Two-Locus–Trait Models: Application to Mite Sensitization

Konstantin Strauch,[1] Rolf Fimmers,[1] Thorsten Kurz,[2] Klaus A. Deichmann,[2] Thomas F. Wienker,[1] and Max P. Baur[1]

[1]Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn; and [2]University Children's Hospital, University of Freiburg, Freiburg, Germany

We present two extensions to linkage analysis for genetically complex traits. The first extension allows investigators to perform parametric (LOD-score) analysis of traits caused by imprinted genes—that is, of traits showing a parent-of-origin effect. By specification of two heterozygote penetrance parameters, paternal and maternal origin of the mutation can be treated differently in terms of probability of expression of the trait. Therefore, a single–disease-locus–imprinting model includes four penetrances instead of only three. In the second extension, parametric and nonparametric linkage analysis with two trait loci is formulated for a multimarker setting, optionally taking imprinting into account. We have implemented both methods into the program GENEHUNTER. The new tools, GENEHUNTER-IMPRINTING and GENEHUNTER-TWOLOCUS, were applied to human family data for sensitization to mite allergens. The data set comprises pedigrees from England, Germany, Italy, and Portugal. With single–disease-locus–imprinting MOD-score analysis, we find several regions that show at least suggestive evidence for linkage. Most prominently, a maximum LOD score of 4.76 is obtained near D8S511, for the English population, when a model that implies complete maternal imprinting is used. Parametric two-trait-locus analysis yields a maximum LOD score of 6.09 for the German population, occurring exactly at D4S430 and D18S452. The heterogeneity model specified for analysis alludes to complete maternal imprinting at both disease loci. Altogether, our results suggest that the two novel formulations of linkage analysis provide valuable tools for genetic mapping of multifactorial traits.

Introduction

In order to understand the basic defects leading to an inherited disease, it is important to dissect the genetic factors underlying the trait. This allows us to identify individuals at risk and, finally, to develop an improved therapy for affected persons. There are two approaches to genetic linkage analysis for dichotomous traits. Parametric (LOD-score) analysis is based on an explicit genetic model. Nonparametric linkage analysis (NPL) evaluates allele sharing among affected individuals and comes to a result without particular model assumptions. If the model specified for analysis is sufficiently close to the true mode of inheritance (MOI) that governs the trait, then parametric analysis has superior power to detect linkage when compared with nonparametric analysis. Therefore, a key issue in linkage analysis is the specification of the correct genetic model. This becomes increasingly difficult with the number of factors, genetic and environmental, implicated in the etiology of the disease. Thus, identification and modeling of the genetic components of complex disorders is methodologically challenging. Despite the fact that there have been quite a few developments in linkage analysis for genetically complex traits, there is still need for new methods taking into account complex inheritance, desirably in the context of genetic-analysis software. This holds especially for linkage analysis of diseases caused by imprinted genes. Imprinting, an epigenetic factor, is also known as "parent-of-origin effect." Although many chromosomal regions are subject to imprinting, little effort has been made yet to correctly model diseases showing a parent-of-origin effect. Another important issue is linkage analysis of diseases governed by two or more loci. Such diseases have been addressed methodologically in the past. However, no tools available for linkage studies explicitly modeling two disease loci allow the use of more than two or three markers.

The present report focuses on two extensions to linkage analysis: (i) parametric (LOD-score) analysis of diseases subject to a parent-of-origin effect and (ii) parametric and NPL analysis with two-locus–trait models. Both methods have been implemented into the latest

Received January 3, 2000; accepted for publication March 27, 2000; electronically published May 4, 2000.

Address for correspondence and reprints: Konstantin Strauch, Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany. E-mail: strauch@imsdd.meb.uni-bonn.de

0002-9297/2000/6606-0023$02.00

full version of the program GENEHUNTER (Kruglyak et al. 1996; Kruglyak and Lander 1998)—that is, version 1.3. GENEHUNTER presents an excellent multimarker framework for both parametric and nonparametric single-trait-locus linkage analysis and commonly is used in genetic studies at present. It is based on the Lander-Green algorithm (Lander and Green 1987; Kruglyak et al. 1995), for which computation time increases exponentially with the number of individuals in a pedigree but only linearly with the number of markers used for analysis. Therefore, GENEHUNTER is best suited for linkage analysis of complex traits when pedigrees are usually of moderate size, simultaneously considering many markers on each chromosome. The multimarker approach allows for compensation of missing data at a particular marker locus, by spillover from adjacent markers, so that a maximum of linkage information can be obtained from a pedigree.

A disease model needs to be specified for LOD-score analysis. This may be problematic if there is no prior information about the MOI underlying the trait. In such a case, one can maximize LOD scores with respect to disease-model parameters, in which case the LOD scores are then called "MOD scores" (Clerget-Darpoux et al. 1986), or "maximizing the maximum LOD score" (MMLS [Greenberg 1989]). Using this approach, we have applied our extensions to human family data for mite sensitization. The results are presented and discussed herein.

First Extension: Parametric Linkage Analysis with Imprinting

Background

The expression of an imprinted gene depends on whether it is transmitted by the father or the mother. Imprinting can happen because of inactivation of either the paternal or maternal copy of a gene, possibly by methylation of DNA or by differences in chromatin structure (Hall 1990; Ainscough and Surani 1996; Bartolomei and Tilghman 1997). It is determined by chromosomal region. With complete maternal imprinting, for example, individuals heterozygous at the disease locus express the trait if they have inherited the mutation from the father but do not do so if they have received it from the mother. Thus, maternal imprinting means paternal expression. Examples of rare disorders that show a parent-of-origin effect are Beckwith-Wiedemann, Prader-Willi, and Angelman syndromes.

Modeling a Parent-of-Origin Effect

In standard parametric single-trait-locus linkage analysis, the trait model consists of the disease-allele frequency and three penetrance parameters: $P(+/+)$, $P(\text{Het})$, and $P(\text{m/m})$, where "+" denotes the wild-type allele and "m" denotes the mutation; "$P(\text{Het})$" denotes the penetrance for individuals who are heterozygous at the disease locus, irrespective of the parental origin of the mutation. When linkage analysis of imprinted disease genes is performed, however, such a formulation is insufficient. In this case, which parent has transmitted the allele makes a great difference in the probability that a heterozygote will express the trait; hence, no matter what heterozygote penetrance is specified for analysis, it will not be optimal for mapping the imprinted genes.

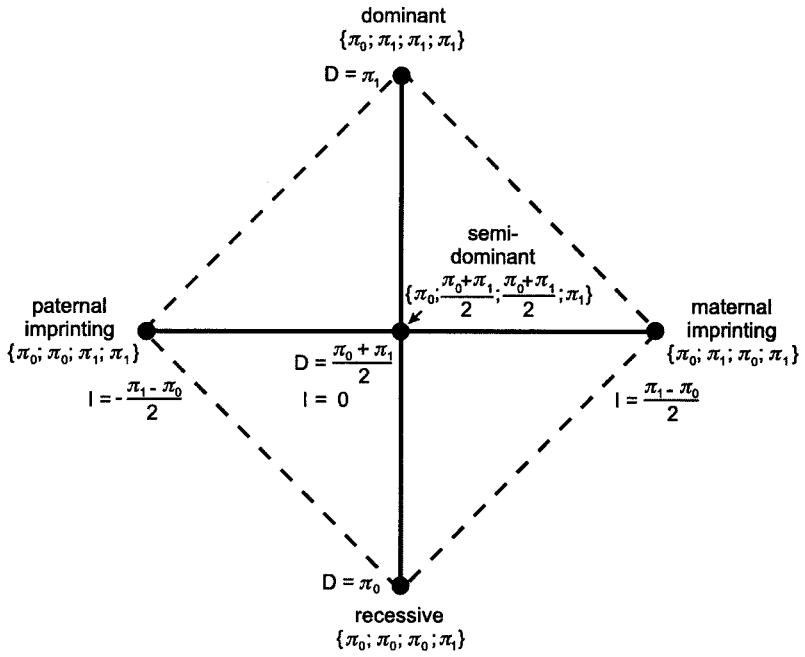
Currently, there is a lack of linkage-analysis tools that adequately model a parent-of-origin effect. On the nonparametric side, affected-sib-pair tests can be performed that evaluate allele sharing separately for male and female meioses (Paterson et al. 1999). Significant differences suggest that imprinting takes place. However, no particular disease model may be tested, and, at this time, nonparametric imprinting methods are not available for extended pedigrees. So, investigators often circumvent this difficulty by maximizing the LOD scores separately over male and female recombination fractions (Smalley 1993). Alternatively, the recombination rate of the assumed imprinting gender is fixed at $\frac{1}{2}$. By this means, nonpenetrant cases are "explained" by fictitious recombinations in the parent who has transmitted the mutation. Another expedient is the definition of separate liability classes for heterozygotes who have inherited the mutation from father versus those who have inherited it from mother (Heutink et al. 1992). Still, for most heterozygotes, the parent-of-origin cannot be recognized at first sight but can be inferred only by likelihood calculation.

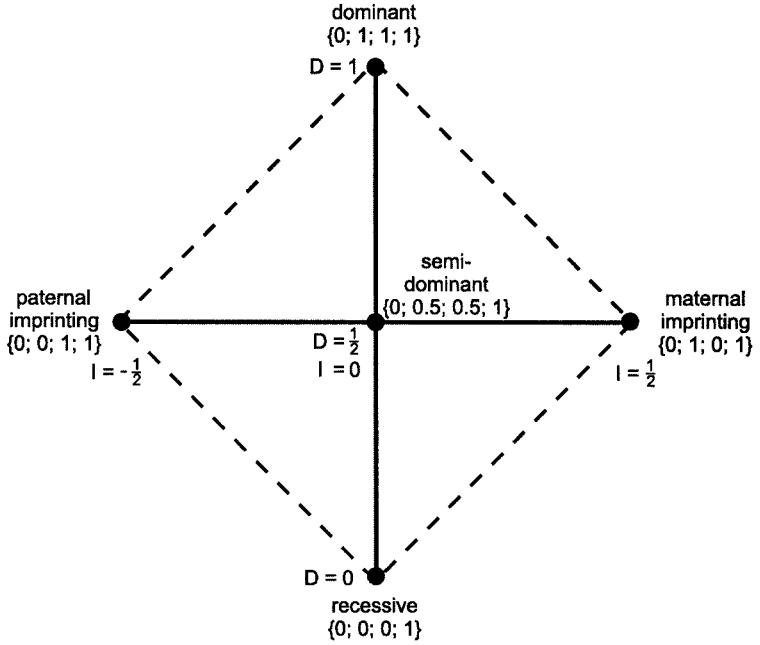
Therefore, to correctly take into account imprinting, we extend the disease model used for LOD-score analysis, in the following way: $P(\text{Het})$, the single-heterozygote penetrance, is replaced by two different penetrances, $P(\text{m/+})$ and $P(\text{+/m})$, with the paternally inherited allele listed first. This allows one to treat paternal and maternal transmission of the disease allele in a different way. Thus, there are four penetrance parameters: $P(+/+)$, $P(\text{m/+})$, $P(\text{+/m})$, and $P(\text{m/m})$. Performing parametric analysis by means of such a four-penetrance model has higher power to detect linkage of imprinted disease genes than does standard LOD-score analysis with three penetrances (Strauch et al. 1999). This is equivalent to the statement, for standard LOD-score analysis, that power to detect linkage is maximal if the analysis model corresponds to the true MOI underlying the trait (Clerget-Darpoux et al. 1986).

The Diamond of Inheritance (DOI)

The parameter space formed by the two heterozygote penetrances, given the penetrances related to both ho-

$$\{P(+/+); P(\mathrm{m}/+); P(+/\mathrm{m}); P(\mathrm{m}/\mathrm{m})\}$$





$$\text{Dominance: } D = \frac{P(\mathrm{m}/+) + P(+/\mathrm{m})}{2} \qquad \text{Imprinting: } I = \frac{P(\mathrm{m}/+) - P(+/\mathrm{m})}{2}$$

$$P(\mathrm{m}/+) = D + I \qquad P(+/\mathrm{m}) = D - I$$

Figure 1 DOI, as a visualization of the parameter space formed by the two heterozygote penetrances of the four-penetrance-trait model that takes into account a parent-of-origin effect. *Top*, General case: $P(+/+) = \pi_0$ and $P(\mathrm{m}/\mathrm{m}) = \pi_1$, with $0 \leqslant \pi_0 < \pi_1 \leqslant 1$. *Bottom*, Special case of no phenocopies and complete penetrance—that is, $P(+/+) = 0$ and $P(\mathrm{m}/\mathrm{m}) = 1$. Dominant, semidominant, and recessive modes of inheritance are displayed on the vertical axis, with the degree of imprinting, I, being 0. For these models, the two heterozygote penetrances are equal. A trait model on the left half of the diamond, with $P(\mathrm{m}/+) < P(+/\mathrm{m})$ and $I < 0$, corresponds to paternal imprinting or maternal expression. Analogously, a model on the right half of the diamond, with $P(\mathrm{m}/+) > P(+/\mathrm{m})$ and $I > 0$, corresponds to maternal imprinting that is equivalent to paternal expression.

mozygous genotypes, is illustrated by the DOI shown in figure 1. The DOI conceptually extends the single-locus–model parameter space for traits that are determined by imprinted genes. Generally, a penetrance is defined as the probability that, given a certain genotype, the trait will be expressed. Here, penetrances are listed in the order $\{P(+/+);P(\mathrm{m}/+);P(+/\mathrm{m});P(\mathrm{m}/\mathrm{m})\}$, with the paternally inherited allele named first. For a more instructive illustration of the relationship between different MOIs, we transform the two heterozygote penetrances, $P(\mathrm{m}/+)$ and $P(+/\mathrm{m})$, into the degree of dominance D and the degree of imprinting I, by the formulas defined in the legend to figure 1. Phenocopy rate $P(+/+)$ and homozygous mutant penetrance $P(\mathrm{m}/\mathrm{m})$ are fixed at π_0 and π_1, respectively, with $0 \leqslant \pi_0 < \pi_1 \leqslant 1$. This general case is presented in the upper panel of figure 1. Dominant ($D = \pi_1$) and recessive MOI ($D = \pi_0$) are the distal points on the dominance scale (vertical axis), with D being the mean of the two heterozygote penetrances. The imprinting scale is perpendicular to the dominance scale, with the degree of imprinting I ranging from $-(\pi_1 - \pi_0)/2$ (complete paternal imprinting) to $(\pi_1 - \pi_0)/2$ (complete maternal imprinting). Note that paternal imprinting means maternal expression, and vice versa. I is calculated as half the difference of the two heterozygote penetrances. In the case of no imprinting, both are equal; that is, $P(\mathrm{m}/+) = P(+/\mathrm{m}) = P(\mathrm{Het}) = D$, and I vanishes. All nonimprinting models are represented by the line connecting the dominant and recessive MOIs. Since the heterozygote penetrances are assumed to be $>P(+/+)$ and $<P(\mathrm{m}/\mathrm{m})$, all genetically possible models lie within the DOI. A point outside would correspond to either heterozygote penetrance below $P(+/+)$ (lower half of graphs in fig. 1), which is meaningless in the context of disease genes, or above $P(\mathrm{m}/\mathrm{m})$ (upper half of graphs in fig. 1), which would mean overdominance or metabolic interference. The lower panel of figure 1 shows the DOI for the special case of no phenocopies and complete homozygous mutant penetrance—that is, for $P(+/+) = 0$ and $P(\mathrm{m}/\mathrm{m}) = 1$. Here, the dominant MOI corresponds to $D = 1$, and the recessive MOI corresponds to $D = 0$. The degree of imprinting I ranges from $-\frac{1}{2}$, for complete paternal imprinting, to $\frac{1}{2}$, for complete maternal imprinting. Altogether, the dominant, recessive, complete-paternal-imprinting, and complete-maternal-imprinting MOIs form the edges of the DOI. This concept clearly demonstrates that paternal or maternal imprinting can by no means be regarded as dominant or recessive. Therefore, looking to the left and right of the central axis of dominant-recessive inheritance is of major importance when the mapping of imprinted disease genes is at issue.

Implementation

To perform LOD-score analysis that correctly models a parent-of-origin effect, we have incorporated the four-penetrance formulation into GENEHUNTER-IMPRINTING. When the imprinting option is activated, the program reads four penetrances (instead of three, as in the nonimprinting case) from the locus data file. Parameters of the disease model need to be specified by the user, for LOD-score analysis. The inheritance-vector approach used by GENEHUNTER is well suited to differentiate individuals who are heterozygous at the disease locus, on the basis of the parental origin of the disease allele; therefore, the appropriate penetrance parameter, as based on the parent who has transmitted the mutation, is either $P(\mathrm{m}/+)$ or $P(+/\mathrm{m})$. GENEHUNTER-IMPRINTING tests H_1 : linkage, under the four-penetrance–imprinting model, versus H_0 : no linkage, under the same model. LOD scores calculated under a particular four-penetrance–imprinting model that are maximized over the recombination fraction between marker and disease locus (or, equivalently, over the map position of the disease locus, in the case of multimarker analysis) have the same distribution under H_0 as do standard three-penetrance LOD scores; therefore, in terms of significance, when a predefined disease model is used, LOD scores obtained with GENEHUNTER-IMPRINTING are directly comparable to standard three-penetrance LOD scores. However, this comparability no longer holds when GENEHUNTER-IMPRINTING is used for calculation of MOD scores, as in the section titled "Application: Sensitization to Mite Allergens" (below).

Second Extension: Parametric and NPL Analysis for Two-Locus–Trait Models

Modeling Two Disease Loci

It is well known that many common diseases in man are caused by several genes. Mapping of such genetically complex traits—including mental disorders, diabetes, multiple sclerosis, asthma, and atopy—remains a daunting task. The first logical step toward adequately modeling of a multigenic trait is linkage analysis that simultaneously considers two disease loci. For the parametric (LOD score) part, such analysis is available by means of the program TMLINK, an extension of the LINKAGE software package (Lathrop et al. 1984; Lathrop and Ott 1990). It is based on the Elston-Stewart algorithm (Elston and Stewart 1971), which can cope easily with large and complex pedigrees but not with more than two or three multiallelic markers. Because of this drawback, a couple of expedients are currently used

if more than one locus is expected to influence expression of a trait. The simplest of these expedients is to specify a single-locus disease model with reduced penetrance or nonzero phenocopy rate $P(+/+)$. Smith (1963) and Ott (1983) proposed a LOD-score–based admixture model that takes into account interfamilial heterogeneity, where, for each family, the trait is assumed to be linked, with probability α, to the marker (or marker map). Several investigators have studied the power to detect linkage for traits that are governed by two genetic loci. Schork et al. (1993) have demonstrated that parametric analysis performed with two trait loci can reach higher power to detect linkage than can single-trait–locus LOD-score analysis. This contradicts findings by other investigators (Durner et al. 1992; Goldin 1992; Vieland et al. 1992*a,* 1992*b,* 1993; Goldin and Weeks 1993; Greenberg et al. 1998), who do not find an increased power to detect linkage if two trait loci are modeled. Part of the LOD-score difference obtained by Schork et al. when comparing single-locus and two-locus LOD-score analysis may be due to the fact that the disease model specified for single-locus analysis is not optimal (Sham et al. 1994). Indeed, for a trait governed by two loci, it has been shown that the power to detect linkage with single-trait locus analysis is highest if the analysis model reflects the MOI at the trait locus linked to the marker, rather than the MOI of the trait per se (Greenberg and Hodge 1989; Greenberg 1990; Durner et al. 1999; also see references cited above). Another reason why Schork et al.'s result differs from the results obtained in the other studies is the fact that they used two markers, one being linked to each disease locus (Schork et al. 1994). The other investigators performed two-trait–locus analysis with only one marker linked to one of the two trait loci. We conclude that power to detect linkage is sensitive to the amount of marker information available at both disease loci and that it is therefore advisable to perform analysis with two trait loci in a multimarker setting. In addition, if the disease is governed by two loci, then two-trait–locus linkage analysis is expected to yield more accurate estimates for the positions of the disease loci than would be expected for analysis with only one trait locus.

For nonparametric analysis, Knapp et al. (1994) have shown that affected-sib-pair tests simultaneously looking at two loci are superior to one-locus tests, in terms of power to detect linkage for traits governed by two loci. Further developments of nonparametric two-trait–locus methods include the two-locus maximum LOD score or maximum-likelihood statistic (MLS) for sib pairs (Cordell et al. 1995; Farrall 1997; Olson 1997), a score statistic developed by Dupuis et al. (1995), as well as the two-locus weighted pairwise correlation method (Zinn-Justin and Abel 1998). A recent generalization of the MLS method for affected relative pairs in extended pedigrees (Cordell et al. 2000) simultaneously analyzes identity-by-descent sharing at several loci across the genome.

Implementation

The findings described in the previous section have motivated us to implement two-trait–locus linkage analysis with multiple markers, both parametric and nonparametric, into the new program GENEHUNTER-TWOLOCUS. It uses two unlinked marker maps—for example, on nonhomologous chromosomes—with one trait locus positioned in each map. Complete inheritance information is extracted by use of all markers on both maps. For calculation of two-trait–locus LOD and NPL scores, the position of the first trait locus is held fixed at one site on the first marker map as specified by the user. The position of the second trait locus is varied on the second marker map in the same way as in a single–disease-locus GENEHUNTER or GENEHUNTER-IMPRINTING analysis. Similar to the single-locus versions, GENEHUNTER-TWOLOCUS employs the inheritance-vector approach. An inheritance vector specifies for every meiosis whether the paternally or maternally inherited allele has been transmitted and, hence, defines the founder alleles of each individual. It allows one to uniquely determine how many alleles a set of individuals shares identical by descent.

Two-trait–locus LOD and NPL scores are calculated in two steps. First, all marker information is evaluated, without consideration of the disease phenotypes. This is done separately for the two marker maps, with each map being handled in a manner similar to that for the single-locus case (Kruglyak et al. 1996). In particular, with the notation used by Kruglyak et al., the two probability distributions over the inheritance vectors $v(x_1)$ and $v(x_2)$—that is, $P(v(x_1) = w_1)$ and $P(v(x_2) = w_2)$—are calculated. They are called "inheritance distributions." Here, $v(x_1)$ and $v(x_2)$ denote the inheritance vectors at positions x_1 and x_2 of the putative disease locus on the first and second marker maps, respectively. The given probabilities are conditional on the genotypes at all markers of that particular map, and, therefore, correspond to the single-locus P_{complete}. Second, a pair of inheritance vectors at the positions of the first and second putative disease loci are assessed by a two-locus scoring function, $S(v_1,v_2,\phi)$. This function takes into account the disease phenotypes of all individuals, comprised by $\phi = (\phi_1, \ldots, \phi_n)$, where n denotes the number of individuals. In general, there is more than one possible inheritance vector at each locus, and hence the expected value of $S(v_1,v_2,\phi)$ needs to be taken over the two inheritance

distributions $v(x_1)$ and $v(x_2)$. In other words, Kruglyak et al.'s single-trait-locus formulation

$$\bar{S}(x,\phi) = \sum_{w \in V} S(w,\phi)P(v(x) = w)$$

is extended here to a two-trait–locus formulation:

$$\begin{aligned}\bar{S}(x_1,x_2,\phi) &= \sum_{w_1,w_2 \in V} S(w_1,w_2,\phi)P(v(x_1) = w_1, v(x_2) = w_2) \\ &= \sum_{w_1,w_2 \in V} S(w_1,w_2,\phi)P(v(x_1) = w_1)P(v(x_2) = w_2) \ ,\end{aligned}$$

where V represents the set of all inheritance vectors. Since the two putative disease loci are positioned on unlinked marker maps, the corresponding inheritance vectors are independent of one another. Therefore, the probability $P(v(x_1) = w_1,\ v(x_2) = w_2)$, which is conditional on the marker genotypes but not on the trait phenotypes, factors into the probabilities for each of the two loci.

Parametric (LOD-Score) Analysis.—For parametric analysis, the two disease loci are assumed to be biallelic. Here, $S(v_1,v_2,\phi)$ is the likelihood ratio $\mathrm{LR}(v_1,v_2)$:

$$\begin{aligned}S(v_1,v_2,\phi) &= \mathrm{LR}(v_1,v_2) \\ &= \frac{P(\phi|v_1,v_2)}{\sum_{w_1,w_2 \in V} P(\phi|w_1,w_2)P_{\mathrm{uniform}}(w_1)P_{\mathrm{uniform}}(w_2)} \ ,\end{aligned}$$

where $P(\phi\,|\,v_1,v_2)$ denotes the likelihood at the two disease loci, given inheritance vectors v_1 and v_2. The denominator expresses the disease-locus likelihood under the hypothesis that both disease loci are unlinked to their corresponding marker maps. This is equivalent to uniform inheritance distributions. The likelihood at the two disease loci can be expressed as follows:

$$\begin{aligned}P(\phi|v_1,v_2) &= \sum_{g_1}\sum_{g_2} P(g_1|v_1)P(g_2|v_2)P(\phi|g_1,g_2) \\ &= \sum_{g_1}\sum_{g_2} P(g_1|v_1)P(g_2|v_2)\prod_{k=1}^{n} P(\phi_k|g_{1,k},g_{2,k}) \ ,\end{aligned}$$

where $g_1 = (g_{1,1},\ldots,g_{1,n})$ and $g_2 = (g_{2,1},\ldots,g_{2,n})$ denote a particular combination of genotypes for all individuals, at the first and second disease locus, respectively. The sums are taken over all genotype combinations at each locus. $P(g_1\,|\,v_1)$ and $P(g_2\,|\,v_2)$ filter out those disease-locus genotypes that are compatible with the given inheritance vectors, and they contain the disease-allele frequencies $f_1(\mathrm{m})$ and $f_2(\mathrm{m})$, respectively. The last term, $P(\phi\,|\,g_1,g_2)$, factors into the two-locus penetrances $P(\phi_k\,|\,g_{1,k},g_{2,k})$ for each individual k, since an individual's genotype does not affect another individual's phenotype. Penetrances are abbreviated as $P(g_{1,k},g_{2,k})$. Finally, the expectation value of the parametric two-locus scoring function $S(v_1,v_2,\phi)$ over the inheritance distributions $v(x_1)$ and $v(x_2)$ must be calculated. This yields the likelihood ratio $\mathrm{LR}(x_1,x_2)$ at positions x_1 and x_2:

$$\begin{aligned}\bar{S}(x_1,x_2,\phi) &= \sum_{w_1,w_2 \in V} \mathrm{LR}(w_1,w_2)P(v(x_1) = w_1)P(v(x_2) = w_2) \\ &= \frac{\sum_{w_1,w_2 \in V} P(\phi|w_1,w_2)P(v(x_1) = w_1)P(v(x_2) = w_2)}{\sum_{w_1,w_2 \in V} P(\phi|w_1,w_2)P_{\mathrm{uniform}}(w_1)P_{\mathrm{uniform}}(w_2)} \\ &= \mathrm{LR}(x_1,x_2) \ .\end{aligned}$$

The numerator of this expression is proportional to the complete likelihood for all markers and both disease loci at positions x_1 and x_2 of the first and second disease locus, respectively. The denominator, as before, represents the likelihood under the hypothesis that both disease loci are unlinked to their corresponding marker maps. Altogether, GENEHUNTER-TWOLOCUS reports the decimal logarithm of this expression—that is, the exact multipoint two-trait–locus LOD score Z:

$$\begin{aligned}Z &= \log_{10}\mathrm{LR}(x_1,x_2) \\ &= \log_{10}\frac{\mathrm{L}(x_1,x_2)}{\mathrm{L}(\text{locus 1 unlinked, locus 2 unlinked})} \ ,\end{aligned}$$

where $\mathrm{L}(x_1,x_2)$ denotes the likelihood for positions x_1 and x_2 of the first and second disease locus, respectively. LOD-score calculation is performed given a specific two-locus imprinting model—that is, allele frequencies at both disease loci, along with a 4×4 matrix of penetrances for combined two-locus genotypes. This, again, allows one to distinguish heterozygotes on the basis of the parental origin of the disease allele and, hence, to realistically model imprinting in the context of two-disease-locus analysis. The trait model needs to be specified by the user. In the nonimprinting case, penetrances reduce to a 3×3 matrix.

NPL.—With respect to NPL analysis, GENEHUNTER-TWOLOCUS uses extensions of the scoring functions S_{pairs} and S_{all} that evaluate sharing of alleles identical-by-descent for affected individuals, simultaneously at both disease loci. NPL scores are given with P values, with the assumption of the null hypothesis that both disease loci are unlinked to their corresponding marker maps. Knapp et al. (1994) have shown, for sib pairs, that a two-locus extension of the mean test, using the sum of the two single–disease-locus mean test scores,

is almost as powerful as the optimal test, for a given disease model. Accordingly, we have chosen to compute the two-locus S_{pairs} and S_{all} as the sum of the corresponding single-locus scores in the current implementation:

$$S(v_1,v_2,\phi) = S(v_1,\phi) + S(v_2,\phi) \ .$$

However, different two-locus NPL scoring functions can easily be incorporated into GENEHUNTER-TWOLOCUS.

Application: Sensitization to Mite Allergens

Parent-of-origin effects are suspected to play a major role in the development of atopy in man (Moffatt and Cookson 1998). We have therefore used GENEHUNTER-IMPRINTING to perform exploratory linkage analysis with pedigree data of a candidate-region search for sensitization to mite allergens. The data set consists of sib pairs as well as of extended pedigrees originating from England, Germany, Italy, and Portugal. There are 94 families comprising 438 individuals altogether, with pedigree size distributed as follows: English population—19 families with 122 individuals, including 7 families with 2 affected sibs, 3 families with 3 affected sibs, 5 families with 4 affected sibs, and 4 extended pedigrees; German population—44 families with 188 individuals, including 33 families with 2 affected sibs and 11 families with 3 affected sibs; Italian population—5 families with 22 individuals, including 3 families with 2 affected sibs and 2 families with 3 affected sibs; and Portuguese population—26 families with 106 individuals, including 23 families with 2 affected sibs and 3 families with 3 affected sibs. Specific IgE antibodies to highly purified crude extract of *Dermatophagoides pteronyssimus* were detected, for each person, by an immunochemiluminometric immunosorbent assay (Magic Lite; ALK). An individual was classified to be affected if a positive titer (according to the manufacturer's recommended specifications) was obtained; otherwise, the individual was considered to be unaffected. One hundred fifty microsatellite markers were typed on chromosomes 1–21, with an average spacing of 10 cM at each candidate region.

Since prior information about the true MOI of the trait is lacking, we did not just perform LOD-score analysis with one disease model; rather, we employed a couple of models for analysis and subsequently calculated MOD scores for genetic regions initially showing a promising LOD score. In particular, we first performed multimarker NPL analysis with option S_{pairs} of GENEHUNTER, as well as LOD-score analysis with GENEHUNTER-IMPRINTING, using five disease models: dominant {.02;.9;.9;.9}, semidominant {.02;.5;.5;.9}, recessive {.02;.1;.1;.9}, paternal imprinting {.02;.1;.9;.9}, maternal imprinting {.02;.9;.1;.9}, with penetrances given in the order $\{P(+/+);P(\mathrm{m}/+);P(+/\mathrm{m});P(\mathrm{m}/\mathrm{m})\}$. These five models represent the four corners and the center of the DOI, although the lower heterozygote penetrance of .1 was chosen to be slightly higher than the phenocopy rate $P(+/+)$. The homozygous mutant penetrance $P(\mathrm{m}/\mathrm{m})$ was arbitrarily set to .9. In the beginning, we assumed the phenocopy rate to be 2% and used the disease-allele frequency $f(\mathrm{m}) = .005$ for all models. Since there is no evidence that prevalence for atopy differs between the two sexes, we used the same penetrances for males and females. Separate multipoint analyses were performed on the English, German, and Portuguese families, as well as on the entire data set, simultaneously using all markers typed for each chromosome. At each marker locus, allele frequencies were assumed to be uniformly distributed. Because of the small number of Italian families, no separate analysis was performed for this population.

If either maximum LOD, HLOD, or NPL score exceeded 2.0, we subsequently maximized LOD scores over disease-model parameters, with GENEHUNTER-IMPRINTING, for that particular population and genetic region. This approach, named "MOD" score analysis, was first proposed by Risch (1984). It not only yields information about evidence for linkage but also provides a valid ascertainment-assumption–free method for estimation of disease-model parameters (Elston 1989; Greenberg 1989; Clerget-Darpoux and Bonaïti-Pellié 1992; Hodge and Elston 1994). Strictly considered, this holds only if the genetic mechanism of the disease specified for analysis—that is, the number of loci—reflects the true genetic mechanism. Yet, Greenberg (1990) has shown by simulation that estimates for disease-model parameters are still approximately correct if a trait determined by two loci is analyzed under a single-locus–trait model. It should be noted that maximization over both the position of the disease locus and the disease-model parameters leads to inflated LOD scores. This needs to be accounted for when one is assessing the significance of the results (see the Discussion section). MOD-score calculation can increase the power to detect linkage, compared with LOD-score analysis with a single model that may be wrong (Clerget-Darpoux et al. 1986; Greenberg et al. 1998; Abreu et al. 1999). This is especially true for multimarker analysis, in which a misspecified model may lead to exclusion of linkage for the disease locus (Risch and Giuffra 1992). Hence, we suppose that it is prudent to perform MOD-score analysis if no prior information about the MOI is available, as is the case in the present study. This holds especially if genes are suspected to be imprinted. Therefore, we have maximized multimarker LOD scores with respect to disease-allele frequency and all four penetrance parameters. We have employed an ad hoc step-

Table 1

Results of Imprinting Models

Marker (Population) and Position	MOD Score	Penetrance					
		P(+/+)	*P*(m/+)	*P*(Het)	*P*(+/m)	*P*(m/m)	*f*(m)
D4S430 (German):							
121.4 cM (imprinting model)	3.46	.03	1.00		.35	1.00	.005
121.4 cM (nonimprinting model)	3.33	.00		.20		1.00	.03
D5S416 (English):							
27.9 cM (imprinting model)	3.48	.00	.30		.10	.30	.07
22.1 cM (nonimprinting model)	3.12	.00		.03		.35	.15
D8S511 (English):							
28.1 cM (imprinting model)	4.76	.04	1.00		.04	1.00	.000005
36.5 cM (nonimprinting model)	2.30	.00		.20		.95	.0005
D11S1314 (total sample):							
75.6 cM (imprinting model)	2.61	.03	.03		.15	.55	.009
77.5 cM (nonimprinting model)	2.18	.03		.10		.45	.01
D11S898 (total sample):							
105.9 cM (imprinting model)	2.29	.05	.05		.25	.25	.02
103.1 cM (nonimprinting model)	1.63	.05		.20		.45	.008
D16S3039 (total sample):							
73.3 cM (imprinting model)	3.37	.02	.05		.15	.35	.005
72.6 cM (nonimprinting model)	2.47	.02		.10		.30	.005
D16S3096 (total sample):							
97.3 cM (imprinting model)	3.06	.02	.05		.25	.25	.004
97.3 cM (nonimprinting model)	1.44	.02		.10		.10	.008
D18S452 (German):							
9.7 cM (imprinting model)	2.36	.02	1.00		.02	1.00	.003
16.7 cM (nonimprinting model)	1.15	.01		.04		.30	.03
D21S265 (Portuguese):							
24.1 cM (imprinting model)	3.30	.02	.20		1.00	1.00	.001
24.1 cM (nonimprinting model)	2.19	.02		.50		1.00	.001

Note.—Genomic regions for which GENEHUNTER-IMPRINTING MOD-score calculation leads to an imprinting disease model, i.e., to penetrances with $P(\mathrm{m}/+) \neq P(+/\mathrm{m})$. At the given position (according to the Généthon map; Dib et al. 1996), the MOD score reaches its maximum.

wise-gradient strategy, varying penetrances by steps of .05. Disease-allele frequency and penetrances that proved to be <.05 were varied by steps ⩽.01, to reach the LOD-score maximum.

Imprinting Results

Table 1 shows the genetic regions for which MOD-score calculation leads to a disease model with $P(\mathrm{m}/+) \neq P(+/\mathrm{m})$. Such models, as can be seen from the DOI, lie off the central axis of dominant-recessive inheritance, which may indicate a parent-of-origin effect. (A detailed description of the mite-sensitization study, also including genetic regions with $P(\mathrm{m}/+) = P(+/\mathrm{m})$ that supposedly are involved in mite sensitization, will be published elsewhere.) It may be asked which criterion should be employed to finally judge whether imprinting, in fact, takes place. A difference between the two heterozygote penetrances obtained by MOD-score analysis may indicate true imprinting, but it may also occur just by chance. One can argue that evidence for a parent-of-origin effect is highest for a large difference between $P(\mathrm{m}/+)$ and $P(+/\mathrm{m})$, corresponding to a point on the extreme left or right of the DOI. Here, the two heterozygote penetrances of the best-fitting imprinting model differ by ⩾.2 at all loci listed in table 1, except for D11S1314 and D16S3039. However, the only way to exhaustively address the question of how large the difference should be is to perform simulations for each particular case, which we have not done. Instead, we will use a different approach to assess whether imprinting takes place. Greenberg and Berger (1994) have investigated to what extent LOD-score differences can be used to infer MOI. In their simulation study, they find that, with a difference of 1.5 between two MOIs (in their case, dominant and recessive), the superior LOD score reflects, with high reliability, the correct MOI and that a difference of 2.5 practically guarantees correct inference of MOI. We suppose that these criteria are not restricted to the dominant and recessive MOIs. Hence, in addition to the MOD scores under imprinting, we also have calculated the MOD scores that were maximized over standard nonimprinting models with three penetrances. This is equivalent to the constraint $P(\mathrm{m}/+) = P(+/\mathrm{m}) = P(\mathrm{Het})$. The nonimprinting results also are shown in table 1. Following the approach used by Greenberg and Berger, we

looked at the difference between imprinting and nonimprinting MOD scores, to gain evidence about a parent-of-origin effect.

$P(m/+) < P(+/m)$, possibly alluding to paternal imprinting, together with a MOD score >3 is found next to D16S3039 and at D16S3096 in the complete data set, as well as at D21S265 in the Portuguese families. Differences between imprinting and nonimprinting MOD scores are moderate next to D16S3039 (.9) and at D21S265 (1.11). A more-pronounced MOD-score difference of 1.62 is obtained at D16S3096, exceeding the 1.5 critical value proposed by Greenberg and Berger. This gives an indication that a paternally imprinted disease gene is located close to D16S3096. A paternally imprinted gene on 11q13 has previously been suspected to determine sensitization to mite allergens (Moffatt and Cookson 1998). In our analysis, we obtain a MOD score of 2.61 in that region near D11S1314, for the complete data set, and obtain a best-fitting disease model with $P(m/+) < P(+/m)$. This result may point in the same direction as does Moffatt and Cookson's finding, although the difference between imprinting and nonimprinting MOD score is as low as .43. Next to D11S898, the model reached by LOD-score maximization for the total data set has penetrances $P(+/+) = P(m/+) = .05$ and $P(+/m) = P(m/m) = .25$, which might allude to expression of the maternally inherited allele exclusively. However, the imprinting MOD score is only 2.29, and the difference between the latter and the nonimprinting MOD score of 0.66 can, at most, be regarded as moderate.

A MOD score >3 and $P(m/+) > P(+/m)$, which may point to maternal imprinting, is found near D4S430 in the German population and at D5S416 in the English population. However, there is only marginal MOD-score increase at both loci when $P(m/+)$ is allowed to differ from $P(+/m)$. The best-fitting penetrances near D18S452 in the German population are $P(+/+) = P(+/m) = .02$ and $P(m/+) = P(m/m) = 1$. Judged on the basis of the disease model, the maternally inherited allele does not seem to influence expression of the trait at all. There also is a noticeable difference, 1.21, between the imprinting and nonimprinting MOD scores. Nevertheless, the imprinting MOD score of 2.36 can be interpreted as only suggestive evidence for linkage. The most striking result, a MOD score of 4.76, is obtained near marker D8S511 in the English population. Penetrance is modeled as complete if the mutation is inherited from the father and is modeled as a phenocopy rate of 4% otherwise. Similar to the former case, the maternally inherited allele does not seem to make any difference in terms of the probability for development of the allergy. In addition, the imprinting MOD score surmounts the nonimprinting result by almost 2.5 units. According to Greenberg and Berger, such a difference practically guarantees correct inference of the MOI. Together with the high MOD score of 4.76, this clearly indicates the existence of a maternally imprinted gene, near D8S511, that is involved in mite sensitization.

One may expect that the nonimprinting disease model yielding the highest maximum LOD score should differ from the best-fitting imprinting disease model only by the heterozygote penetrance, possibly with $P(\text{Het})$ being the average of $P(m/+)$ and $P(+/m)$. This does not hold in general, as can be seen from the results. It is approximately true for D21S265, D16S3039, and D11S1314. At the other loci, except for D11S898, MOD-score calculation for nonimprinting models leads to a higher disease-allele frequency and lower penetrances than are seen for imprinting. This yields prevalences of the same order of magnitude for best-fitting imprinting and nonimprinting models, at most loci. The fact that a higher disease-allele frequency is obtained with nonimprinting MOD-score analysis might be explained as follows. If we assume that the imprinting model reflects the true MOI of the particular gene, then nonimprinting analysis is equivalent to model misspecification. Risch and Giuffra (1992) have shown for multimarker analysis that a misspecified trait model leads to strongly reduced LOD scores. This can be compensated, in part, by specifying that the disease-allele frequency for analysis be higher than the true value. Therefore, maximization of LOD scores over incorrect nonimprinting models could, in turn, yield a higher estimate for the disease-allele frequency.

Two-Trait–Locus Analysis of Mite Sensitization

Motivation

In our single–disease-locus analysis of the mite-sensitization phenotype, we find two regions for the German sample that have a MOD score >2: D4S430 (3.46) and D18S452 (2.36). We therefore judged this sample to be perfectly suited for a two-trait–locus study and decided to perform multipoint NPL and LOD-score analysis of data on chromosomes 4 and 18, with GENEHUNTER-TWOLOCUS. For both D4S430 and D18S452, single-locus MOD-score calculation revealed a best-fitting disease model with a fully penetrant gene subject to maternal imprinting and with a nonzero phenocopy rate (see above). Complete penetrance implies that each of the two disease genes independently results in sensitization to mite allergens. The existence of phenocopies suggests that the trait may also have causes other than the particular gene in consideration. Obviously, this points to the other of the two genes but may also suggest further genetic or environmental factors. Hence, to begin, we assumed the trait to be governed by a two-locus heterogeneity model, as defined by Risch (1990) or Neu-

Table 2

Two-Locus Heterogeneity/Maternal-Imprinting Model

First-Locus Genotype	Penetrance When Second-Locus Genotype Is			
	+/+	m/+	+/m	m/m
+/+	.00	1.00	.00	1.00
m/+	1.00	1.00	1.00	1.00
+/m	.00	1.00	.00	1.00
m/m	1.00	1.00	1.00	1.00

Note.—The two-locus heterogeneity/maternal imprinting model. A paternally inherited allele at either locus leads to expression of the trait.

man and Rice (1992). This model assumes individuals to be affected if a disease allele is present at either of the two loci. However, in our analysis with GENEHUNTER-TWOLOCUS, we extended the heterogeneity model for maternal imprinting; it takes into account that only a paternally inherited mutation at either locus leads to expression of the trait.

Two-Locus–Trait Models With and Without Imprinting

In the usual setting, a two-locus–trait model includes the disease-allele frequency at each locus and a 3×3 penetrance matrix. Here, this model is extended to incorporate imprinting effects at both loci. If heterozygotes are differentiated on the basis of parental origin, there are four possible genotypes at each locus, which corresponds to an expanded, 4×4 penetrance matrix. For analysis of chromosome 4 (for the first disease locus) and chromosome 18 data (for the second disease locus), we used a heterogeneity model with complete maternal imprinting at both loci, with the first row and column of the penetrance matrix being {0;1;0;1}. This vector contains the probabilities of expression of the trait when the effect of only one locus is regarded—that is, the condition for individuals who are homozygous wild-type at the other disease locus. In the past (Risch 1990; Neuman and Rice 1992), these probabilities have been named "marginal penetrances." The other matrix elements are obtained by simultaneously considering the two single-locus effects, by means of formula $P_{i,j} = P_{i,1} + P_{1,j} - P_{i,1}P_{1,j}$. Here, i and j are in the range of 1–4 and denote the genotype at the first and second trait locus, respectively. The heterogeneity/maternal-imprinting model is shown in table 2. To investigate whether this particular two-locus model is adequate, we chose the approach of a two-locus MOD-score analysis and maximized two-locus LOD scores over allele frequencies at both disease loci and over all 16 penetrances. Initially, the disease-allele frequencies were assumed to be equal to those obtained by single–disease-locus MOD-score analysis. Penetrances were varied by steps of .01, and disease-allele frequencies were varied by steps of .001.

We recalculated the two-locus MOD score for nonimprinting-trait models. These are constrained by $P(g_1,\text{m/+}) = P(g_1,\text{+/m}) = P(g_1,\text{Het})$, with g_1 being the genotype at the first trait locus and with $P(\text{m/+},g_2) = P(\text{+/m},g_2) = P(\text{Het},g_2)$, with g_2 denoting the genotype at the second trait locus. In other words, the two penetrances in the upper middle, in the lower middle, on the middle left, and on the middle right of the matrix are each set to be equal, as are the four penetrances in the center. This leads to the well-known two-locus model with nine penetrances.

Two-Trait–Locus Results

With NPL analysis using the two-locus version of S_{pairs}, we obtain an NPL score of 3.66, which corresponds to an exact P value of .0001, occurring precisely at D4S430 and D18S452. The single-locus NPL scores are 3.28 ($P = .0005$) for D4S430 and 1.96 ($P = .0274$) for D18S452. LOD-score maximization with respect to disease-model parameters reveals the assumed heterogeneity/maternal-imprinting model given in table 2 to be indeed the best-fitting model. Only the disease-allele frequencies prove to be slightly lower than assumed—that is, $f_1(m) = .003$ for D4S430 and $f_2(m) = .002$ for D18S452. The two-locus LOD score reaches 6.09, which is 0.27 higher than the 5.82 sum of the corresponding maximized single–disease-locus LOD scores. Most interestingly, whereas the maximum single-locus LOD-score positions are 3.7 cM centromeric of D4S430 and 8 cM telomeric of D18S452, the two-locus LOD score is maximal exactly at these two markers. This is equivalent to no recombinations occurring between either D4S430 or D18S452 and the corresponding disease loci. Two-locus nonimprinting analysis yields a MOD score of 4.21, with the disease model given in table 3. The associated disease-allele frequencies are $f_1(m) = .004$ and $f_2(m) = .006$. Unlike the best imprinting model, the nonimprinting model contains reduced penetrances as well as nonzero phenocopy rates. As for the maximum-likelihood positions of the two disease loci, the locus on chromosome 18 is positioned precisely at D18S452, the

Table 3

Nonimprinting Model Obtained by Two-Locus MOD-Score Analysis

First-Locus Genotype	Penetrance When Second-Locus Genotype Is		
	+/+	Heterozygous	m/m
+/+	.00	.04	.88
Heterozygous	.10	.10	1.00
m/m	1.00	1.00	1.00

same position as in the imprinting model, whereas the chromosome 4 locus now appears to be 3 cM centromeric of D4S430.

Discussion

Assessment of the significance of LOD scores is a difficult issue. There is already debate with regard to single–disease-locus LOD scores when one model is used, and things are even more complicated with LOD scores that have been maximized over disease-model parameters (Clerget-Darpoux et al. 1990). MacLean et al. (1993*a*) have shown, for phase-known double backcrosses, that maximization of the LOD score over one penetrance parameter leads to a χ^2 distribution with one extra degree of freedom (df). Still, in general, it is not possible to account for LOD-score inflation by assuming additional df, since one maximizes the LOD score rather than the likelihood (Morton 1998). Weeks et al. (1990) and Hodge et al. (1997) have found, by simulation, that, for MOD scores, a critical LOD score of 3 should be adjusted by some value in the range of 0.3–1.0, with the upper bound being rather conservative; however, the correction of 1.0, given by Weeks et al., is due not only to MOD-score calculation but also to maximization over several diagnostic schemes. Therefore, even though an additional penetrance parameter is introduced in GENEHUNTER-IMPRINTING, we consider the MOD score of 4.76 obtained near D8S511 in the English families to be clearly remarkable. Moreover, we propose that the other loci given in table 1, especially those with a MOD score >3, show at least suggestive evidence for linkage to mite sensitization. Altogether, the results indicate that the atopic phenotype of mite sensitization in humans is genetically heterogeneous, both within and between populations. This corresponds to findings in previous studies of asthma and atopy (Barnes and Marsh 1998; Howard et al. 1999). For the first time, however, we find a clear indication that paternally expressed genes are involved in atopy.

Even more difficult, if not impossible, is assessment of the significance of two-locus LOD scores. If maximized over two recombination fractions, $2\ln(10)$ times the LOD score is distributed as a $\frac{1}{4} : \frac{1}{2} : \frac{1}{4}$ mixture of χ^2 with 2 df, χ^2 with 1 df, and a point mass at zero under the null hypothesis that both disease loci are unlinked to the markers (Self and Liang 1987). With these assumptions, a single-locus LOD score of 3 corresponds to a two-locus LOD score of 3.5. Schork et al. (1993) have commented on this, arguing that significance depends on assumptions about prior probability of linkage to two loci. In the present study of mite sensitization, inflation of two-locus LOD scores should be expected, since we maximized LOD scores with respect to both disease-allele frequencies and all penetrance parameters. Similar to what is seen for single-locus MOD scores, it is impossible to express this inflation in terms of additional df of the underlying χ^2 distribution. Still, penetrances of the initially assumed imprinting model prove to be the best guess. We also would like to add to Schork et al.'s discussion from a different point of view. The sharp model, pointing to heterogeneity and maternal imprinting, which is obtained by two-locus MOD-score analysis, has intuitive appeal in terms of its biological aspects. It represents two genes that are subject to complete maternal imprinting and that can independently lead to development of the atopy. Clearly, the model is more adequate than the two-locus nonimprinting model that contains both reduced penetrances and nonzero phenocopy rates. The improvement of the model proceeds with a MOD-score difference of almost 2. Altogether, the MOD score of 6.09 is an exceptionally clear result. It raises the importance of the single–disease-locus results for D4S430 and D18S452 from more or less suggestive to absolutely remarkable.

MacLean et al. (1992) have proposed a multi–disease-locus test that combines single–disease-locus LOD scores. Cox et al. (1999) have developed the two-locus equivalent for NPL scores, using the result at the first disease locus to weight the scores, of different pedigrees, at the second locus. Furthermore, they calculate family correlations between NPL scores at two disease loci, as do MacLean et al. (1993*b*) for LOD scores. This allows one to imply the degree of interaction between disease loci and, hence, to distinguish between heterogeneity and multiplicative action. The parametric part of GENEHUNTER-TWOLOCUS requires specification of a considerable number of parameters but allows the user to perform genetic modeling in a much more specific way, even including parent-of-origin effects. In our example, this proves to be crucial for detection of linkage. In cases in which no prior information or hypothesis about the two-locus disease model is available, two-locus NPL analysis may be the method of choice. Whereas GENEHUNTER-IMPRINTING does not need more computation time than is required by the nonimprinting version, analysis with GENEHUNTER-TWOLOCUS is computationally more intensive than single-trait-locus analysis. In GENEHUNTER-TWOLOCUS analysis, the number of effective meioses in a pedigree ([2 × nonfounders] − founders [Kruglyak et al. 1996]) should not exceed 11 or 12. In any case, for the small pedigrees that Cox et al. have proposed as being optimal for their method, linkage analysis using GENEHUNTER-TWOLOCUS is without problems.

It is well known that a considerable portion of the human genome is subject to imprinting, although not all those genetic regions have been discovered yet (Hall 1990; Ainscough and Surani 1996; Bartolomei and Tilghman 1997). Clerget-Darpoux et al. (1986) have

shown that it is crucial to correctly specify the MOI—that is, whether inheritance is dominant or recessive—since power to detect linkage is strongly reduced otherwise. This is reflected by the fact that the dominant and recessive MOIs are opposite points in the DOI. By the same token, the DOI illustrates that analysis of a maternally or paternally imprinted gene by means of a dominant or recessive model is an aggravating misspecification as well. This explains why ignoring the imprinting in the analysis model—if, in fact, imprinting is present—can drastically reduce the power to detect linkage, as has been shown by Strauch et al. (1999) and as has been reflected in the results of the present study. We therefore propose that, for parametric linkage analysis, the maternal- as well as the paternal-imprinting model, being the two other edges of the DOI, should become as standard as the dominant and recessive models. In addition, we suggest that genetic linkage studies that did not show linkage be revisited with LOD-score analysis correctly modeling a parent-of-origin effect.

With our two novel formulations of linkage analysis, we aim at combining the advantages of modeling genetically complex diseases in an appropriate way, with extraction of as much inheritance information as possible by multimarker analysis. We think that these tools can help us to dissect the genetic components of a wide variety of complex traits. (Both programs—GENEHUNTER-IMPRINTING and GENEHUNTER-TWOLOCUS—can be obtained by contacting the corresponding author.)

Acknowledgments

We would like to thank Birgit Buchholz and Katharina Faleovalu for graphical assistance, as well as two anonymous reviewers for their helpful comments. This work was supported by Deutsche Forschungsgemeinschaft grants GRK246/TP07, SFB400/Z2, and Ba660/6-3 (to K.S., R.F., T.F.W., and M.P.B.), as well as by Bundesministerium für Forschung und Technologie grant "Klinische Forschergruppe: Pathomechanismen der allergischen Entzündung" 01GC9701/5 (to T.K. and K.A.D.).

References

Abreu PC, Greenberg DA, Hodge SE (1999) Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. Am J Hum Genet 65:847–857

Ainscough JFX, Surani MA (1996) Organization and control of imprinted genes: the common features. In: Russo VEA, Martienssen RA, Riggs AD (eds) Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 173–194

Barnes KC, Marsh DG (1998) The genetics and complexity of allergy and asthma. Immunol Today 19:325–332

Bartolomei MS, Tilghman SM (1997) Genomic imprinting in mammals. Annu Rev Genet 31:493–525

Clerget-Darpoux F, Babron M-C, Bonaïti-Pellié C (1990) Assessing the effect of multiple linkage tests in complex diseases. Genet Epidemiol 7:245–253

Clerget-Darpoux F, Bonaïti-Pellié C (1992) Strategies based on marker information for the study of human diseases. Ann Hum Genet 56:145–153

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

Cordell HJ, Todd JA, Bennett ST, Kawaguchi Y, Farrall M (1995) Two-locus maximum LOD score analysis of a multifactorial trait: joint consideration of *IDDM2* and *IDDM4* with *IDDM1* in type I diabetes. Am J Hum Genet 57: 920–934

Cordell HJ, Wedig GC, Jacobs KB, Elston RC (2000) Multilocus linkage tests based on affected relative pairs. Am J Hum Genet 66:1273–1286

Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A (1999) Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nat Genet 21:213–215

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Dupuis J, Brown PO, Siegmund D (1995) Statistical methods for linkage analysis of complex traits from high-resolution maps of identity by descent. Genetics 140:843–856

Durner M, Greenberg DA, Hodge SE (1992) Inter- and intra-familial heterogeneity: effective sampling strategies and comparison of analysis methods. Am J Hum Genet 51:859–870

Durner M, Vieland VJ, Greenberg DA (1999) Further evidence for the increased power of LOD scores compared with nonparametric methods. Am J Hum Genet 64:281–289

Elston RC (1989) Man bites dog? the validity of maximizing lod scores to determine mode of inheritance. Am J Med Genet 34:487–488

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Farrall M (1997) Affected sibpair linkage tests for multiple linked susceptibility genes. Genet Epidemiol 14:103–115

Goldin LR (1992) Detection of linkage under heterogeneity: comparison of the two-locus vs. admixture models. Genet Epidemiol 9:61–66

Goldin LR, Weeks DE (1993) Two-locus models of disease: comparison of likelihood and nonparametric linkage methods. Am J Hum Genet 53:908–915

Greenberg DA (1989) Inferring mode of inheritance by comparison of lod scores. Am J Med Genet 34:480–486

——— (1990) Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: inferring mode of inheritance and estimating penetrance. Genet Epidemiol 7:467–479

Greenberg DA, Abreu P, Hodge SE (1998) The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet 63:870–879

Greenberg DA, Berger B (1994) Using LOD-score differences to determine mode of inheritance: a simple, robust method even in the presence of heterogeneity and reduced penetrance. Am J Hum Genet 55:834–840

Greenberg DA, Hodge SE (1989) Linkage analysis under "random" and "genetic" reduced penetrance. Genet Epidemiol 6:259–264

Hall JG (1990) Genomic imprinting: review and relevance to human diseases. Am J Hum Genet 46:857–873

Heutink P, van der Mey AGL, Sandkuijl LA, van Gils APG, Bardoel A, Breedveld GJ, van Vliet M, et al (1992) A gene subject to genomic imprinting and responsible for hereditary paragangliomas maps to chromosome 11q23-qter. Hum Mol Genet 1:7–10

Hodge SE, Abreu PC, Greenberg DA (1997) Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. Am J Hum Genet 60: 217–227

Hodge SE, Elston RC (1994) Lods, wrods, and mods: the interpretation of lod scores calculated under different models. Genet Epidemiol 11:329–342

Howard TD, Wiesch DG, Koppelman GH, Postma DS, Meyers DA, Bleecker ER (1999) Genetics of allergy and bronchial hyperresponsiveness. Clin Exp Allergy 29 Suppl 2:86–89

Knapp M, Seuchter SA, Baur MP (1994) Two-locus disease models with two marker loci: the power of affected-sib-pair tests. Am J Hum Genet 55:1030–1041

Kruglyak L, Daly MJ, Lander ES (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am J Hum Genet 56: 519–527

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

Kruglyak L, Lander ES (1998) Faster multipoint linkage analysis using Fourier transforms. J Comput Biol 5:1–7

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84: 2363–2367

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lathrop GM, Ott J (1990) Analysis of complex diseases under oligogenic models and intrafamilial heterogeneity by the LINKAGE programs. Am J Hum Genet Suppl 47:A188

MacLean CJ, Bishop DT, Sherman SL, Diehl SR (1993*a*) Distribution of LOD scores under uncertain mode of inheritance. Am J Hum Genet 52:354–361

MacLean CJ, Ploughman LM, Diehl SR, Kendler KS (1992) A new test for linkage in the presence of locus heterogeneity. Am J Hum Genet 50:1259–1266

MacLean CJ, Sham PC, Kendler KC (1993*b*) Joint linkage of multiple loci for a complex disorder. Am J Hum Genet 53: 353–366

Moffatt MF, Cookson WOCM (1998) Maternal effects in atopic disease. Clin Exp Allergy 28 Suppl 1:56–61

Morton NE (1998) Significance levels in complex inheritance. Am J Hum Genet 62:690–697

Neuman RJ, Rice JP (1992) Two-locus models of disease. Genet Epidemiol 9:347–365

Olson JM (1997) Likelihood-based models for genetic linkage analysis using affected sib pairs. Hum Hered 47:110–120

Ott J (1983) Linkage analysis and family classification under heterogeneity. Ann Hum Genet 47:311–320

Paterson AD, Naimark DMJ, Petronis A (1999) The analysis of parental origin of alleles may detect susceptibility loci for complex disorders. Hum Hered 49:197–204

Risch N (1984) Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes. Am J Hum Genet 36:363–386

——— (1990) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Risch N, Giuffra L (1992) Model misspecification and multipoint linkage analysis. Hum Hered 42:77–92

Schork NJ, Boehnke M, Terwilliger JD, Ott J (1993) Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. Am J Hum Genet 53:1127–1136

——— (1994) Reply to Sham et al. Am J Hum Genet 55: 856–858

Self SG, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Sham PC, MacLean CJ, Kendler KS (1994) Two-locus versus one-locus LODs for complex traits. Am J Hum Genet 55: 855–856

Smalley SL (1993) Sex-specific recombination frequencies: a consequence of imprinting? Am J Hum Genet 52:210–212

Smith CAB (1963) Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet 27: 175–182

Strauch K, Fimmers R, Windemuth C, Hahn A, Wienker TF, Baur MP (1999) Linkage analysis with adequate modeling of a parent-of-origin effect. Genet Epidemiol 17 Suppl 1: S331–S336

Vieland VJ, Greenberg DA, Hodge SE (1993) Adequacy of single-locus approximations for linkage analyses of oligogenic traits: extension to multigenerational pedigree structures. Hum Hered 43:329–336

Vieland V, Greenberg DA, Hodge SE, Ott J (1992*a*) Linkage analysis of two-locus diseases under single-locus and two-locus analysis models. Cytogenet Cell Genet 59:145–146

Vieland VJ, Hodge SE, Greenberg DA (1992*b*) Adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 9:45–59

Weeks DE, Lehner T, Squires-Wheeler E, Kaufmann C, Ott J (1990) Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis. Genet Epidemiol 7:237–243

Zinn-Justin A, Abel L (1998) Two-locus developments of the weighted pairwise correlation method for linkage analysis. Genet Epidemiol 15:491–510