

Mapping Complex Genetic Traits in Humans: New Methods Using a Complete RFLP Linkage Map

E.S. LANDER* † ‡ AND D. BOTSTEIN †

*Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142; †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ‡Harvard University, Cambridge, Massachusetts 02138

It has been clear since the rediscovery of Mendel that humans obey laws of heredity identical with those of other organisms. The central features of Mendelism were observable in humans by following simply inherited common traits, including some diseases. However, the systematic study of human heredity using the standard concepts (complementation and recombination, tests of epistasis, etc.) has been infeasible in humans for two reasons: (1) because *Homo sapiens* is not an experimental animal that can be manipulated at will and (2) because few genetic markers had been found that were heterozygous often enough to allow random matings to be informative. The advent of recombinant DNA technology led to the suggestion (Botstein et al. 1980) that common polymorphisms in DNA sequence (conveniently observed as restriction fragment length polymorphisms, or RFLPs) could be used as generally informative genetic markers allowing the systematic study of heredity in humans, including the construction of a true linkage map of the entire human genome.

The application of RFLP technology to simple Mendelian diseases has proceeded rapidly. RFLPs have been found closely linked to the autosomal dominant Huntington's disease (Gusella et al. 1983) and polycystic kidney disease (Reeder et al. 1985), the autosomal recessive cystic fibrosis (Knowlton et al. 1985; Tsui et al. 1985; Wainwright et al. 1985; White et al. 1985), and the X-linked recessive Duchenne muscular dystrophy (DMD) (Davies et al. 1983). In turn, closely linked DNA markers have made it possible to localize the disease genes to chromosomal regions (Gusella et al. 1983; Knowlton et al. 1985; Reeder et al. 1985; Wainwright et al. 1985; White et al. 1985), to undertake prenatal diagnosis of fetuses known to be at risk, and to begin chromosome walks in an attempt to clone the disease genes (Kunkel et al.; Worton et al.; all this volume). In addition, the RFLP linkages have shed light on the formal genetics of the disorders—showing, for example, that mutations at a single locus account for all (or almost all) cases of cystic fibrosis (Donis-Keller et al.; Tsui et al.; Williamson et al.; White et al.; all this volume) or of Huntington's disease (Gusella et al., this volume). These successes make clear that RFLPs can be found linked to any common human disease that shows simple Mendelian transmission and is caused by a single genetic locus.

Much of what we want to know about human heredity, however, concerns traits whose underlying genetics is less favorable for analysis. Some apparently identical clinical conditions can result from mutations at any one of several genes—a circumstance called *genetic heterogeneity*. Some traits have *incomplete penetrance*, with only a fraction of those carrying the appropriate mutant genotype actually displaying the trait. Conversely, some genotypes predispose individuals to a disease, but those of normal genotype may be affected as well, just at a lower rate. Environment may play an important role in the expression of the trait. In addition, *gene interactions* can occur, in which a phenotype results from the interaction of alleles at more than one locus. These complexities can underlie even the most clear-cut phenotypes.

These complex modes of inheritance are common in genetically well-studied organisms, such as bacteria, yeast, nematodes, and fruit flies, and evidence is accumulating that humans are no different. Geneticists usually surmount the problems by isolating pure-breeding stocks, each carrying a mutation at a single locus, and then arranging crosses at will. In the case of human genetics, however, we must take crosses as we find them.

Unfortunately, even with RFLPs as markers, traditional methods for genetic linkage analysis are quite inefficient at mapping genetically complex traits: A prohibitively large sample may be required before linkage can be detected or other genetic properties tested. Furthermore, traditional linkage analysis absolutely requires families with two or more affected individuals. Yet, many traits of biological or medical interest are quite rare, with most cases being sporadic. Collecting enough pedigrees with multiple affected subjects may be impossible.

It may thus appear, at first sight, that the vast majority of human heredity must remain refractory to genetic mapping, due to complexity or rarity. On the contrary, it is the main thesis of this paper that this need not be so. The important point is that studying the segregation patterns of a large number of mapped RFLPs simultaneously (rather than one at a time) can make it feasible to map many traits that are genetically complex and/or rare. We discuss below new mathematical techniques, unusual genetic resources, and clinical ap-

proaches that can potentially be brought to bear on the problem.

The RFLP Linkage Map

More-powerful methods of genetic analysis all require a linkage map of RFLPs covering the entire human genome. Such a genetic map was proposed at a time when only a few human RFLPs were known to exist, but the subsequent isolation of more than 1000 RFLPs (Willard et al. 1985; Donis-Keller et al., this volume), some as polymorphic as the HLA antigen system, in less than a decade has made clear the feasibility of the map. Indeed, construction of such a map is already well underway as part of an international collaboration (J. Dausset, pers. comm.). The eventual existence of a human RFLP map thus seems assured.

We shall assume, for simplicity, the availability of "perfect" RFLP linkage maps. By this, we mean linkage maps of RFLPs evenly spaced throughout the genome (e.g., one every 20 cM in a human genome of 3300 cM), with each RFLP so highly polymorphic that it is rarely found homozygous. Obtaining roughly even spacing presents no difficulty, although it obviously requires sorting through more than the bare minimum number of RFLPs. On the other hand, assuming that each RFLP is so highly polymorphic is unrealistic. However, if the RFLPs are informative only half the time, for instance, one can compensate by doubling the number used. (This trade-off between marker density and informativeness is accurate to a first-order approximation for family studies.) Thus, the actual number of RFLPs needed to simulate a perfect 20-cM map will vary with their informativeness.

Linkage Analysis via Likelihood Ratios

One cannot simply "count recombinants" in a human cross, because of uncertainty about the genotype of individuals or, especially, about the "phase" of markers in the parents (i.e., whether the markers are *cis* or *trans* to each other). Human linkage analysis is therefore done by the method of likelihood ratios (Fisher 1935; Haldane and Smith 1947; Morton 1955). One compares the probability that the observed data would arise under one hypothesis (e.g., linkage at 10% recombination between two markers) with the probability that it would arise under an alternative hypothesis (typically, nonlinkage). The ratio of these probabilities is called the *odds ratio* for one hypothesis relative to the other. Odds ratios from independent families can be multiplied together; when the odds become overwhelming, linkage is considered proven. By convention, one requires that the pedigree data yield an odds ratio of 1000:1 in favor of linkage over nonlinkage. (This threshold is less stringent than it may appear. A priori, the odds are 50:1 against any two loci being linked. Thus, even though the data may yield an odds ratio of 1000:1, the actual a posteriori odds in favor of

linkage are only about 20:1, corresponding to the 5% confidence level.) Finding 100:1 odds against linkage is the conventional threshold for rejecting linkage. Once linkage is established, one estimates the recombination fraction as the value θ at which the likelihood ratio is largest—the so-called *maximum likelihood estimate*.

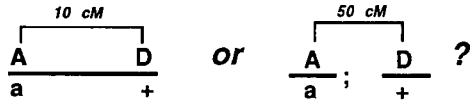
For convenience, human geneticists work with the \log_{10} of the odds ratio, called the *LOD score*. LOD scores from successive pedigrees are thus added until the score grows to 3 (signifying linkage) or falls to -2 (indicating nonlinkage). Figure 1a illustrates the simple case of a phase-known meiosis with a dominant trait and a RFLP marker locus. If the loci are actually linked at 10%, then with probability 0.9 the LOD score will be $\log_{10}(9/5) = +0.25$, and with probability 0.1 it will be $\log_{10}(1/5) = -0.70$. The *expected LOD score*, or ELOD, will be $+0.16$. For the LOD score to reach the threshold of 3 required to demonstrate linkage, one would need to study about 19 ($\approx 3/0.16$) such meioses. If the loci are actually unlinked, the ELOD will be -0.22 and about 9 ($\approx -2/-0.22$) meioses will suffice to show nonlinkage. In this simple example the LOD score approach is hardly necessary, but it can be extended straightforwardly to exceedingly complicated pedigrees plagued by many uncertainties, with likelihoods of the pedigrees under competing hypotheses being calculated by computer.

Interval Mapping

With a complete RFLP linkage map of the human genome, one can test much more precise hypotheses than one can study using only unmapped marker loci. Because the hypotheses are more demanding, they are easier to prove, if correct, or to disprove, if not. The simplest example, which we call *interval mapping*, is illustrated in Figure 1b. In studying a dominant trait caused by a single genetic locus, we can compare the hypothesis that the trait maps in the middle of the interval between an adjacent pair of RFLPs with the hypothesis that it is unlinked to the interval. The hypotheses make very different predictions about the alleles we expect to see in an affected individual (see Fig. 1b). If the first hypothesis (linkage) is correct, the ELOD will be 0.22 per meiosis. Only about 13 meioses are thus needed to prove linkage, as opposed to 19 when a single unmapped marker is used. If the trait is unlinked, only 4 (as against 9) meioses are needed to show nonlinkage and allow one to move on to a new interval.

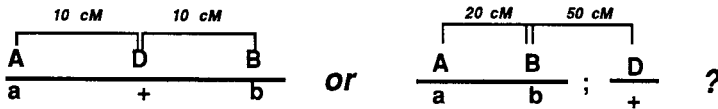
Intuitively, we can understand the power of interval mapping by thinking of two flanking markers as comprising a single "virtual" RFLP. This "virtual" RFLP is informative in meioses in which the flanking markers do not recombine and, in such meioses, it segregates as if it were very tightly linked to all loci in the interval: It will appear to "recombine" with a locus in the interval only when a double crossover occurs (the chance of which is small, even for relatively large intervals). In essence, we need to consider only two alternatives in searching for linkage: An interval is either

a. Using Single Markers



Possible Genotypes of Affected Individual	Frequency of Genotype		Ratio of Odds Linked/ Unlinked	Expected Contribution to LOD Score	
	If Linked as shown	If Unlinked		If linked	If unlinked
A	.90	.50	9 / 5	0.9 log 9/5	0.5 log 9/5
a	.10	.50	1 / 5	+ 0.1 log 1/5	+ 0.5 log 1/5
Expected LOD Score.....				+ 0.16	- 0.22

b. Using Interval Mapping



Possible Genotypes of Affected Individual	Frequency of Genotype		Ratio of Odds Linked/ Unlinked	Expected Contribution to LOD Score	
	If Linked as shown	If Unlinked		If linked	If unlinked
A B	.81	.40	81 / 40	.81 log 81/40	.40 log 81/40
A b	.09	.10	9 / 10	.09 log 9/10	.10 log 9/10
a B	.09	.10	9 / 10	.09 log 9/10	.10 log 9/10
a b	.01	.40	1 / 40	.01 log 1/40	.40 log 1/40
Expected LOD Score.....				+ 0.224	- 0.520

Figure 1. Example of calculation of expected LOD score (ELOD), using a single genetic marker (*a*) and using interval mapping (*b*) in the case of a single phase-known meiosis and a dominant trait (D). In *a*, we compare the hypothesis that the trait is linked to the marker at 10 cM to the hypothesis that it is unlinked. In *b*, we compare the hypothesis that the trait lies between two markers known to be separated by 20 cM to the hypothesis that it is unlinked. (For simplicity, cM and recombination fraction are taken to be the same; no mapping function is used. In all other figures, the Haldane mapping function—corresponding to no interference—is used. Allowing for interference slightly increases the power of the methods.)

very tightly linked or else it is completely unlinked. This sharp dichotomy makes analysis easier and more powerful (even more so when genetic complications are involved, as we shall see).

In Figure 2, *a* and *b*, the number of families needed to map a simple Mendelian dominant or recessive trait using interval mapping is compared with the number needed using the traditional single-marker approach. The precise comparison depends on the spacing between consecutive RFLPs and (in the case of a recessive trait) the number of affected children per family, but one can expect to need, roughly, a sample only 40–60%

as large. Clearly, interval mapping makes much more efficient use of the limiting resource, that is, families with affected members.

In planning a linkage study, it is prudent to collect more than just the expected number of families needed. Figure 2, *c* and *d*, therefore shows the sample required to ensure a 95% chance of detecting linkage, should it be present. It is striking to note (Fig. 2d) that a sample of families with three affected children, which is large enough to ensure a 95% chance of finding linkage using interval mapping, will not suffice to provide even odds of finding linkage by single-marker methods.

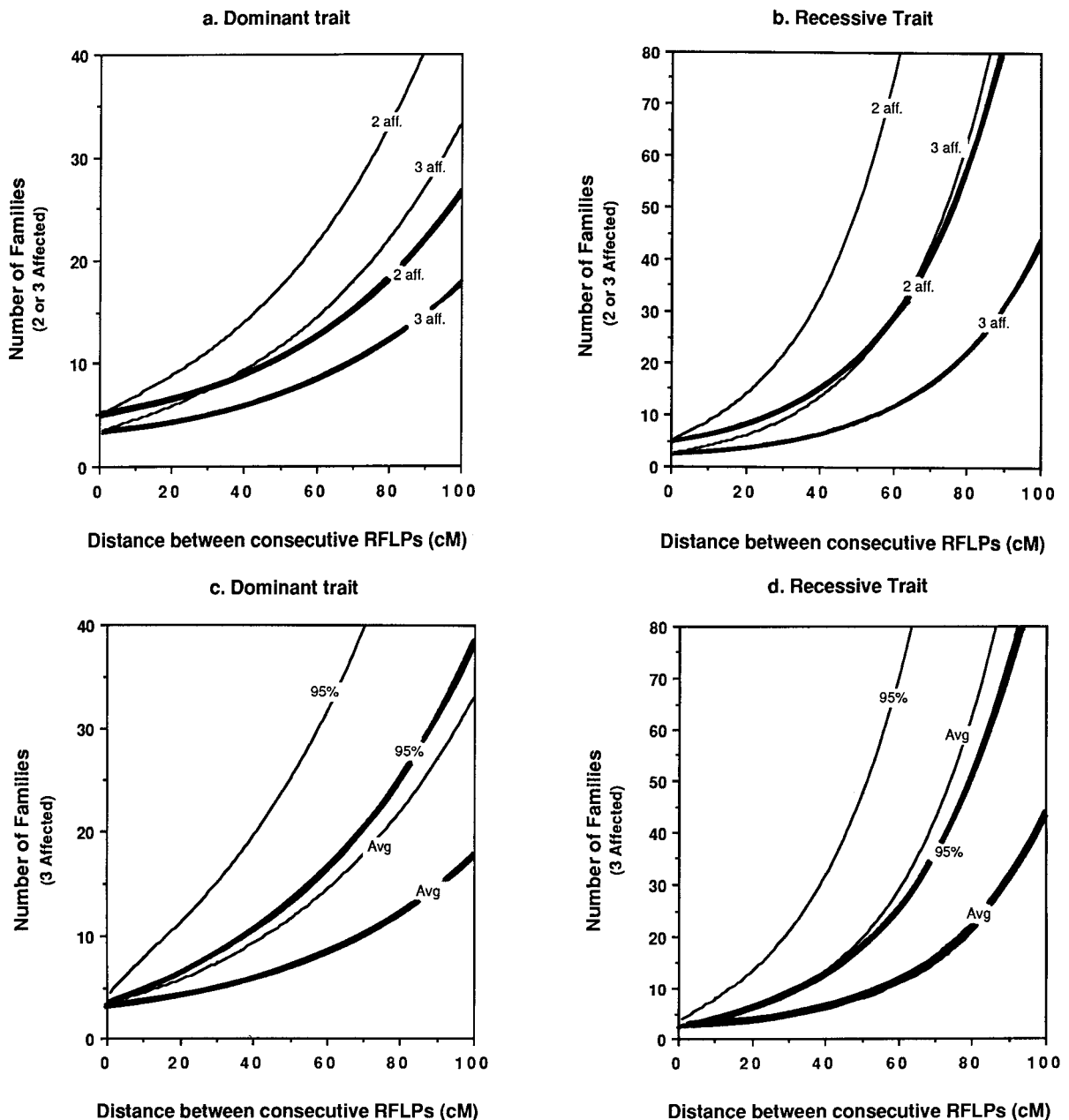


Figure 2. The number of families needed to map simple Mendelian traits with completely informative RFLPs—comparison of single-marker method (thin curves) with interval mapping (thick curves). Phase is assumed to be known for the RFLP markers and for a dominant trait, but the phase of a recessive trait is assumed to be unknown. For a dominant trait, the graphs show the expected number of meioses needed to reach a LOD score of 3.0 (a) and the number needed to ensure a 95% certainty of reaching a LOD of 3.0 (c). For a recessive trait, the graphs show the expected number of families of different types (classified by the number of affected sibs) to reach a LOD score of 3.0 (b) and the number needed to ensure a 95% certainty (d). The number of affected individuals in the family is relevant because the phase of the trait is presumed to be unknown in the recessive case; sibs thus contribute information about phase. In the phase-known dominant case, meioses can be treated independently.

Genetic Complexities

The current (sixth) edition of McKusick's (1983) catalog *Mendelian Inheritance in Man* lists 3,368 traits known or believed to mendelize to some extent. A few, like cystic fibrosis or Huntington's disease, exhibit perfect Mendelian transmission (and are now known to map to a single locus). Many, perhaps most, surely involve more complex genetics, which can greatly complicate the search for linkage. We consider the pros-

pects and methods for mapping traits that (1) are genetically heterogeneous, (2) show incomplete penetrance, (3) amount to predispositions, or (4) involve gene interactions. More-detailed treatments will appear elsewhere (Lander and Botstein 1986 and in prep.).

Genetic heterogeneity: The method of simultaneous search. By far the most serious obstacle to linkage studies will be genetic heterogeneity, the situation of a phenotype that can be caused by mutations at any one

of several loci. The classic paradigms of heterogeneity, well-studied in lower organisms, are (1) the interruption of a single biochemical pathway at any of its steps and (2) the loss or disruption of a heteromultimeric protein complex by mutation in the structural gene for any subunit, although other possibilities (including regulatory genes, posttranslational modifications, etc.) are not hard to imagine. For example, hereditary methemoglobinemia, once thought to be a homogeneous clinical entity, can be produced by mutations in either the α or β chains of hemoglobin or in NADH dehydrogenase (Stanbury et al. 1983). Elliptocytosis (Morton 1956) and Charcot-Marie-Tooth disease (Bird et al. 1983; Dyck et al. 1983; see also Ott 1985) are both known to be genetically heterogeneous, because in each case linkage to a marker has been seen in some large pedigrees but is absent in others. Xeroderma pigmentosum and ataxia telangiectasia are almost certainly genetically heterogeneous, since *in vitro* assays on cell fusions reveal nine and five complementation groups, respectively (Jaspers et al. 1982, 1985; Keijzer et al. 1979). In the early 1950s, two albinos married and produced normal children (Trevor-Roper 1952), confirming that albinism is heterogeneous (as had already been suspected from phenotypic distinctions and population genetics: i.e., much higher consanguinity among parents than expected for a single gene). There is also very good evidence that congenital deafness, congenital blindness, and coronary heart disease are heterogeneous in cause (Stanbury et al. 1983).

Heterogeneity has traditionally been the human geneticist's nightmare because evidence for linkage to a locus in one family will be offset by evidence against linkage in another family. Even a modest degree of heterogeneity may cause the traditional LOD score to be negative even for a marker tightly linked to one of the loci involved.¹ One solution is to concentrate only on a pedigree so large that linkage can be proven without the need for other evidence. Such situations are quite rare, as well as tedious to collect.²

Another solution is to allow explicitly for heterogeneity in the hypotheses to be tested, by supposing that only a certain fraction α of families are of the linked type. Such *admixture* methods (Ott 1985; Cavalli-Sforza and King 1986) have been used with single markers to detect linkage and/or test for heterogeneity. Single-marker admixture methods, however, are especially weak at detecting that a trait is heterogeneous: It is very difficult to distinguish whether (1) a marker is loosely linked to the disease-causing locus in all families or (2) it is tightly linked to the disease-causing locus in some families but unlinked in others.

We can modify the admixture method by incorporating interval mapping. Fewer families are required to detect linkage. The savings is of the same proportion as for a homogeneous trait, although it is more important here because the number of families involved is larger (Fig. 3a). Even more dramatic is the improvement in detecting the presence of heterogeneity—the number of families needed decreases from 5- to 50-fold

in some typical cases (Fig. 3b). The reason is clear: Since interval mapping is essentially a test of very tight linkage (to a “virtual” RFLP), we no longer can confuse heterogeneity with loose linkage.

A complete RFLP map affords an even more powerful strategy for detecting linkage to a genetically heterogeneous trait. We call this approach *simultaneous search*, because it involves studying the segregation of several candidate loci simultaneously to see whether, as a set, they account for the transmission of the trait. From a mathematical point of view, we compare the likelihood of the hypothesis that in each family the trait cosegregates with one of the candidate loci (although which one may depend upon the family) with the likelihood of the hypothesis that the trait is unlinked to any of the loci. We examine in turn all pairs of loci (then triples, quadruples, etc.) to find a set that together explains the transmission in all of the families being studied. Because we are considering a larger number of hypotheses than usual, the threshold for the LOD score must be set higher than 3 to guard against false positives. A more detailed explanation of the mathematics appears in Lander and Botstein (1986).

From a more intuitive point of view, we are searching for a set of loci with the distinctive property that, in every family, one of the loci segregates as if it were tightly linked to the trait. Figure 4a illustrates a simplified situation in which the genome consists of nine intervals, A–I, and we have 20 isolated opportunities to observe a crossover between each interval and some dominant trait. If the trait were homogeneous, we would expect that some interval would show zero crossovers or maybe one (recall that interval mapping exhibits very tight linkage). The sample “observed” data are not consistent with this expectation. The data are consistent with the number of crossovers expected if the trait were sometimes caused by a locus in interval B and sometimes by one in interval F. However, the evidence is far from convincing. A better way to analyze the data is to ask how often the trait recombines with both of the two intervals: For the correct pair of intervals, B and F, we get the expected result of zero (Fig. 4b). Intuition suggests (and mathematics confirms) that this is more powerful evidence. This illustrates the basic principle behind simultaneous search, although in practice it is preferable to use families with several meioses rather than isolated meioses, as in the example.

We have calculated the number of families required to map heterogeneous diseases by simultaneous search (Fig. 4c). The results suggest that heterogeneous diseases caused by as many as five loci can be resolved with a feasible number of families.

¹Accordingly, the practice of *exclusion mapping* (declaring that a trait is “excluded” from a region once a LOD score of -2 is reached) is of dubious value in the general case.

²Moreover, in each large pedigree one maps just one of the alternative causative genes, which may be a minor one in the population.

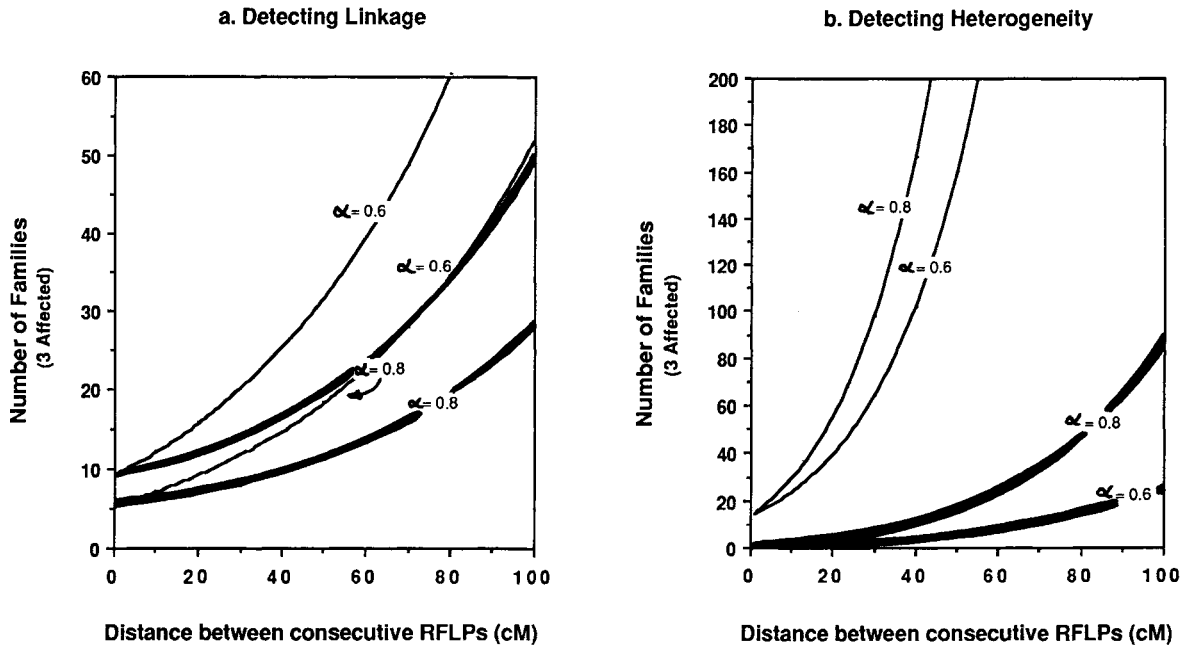


Figure 3. (a) The number of families with three affected sibs needed to map a heterogeneous dominant trait via admixture methods—comparison of single-marker method (thin curve) with interval mapping (thick curve). The curves pertain to a trait in which the fraction (α) of families segregate for a linked locus. (b) The number of families with three affected sibs needed to detect that a trait is heterogeneous via admixture methods—comparison of single-marker method (thin curve) with interval mapping (thick curve). Phases are assumed to be known.

The advantage of simultaneous search is that the set of loci that causes a trait often can be recognized before any single locus could be recognized by studying it alone. Simultaneous search offers no advantage if only one cause is a genetic locus, with other families showing the trait for nongenetic reasons, such as a virus. In such cases, admixture methods must suffice.

We should note that, although heterogeneity wreaks havoc with traditional linkage analysis, it is not even detectable by segregation analysis (the studying of pedigrees to see if they fit Mendelian patterns of transmission). No matter how many independent loci can cause a recessive trait, each family segregates the expected Mendelian 3:1 ratio. This is in sharp contrast to genetic traits showing incomplete penetrance. Incomplete penetrance disrupts the expected segregation ratios (making it hard to prove Mendelian transmission by segregation analysis), but it does not seriously complicate linkage mapping.

Incomplete penetrance: Only affected individuals contribute significantly. A genetic trait that requires additional factors to become manifest may indicate incomplete penetrance. These factors might include environment, genetic background, or chance. A simple example is retinoblastoma, in which a second somatic alteration leads to cancer in individuals carrying the dominant allele; the event is quite frequent, so penetrance is high (Macklin 1959; Cavenee et al. 1983). Wernicke-Korsakoff syndrome seems to be due to a mutation causing transketolase to bind thiamine pyrophosphate less avidly than normal, but a clinical phenotype (alcohol-induced encephalopathy) is apparent

only in patients with a dietary thiamine deficiency (Blass and Gibson 1977). As this example illustrates, the degree of penetrance can vary widely with the environment and with the precise definition of the phenotype (encephalopathy vs. thiamine pyrophosphate binding). Finally, diseases with late age of onset must always be treated as incompletely penetrant.

When penetrance is incomplete, the genotype of the unaffected individuals in a sibship is in doubt. Not surprisingly, relatively little can be learned by studying the markers that they have inherited. In fact, the problem of the uncertain genotype of unaffected individuals is evident even for a recessive trait of complete penetrance. There are three possible genotypes for an unaffected sib, but only one for an affected sib. For a tightly linked marker then, the odds ratios will be 4:3 for an unaffected and 4:1 for an affected sib, making the contribution to the LOD score only one fifth as large for unaffected sibs. Even a small degree of incomplete penetrance creates a similar problem for dominant traits and further exacerbates it for recessive traits. Figure 5 shows how quickly the contribution to the LOD score plummets as the genotype of an unaffected sib becomes uncertain. Below about 80% penetrance, unaffected sibs are virtually useless. In assessing the value of a pedigree for linkage analysis then, it is usually accurate to look only at the affected sibs.³

Incomplete penetrance, however, does not confound

³It is nevertheless advisable to collect the unaffected sibs, where practical, since they may aid in determining the phases of the RFLP markers used for mapping.

a. How often does each interval segregate away from the disease?

b. How often do both B and F segregate away from the disease?

Interval	A	B	C	D	E	F	G	H	I
Crossovers "Observed"	11	6	8	13	10	4	12	8	9
Crossovers Expected:									
if trait in B	10	0	10	10	10	10	10	10	10
if trait in F	10	10	10	10	10	0	10	10	10
if in B <u>or</u> F	10	5	10	10	10	5	10	10	10
Total Crossovers Possible: 20									

	A	B	C	D	E	F	G	H	I
A		3	5	7	5	3	9	5	5
B			3	4	5	0	5	3	3
C				3	4	2	5	3	4
D					7	3	12	7	8
E						2	8	4	5
F							3	2	2
G								7	12
H									4

c. Number of families needed to map a heterogeneous dominant trait

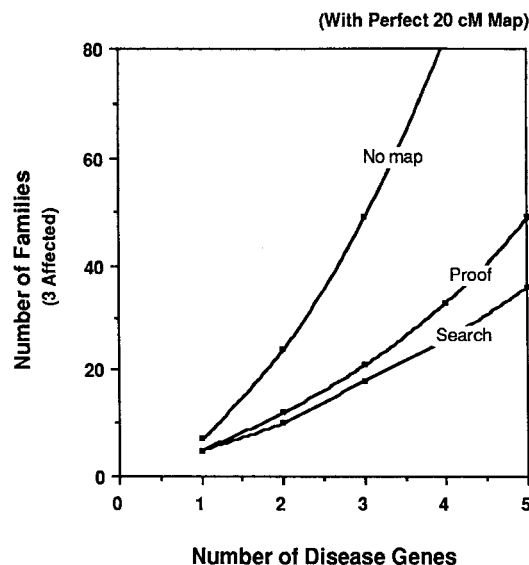


Figure 4. Illustration of the principle behind the simultaneous search method and the power of this approach for mapping heterogeneous traits. In *a*, a hypothetical genome is divided into nine intervals (A,B . . . I) flanked by completely informative RFLP markers. Twenty informative meioses are scored, resulting in the "observed" data for crossovers with a dominant disease gene shown in the second line. The number of crossovers expected is shown, under three different hypotheses about the location of the disease gene. The third hypothesis, that the disease is caused equally often by a gene in B and a gene in F, fits the data best, but the fit is not convincing. Instead, in *b*, we record the number of meioses in which the trait is seen to recombine with *both* of a pair of intervals. Here, the pair of intervals B and F stand out more markedly. The number of families with three affected sibs required to map the genes causing a heterogeneous recessive trait or a heterogeneous dominant trait is shown in *c* and *d*, respectively. In each case, the genes causing the trait are assumed to be equally frequent. The curve marked "Search" pertains to proving that the set of loci shows linkage according to simultaneous search; the curve marked "Proof" pertains to subsequently proving the involvement of each of the loci in the set. The curve marked "No map" pertains to using admixture methods alone. Phase is assumed to be known for RFLP markers and for dominant traits but not for recessive traits.

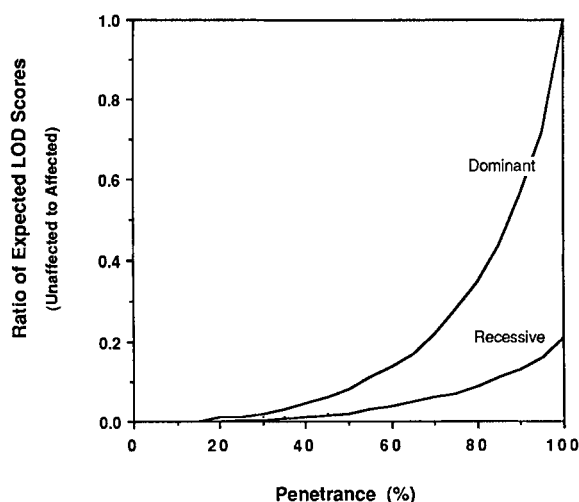


Figure 5. Relative values for linkage mapping for affected and unaffected sibs as a function of penetrance of a disease gene. Specifically, the graph shows the ratio of the ELOD for an unaffected sib to that for an affected sib, when a marker is available at 0 cM. All phases are assumed to be known.

linkage analysis in the serious way that heterogeneity does. The practical consequences are limited to the following:

1. Finding families with several affected sibs (and an affected parent, for a dominant trait) will be more difficult than if the trait were fully penetrant. A possible solution is to identify a more penetrant-related phenotype, a strategy we return to below.
2. Given the same set of pedigrees, however, a recessive trait of incomplete penetrance should not be significantly harder to map than a fully penetrant one. Unaffected sibs are of marginal value in both cases.
3. Given an affected parent, a dominant trait may require up to twice as many meioses be studied, since unaffected individuals no longer contribute. Families without an affected parent contribute less information, since the parental source of the disease gene is unknown.

Predispositions: The need to enrich for genetically caused cases. Genetic predispositions are not well understood in general, but some spectacular examples exist. Some rare families clearly segregate a dominant trait, causing at least a 100-fold increase in the risk for specific cancers (McKusick 1983). Homozygotes for a mutant allele of the gene for α_1 -antitrypsin show a 20-fold higher risk of emphysema than normal (Stanbury et al. 1983), although it should be noted that most cases of emphysema are due to cigarette smoking, not α_1 -antitrypsin deficiency. The relative risk of developing type-1 diabetes is 9-fold higher in individuals with HLA genotype *Dw3/Dw4*; similarly *DQ* alleles have been associated with myasthenia gravis (Stanbury et al. 1983; see also Bell et al., this volume). Genes involved in predispositions may identify steps in the pathway of pathogenesis.

Mapping a predisposition to a common disease may be difficult because of the need to discern the “genetic” families (those segregating the predisposition) from the “nongenetic” families (those with more than one affected individual just by chance). Indeed, *most* families with two affected members may be due to *nongenetic* causes if the predisposition is weak, the genotype causing it is rare, or the disease is common in the general population. For example, since one in four Americans develops cancer over a lifetime, two sibs with cancer hardly suggests that the family segregates for a predisposing locus. Linkage analysis can be undertaken only if one can enrich for families with a genetic cause, amid an excess of nongenetic cases (cf. Fig. 3a concerning the admixture of causes).

The best way to enrich is to seek families with many affected individuals: If nongenetic cases occur independently in the general population, then a cluster of, for instance, four or five cases more likely will be due to inheritance rather than chance.

To illustrate the point intuitively, suppose that some families segregate a dominant trait causing a 20-fold increased risk for a disease. The chance that a child in such a family develops the disease is 10-fold higher than for a normal family (a 50% chance of inheriting the trait multiplied by a 20-fold increased risk); the chance of two children developing the disease is 100-fold higher than for a normal family, and the chance of three children in the family is 1000-fold higher.⁴ Thus, even if genetically predisposed families represent only 1% of the population, they will comprise more than 90% of the families with three affected members.

The real difficulty occurs if nongenetic cases do not occur independently (e.g., due to common environmental exposure). Finding multiple cases in a family does not then ensure as great an enrichment. Because of this, in designing a linkage study, it will be important to exploit what is known about the epidemiology of the disease in order to enrich most effectively (e.g., choosing families without obvious environmental risk factors).

In many cases, it should be possible to map genes causing predispositions of at least 10-fold (calculations not shown). Weaker predispositions, however, will likely be very difficult to map.

Gene interactions. Finally, traits may result from genetic interactions between alleles at several loci. Examples abound in other organisms but have been hard to elucidate in humans. It is known, though, that α -thalassemia partly suppresses β -thalassemia. It is also suspected that at least some diseases that show partial association with HLA genotype may involve other loci as well.

The most fundamental gene interaction is a *synthetic trait*, one caused only when mutant alleles are present

⁴The exact risks also depend slightly on the number of unaffected members in the family and the incidence of the disease in the general population, but we ignore this for the sake of illustration.

at all of several loci (to be distinguished from a heterogeneous trait, which is caused by mutations at any of several loci). At some of the *component loci*, the alleles may act dominantly; at others, they may act recessively. Some may be rare; others may occur at a high frequency. The incidence of the trait in the population will reflect the product of the frequencies of the component loci.

By far the most difficult case for linkage mapping is a component locus at which the allele (mutant) frequency is very high (e.g., 50%): A large fraction of parents of affected individuals will be homozygotes at the locus but will be unaffected themselves since they lack other components. Since an affected child may have inherited either chromosome from such a parent, the trait will *not* map to the locus in such a family. Analytically, it is the same as admixture with a large proportion of families having a nongenetic cause (cf. Fig. 3a above).

On the other hand, the trait will show clear linkage to each component locus at which the allele frequency is not too high (e.g., less than 10%). Linkage analysis is relatively straightforward:

1. Recessive components involve no special considerations and require the same number of families as for a simple recessive trait (Fig. 2a).
2. Dominant components are slightly harder, since one will typically not know which parent contributed the mutant allele. The number of families needed to map a dominant trait under these circumstances is shown in Figure 6.

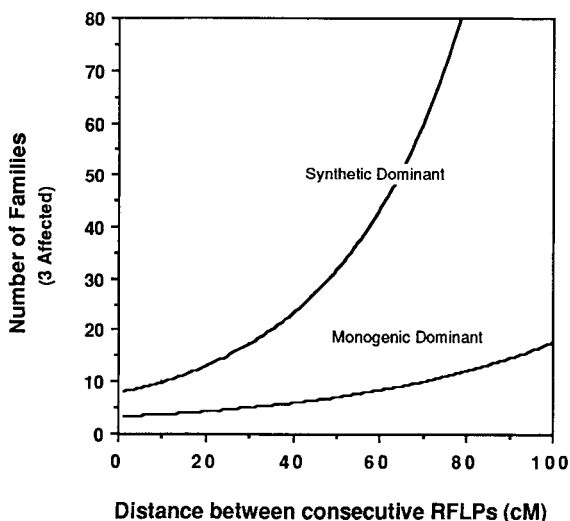


Figure 6. The number of families needed to map a dominant component allele involved in causing a synthetic trait via gene interaction. Which parent contributes the allele, as well as the phase of the allele, is assumed to be unknown. The phase of RFLP markers is assumed to be known. The allele is assumed to be present at less than 5% in the general population, so that one parent may be assumed to be a heterozygote and one parent homozygous normal (see text). For comparison, the number of families needed when a trait is caused monogenically by a dominant allele is also shown.

Component loci with intermediate allele frequency (10–50%) require special precautions and careful selection of families. For example, contrary to intuition, a very large proportion of affected individuals in a family may be *undesirable* since they may enrich for parents who are homozygotes. (We shall elaborate on these issues elsewhere.) Overall, it appears that component loci involved in synthetic interactions should be mappable if the frequency of the mutant allele at the locus is under 20%.

More-complex genetic interactions can be analyzed in a similar fashion. *Multifactorial inheritance* (which usually refers to additive interactions between a number of loci) can be modeled as involving a combination of heterogeneity and synthetic interactions. A detailed treatment of these issues, however, is beyond the scope of the present work.

Mapping without Family Studies

For many of the genetic disorders in McKusick's catalog, it will be difficult to collect enough families, each with several affected members, for traditional linkage analysis. The disorders are rare, occurring sporadically in isolated cases or small clusters.

We propose two methods for linkage analysis when family studies are not sufficient or not possible at all. *Homozygosity mapping*, appropriate for a recessive disease, exploits the unusual genetic constitution produced by inbreeding. *Disequilibrium mapping* attempts to use special features of the population genetics of genetic isolates.

Homozygosity mapping: Inbred individuals contain a surprising amount of information. In his classic study on inborn errors of metabolism, Garrod (1902) observed that an unusually high proportion of patients with alkaptonuria were progeny of consanguineous marriages. Almost immediately, Bateson (1902) proposed a Mendelian explanation: "The mating of first cousins gives exactly the conditions most likely to enable a rare, and usually recessive, character to show itself." The rarer the disease gene, the more pronounced will be the frequency of inbreds among those affected. Consanguinity was found in 82% of all cases of Brazilian type achropody (McKusick 1983) and at least 25% of all cases of Tangier disease (Stanbury et al. 1983). By contrast, consanguineous marriages account for only about 3% of cases of cystic fibrosis, for which about 1 in 20 people is a carrier (Romeo et al. 1985).⁵

Children of first-cousin marriages provide a very sensitive assay for the location of a recessive disease gene: They will be homozygous by descent not simply at the disease locus, but throughout an entire chromo-

⁵Of course, the rate of consanguinity among parents of affected children depends on the rate of consanguineous marriages in the general population, which varies widely between different countries.

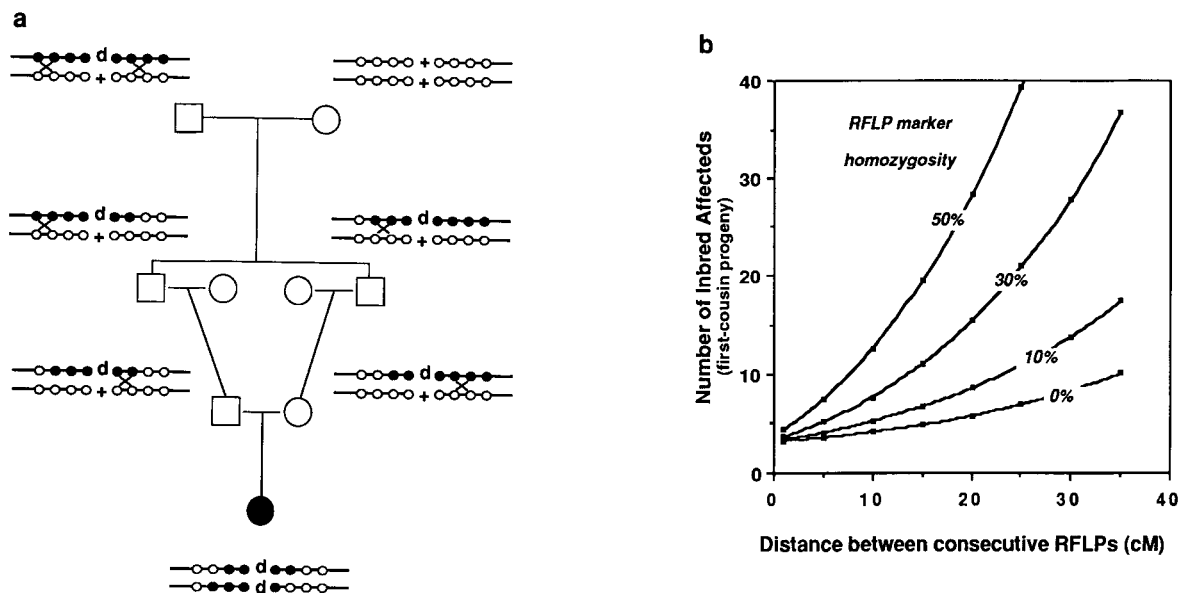


Figure 7. Illustration of the principle behind homozygosity mapping and the power of this approach for mapping a recessive trait. In *a*, the pedigree of a first-cousin marriage is given, showing the descent of marker alleles on the chromosome bearing the recessive disease allele. Homozygosity by descent for a contiguous stretch of markers in the affected child is shown. In *b*, the expected number of affected first-cousin progeny needed to map a recessive trait (i.e., to reach a LOD score of 3.0) as a function of the marker density of a RFLP map is shown. The various curves pertain to RFLPs of different degrees of informativeness: ones that are found homozygous 50%, 30%, 10%, and 0% of the time in the general population.

somal region nearby⁶ (Fig. 7a). Of course, a region unrelated to the disease might become homozygous by descent in the inbred child simply by chance, but such regions will differ from child to child. Since the chance of an unrelated region becoming homozygous by descent is only 1/16 (the *coefficient of inbreeding* for a first-cousin marriage), then if a region is seen to be homozygous by descent in three unrelated inbred children, the odds are 16³:1 (i.e., 4096:1) that it contains the disease gene. In principle then, three inbred children suffice to map a homogeneous recessive disease.

All we need to carry out this strategy, which we call *homozygosity mapping*, is an assay for whether a region is homozygous by *descent*. A RFLP map provides such an assay: We check whether all the RFLPs in a stretch of the genome are homozygous. The power of homozygosity mapping depends on the accuracy of this assay, which in turn depends on the spacing between consecutive RFLPs and their degree of polymorphism (specifically, the chance that a RFLP will be found homozygous in the general population just by chance). Using a 5-cM map of RFLPs that are homozygous 50% of the time in the population, one can map a recessive trait based on about eight inbred children (Fig. 7b). A 10-cM map of RFLPs homozygous 30% of the time, or a 20-cM map of RFLPs homozygous 10% of the time, would be equivalent. Doubling the density of the RFLPs cuts the requirement to about five children in each case.

⁶There is a chance that consanguinity is not the cause of the genetic disorder, but for first-cousin marriages it is small, unless the recessive allele is very common. This possibility must be included in the analysis.

Using homozygosity mapping with a high-quality RFLP map, a single affected child born of first cousins is roughly as informative as a nuclear family with three affected children.⁷ The advantage is that, for many traits, inbred children will be easier to locate than families with three affected children. Of course, there is no need to choose between homozygosity mapping and traditional family studies; information can be combined by simply adding the LOD scores. In those cases where families with three or more affected sibs are particularly helpful (such as simultaneous search for a heterogeneous trait), inbreds should similarly prove valuable.

Children of second-cousin marriages are roughly as informative. (The probability that an unrelated region of the genome is homozygous by descent is even smaller, 1/64. On the other hand, there is a greater chance that markers near the disease locus will have recombined and/or that the disease will not be the fault of inbreeding.) For more distant consanguinity in a large population, the probability that the disease is due to homozygosity by descent (rather than random chance) falls off exponentially, making such cases much less powerful for homozygosity mapping.

Disequilibrium mapping: Recent genetic isolates may contain enough information to contribute to linkage analysis. When a fresh mutation occurs and spreads through a population, it tends to be coinherited along

⁷Two affected sibs from a first-cousin marriage are equivalent to a family with four affected sibs, and so forth.

with whatever alleles were present on the ancestral chromosome at nearby loci. Eventually, the alleles at nearby loci will be randomized by the process of recombination. But, at earlier times, one will observe *linkage disequilibrium*, an excess of one allele on chromosomes bearing the mutation. As Bodmer points out in his introduction to this volume, linkage disequilibrium should be detectable in the general population only for markers within about 0.1 cM (or about 100 kb of DNA) and then only if the mutation arose only once. Thus, searching for linkage disequilibrium in the general population will be a fruitless strategy for detecting linkage in the first instance (although disequilibrium may be useful once a locus has been narrowed down to a 1-cM interval).

There are some special situations, however, in which *disequilibrium mapping* may be possible. Some isolated populations show unusually high rates of certain genetic diseases, due to the so-called *founder effect*: The mutation was present in one of the original founders and spread by genetic drift while the population was still small. Finland is a particularly well-studied example. Congenital chloride diarrhea, congenital nephrotic syndrome, cornea planta congenita recessiva, and neuronal ceroid lipofuscinosis all occur at higher rates in Finland than in the rest of the world; the clustered ancestry of cases (Fig. 8a) also supports a founder effect in isolates formed about 300 years ago (Norio et al. 1973; A. de la Chapelle, pers. comm.).

The potential power of disequilibrium mapping depends on several factors: (1) the number of chromosomes descended from the founder available for study; (2) the number of generations since the isolate was founded; (3) the spacing between mapped RFLP markers; and (4) the informativeness of the RFLP markers

in the isolated population. Figure 8b illustrates the case of a recessive disease, inherited from a founder 10 generations ago, that now affects 15 individuals (thereby providing 30 chromosomes to examine for disequilibrium). One could expect to map the disease by disequilibrium mapping (i.e., attain a LOD score of 3) if one had an 8-cM map of RFLPs that were 50:50 two-allele systems, a 12-cM map of RFLPs that were three-allele systems, or a 17-cM map of RFLPs that were five-allele systems in the isolated population.

Disequilibrium mapping, alone among the methods discussed here, depends on population genetics rather than simple transmission genetics. Thus, it must be used with special care. It is crucial, for example, that the RFLP allele frequencies found in affected individuals be compared to the frequencies expected for the *isolate*, not for the general population. In passing through the bottleneck when an isolate is founded, allele systems can become either more or less polymorphic due to genetic drift. Populations that passed through very narrow bottlenecks will probably be too nearly monomorphic; HLA and other highly polymorphic systems should provide a helpful indication of this. Test cases will be needed before the power of disequilibrium mapping can be fully assessed.

How Good a Map Is Needed?

How good a RFLP map of the human genome is needed? Before answering this question, we should emphasize that we have been considering "perfect" RFLP maps, that is, ones with fully informative markers evenly spaced along the human genome. In reality, few fully informative RFLPs have been discovered, making it necessary to use many more partially informative

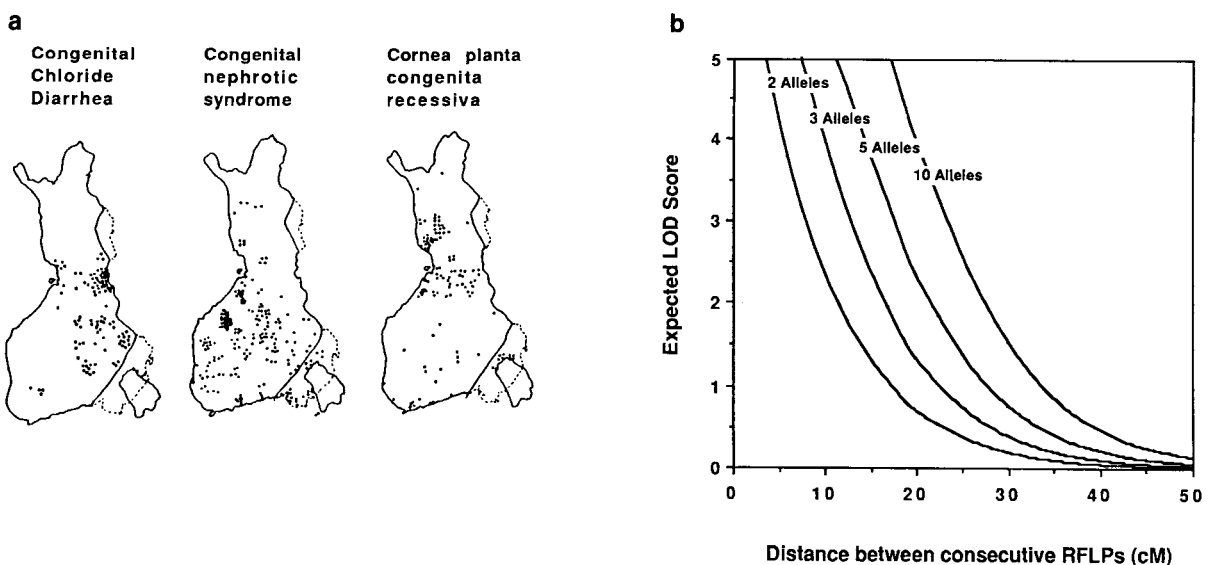


Figure 8. Disequilibrium mapping. (a) Maps of Finland showing geographic clustering of inherited diseases, presumably due to a founder effect. (b) The LOD score expected for an isolate in which 30 chromosomes, descended from an ancestor 10 generations ago, can be studied. The graph plots the LOD score against the marker density of a RFLP map. The various curves pertain to the informativeness of the markers: RFLPs consisting of 2, 3, 5, and 10 equally frequent alleles are illustrated.

RFLPs to approximate a "perfect" RFLP map of any given density.

With this caveat in mind, how good a "perfect" RFLP map of the human genome is needed? The answer varies with the application:

1. For mapping simple homogeneous Mendelian traits, a perfect 40-cM RFLP map should suffice (Fig. 2a,b). Using a perfect 20-cM map would reduce the number of families needed by about half.
2. For more complex genetic traits, using a perfect RFLP map with spacing of 20 cM or less seems advisable (Fig. 4c). This should decrease the number of families needed for these more complex cases to a more manageable level. Using a perfect 10-cM map would reduce the number needed by about half.
3. For homozygosity mapping and disequilibrium mapping, the equivalent of perfect RFLP maps of 10 cM or less are advantageous (Figs. 7b and 8b).

Thus, there are uses for RFLP linkage maps at all scales. That it is feasible to make RFLP maps of any desired resolution has been amply demonstrated by Donis-Keller and colleagues with their recent discovery and mapping of 50 RFLPs on chromosome 7, a genetic length of 140 cM (Donis-Keller et al., this volume).

Despite this achievement, the equivalent of even a perfect 40-cM RFLP map is not yet available for the vast majority of the human genome. The collection of many more RFLPs remains a major task ahead and it will be important to concentrate on finding ones that are highly informative. In fact, fewer than 200 of the 1000 reported RFLP markers come close to being fully informative (Donis-Keller et al.; White et al.; both this volume). The alternative to highly informative markers—using a much greater number of poorly informative ones—may be mathematically equivalent but is clearly inferior from the point of view of the laboratory work involved in applying the map to any specific trait.

DISCUSSION

Even though a complete RFLP map of the human genome is still some years in the future, it has become clear that such a map will be a powerful tool both for the study of human biology and the practice of clinical genetics. We have summarized above ideas and calculations that show the power of a RFLP map in mapping loci that cause, or contribute to, human traits and diseases. The main conclusions are:

1. Simple Mendelian traits can be mapped using fewer families with a RFLP map than is possible using markers one at a time.
2. In some cases, traits can be mapped without family studies at all, but only with a reasonably complete RFLP map.
3. The genetic contributions to diseases or traits of surprisingly complex etiology will be accessible to study, but, again, only with a reasonably complete

RFLP map. In particular: (a) Mapping heterogeneous traits, caused by any of up to four or five different loci, should be possible, but will require a larger number of families. If more than five genes are involved, the task grows much more difficult. (b) Incomplete penetrance is not a major obstacle to linkage analysis, unless penetrance is so low that clustered cases are rare. (c) Mapping predispositions to common diseases should be possible when one can enrich the sample for families with genetic involvement. (d) It should be possible to map two or more loci that interact to cause a trait, except for those component loci at which the trait-causing allele occurs at very high frequency in the population.

Concerning the kinds of genetic resources best suited to analysis with a RFLP map, some conclusions can be drawn as well:

1. Although the traditional focus has been on very large pedigrees, families with two or more affected sibs can also suffice for linkage analysis of a simple Mendelian trait using a RFLP map. For heterogeneous traits, it is preferable to seek families with at least three affected.
2. The value of a family depends crucially on the accuracy of the diagnosis in each member asserted to be affected. Misdiagnosis can seriously hamper linkage analysis.
3. Unaffected family members contribute only slightly to linkage analysis (except for dominant traits displaying full penetrance). Nevertheless, they should be collected when feasible, because they aid in determining the phase of the RFLP markers.
4. Since the amount of DNA eventually required from each individual is not entirely predictable in advance (depending as it does on the complexity of the genetic situation), the practice of "immortalizing" peripheral blood lymphocytes with Epstein-Barr virus (Anderson and Gusella 1984) is recommended. This relatively simple procedure ensures an essentially unlimited supply of DNA and ensures that each donor has to be sampled only once.
5. Affected children of consanguineous marriages are valuable genetic resources, especially for linkage analysis of less common diseases. In collecting samples, it may be profitable to consider the many regions of the world where consanguinity is not infrequent. Also, under appropriate conditions, isolated populations showing founder effects may be useful for linkage mapping.

These are simply general comments. A more precise analysis must be performed, of course, when designing a particular study. Toward this end, we and others have written computer programs to estimate the resources needed, given assumptions about the genetics of a trait.

No matter what methods are employed, there will occur genetic situations beyond the limits of linkage analysis—because of too much heterogeneity, too high a background of nongenetic cases, too complex a genetic interaction. This will become known only in ret-

respect, when linkage is not detected. Certain clinical approaches may help when such problems arise or can be anticipated.

If several genes interact to produce a phenotype, the best route may be through the identification of "subphenotypes" in unaffected family members. With luck, the subphenotype may represent an effect of just one of the interacting genes and thus may be easier to map. A potential example is the photosensitivity frequently found in unaffected close relatives of epileptics (Doose et al. 1982; Botstein and Donis-Keller 1984); investigations of this subphenotype suggest that it might be advantageous to try to map it before, or at the same time as epilepsy. Similarly, clinically measurable aspects of cholesterol metabolism may provide useful subphenotypes to map in families with heart disease (M. Brown, pers. comm.). Finally, if a trait is heterogeneous but a simultaneous search turns up no candidate loci, it may be advantageous to find ways to clinically subdivide the disease into classes, each of which may be mappable.

CONCLUSION

The influence of genetic arguments in the understanding of human biology has been mainly theoretical, primarily because experimental genetics could rarely be applied directly to humans. As many of the articles in this volume make clear, the advent of RFLP mapping has begun to change this, at least in the case of simple Mendelian diseases. Our expectation is that a complete RFLP map will allow even more powerful analytic methods—methods that will draw on molecular biology, mathematics, clinical medicine, and population genetics.

By no means will complex genetic traits be easy to analyze. Our results merely offer hope that analysis is possible, at least in some cases. Despite all the obstacles, however, it should be remembered that human beings have certain unique advantages for genetics: No other organism is as sensitive to its phenotypes, nor as cooperative about reporting them. Genetic analysis by linkage methods may be able to help in the elucidation of mammalian biology, both normal and abnormal, just as it has in the understanding of biological processes in bacteria, yeast, flies, and mice.

ACKNOWLEDGMENTS

We thank David Page and Helen Donis-Keller for many helpful discussions and for their encouragement. This work was partially supported by grants from the System Development Foundation to E.S.L. and from National Institute of General Medical Sciences to D.B.

REFERENCES

- Anderson, M.A. and J.F. Gusella. 1984. Use of cyclosporin A in establishing Epstein-Barr virus transformed human lymphoblastoid cell lines. *In Vitro* **20**: 856.
- Bateson, W. 1902. *Mendel's principles of heredity*. Cambridge University Press, England.
- Bird, T.D., J. Ott, E.R. Giblett, P.F. Chance, S.M. Sumi, and G.H. Kraft. 1983. Genetic linkage evidence for heterogeneity in Charcot-Marie-Tooth neuropathy. *Ann. Neurol.* **14**: 679.
- Blass, J.P. and G.E. Gibson. 1977. Abnormality of a thiamine-requiring enzyme in patients with Wernicke-Korsakoff syndromes. *N. Engl. J. Med.* **297**: 1367.
- Botstein, D. and H. Donis-Keller. 1984. A molecular approach to defining the inherited components in epilepsy and other diseases of uncertain etiology. *Epilepsia* **25**: S150.
- Botstein, D., R.L. White, M.H. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**: 314.
- Cavalli-Sforza, L. and M.C. King. 1986. Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. *Am. J. Hum. Genet.* **38**: 599.
- Cavenee, W.R., T.P. Dryja, R.A. Phillips, W.F. Benedict, R. Godbout, B.L. Gallie, A.L. Murphree, L.C. Strong, and R.L. White. 1983. Expression of recessive alleles by chromosome mechanisms in retinoblastoma. *Nature* **305**: 779.
- Davies, K.E., P.L. Pearson, P.S. Harper, J.M. Murray, T. O'Brien, M. Sarfrazi, and R. Williamson. 1983. Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acids Res.* **11**: 2303.
- Doose, H. 1982. *Genetic basis of the epilepsies* (ed. V.E. Anderson et al.), p. 2839. Raven Press, New York.
- Dyck, P.J., J. Ott, S.B. Moore, C.J. Swanson, and E.H. Lambert. 1983. Linkage evidence for genetic heterogeneity among kinships with hereditary motor and sensory neuropathy type I. *Mayo Clin. Proc.* **58**: 430.
- Fisher, R.A. 1935. The detection of linkage with dominant abnormalities. *Ann. Eugen.* **6**: 187.
- Garrod, A.E. 1902. A study in chemical individuality. *Lancet* **II**: 1616.
- Gusella, J.F., N.S. Wexler, P.M. Conneally, S.L. Naylor, M.A. Anderson, R.E. Tanzi, P.C. Watkins, K. Ottina, M.R. Wallace, A.Y. Sakaguchi, A.B. Young, I. Shoulson, E. Bonnila, and J.B. Martin. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**: 234.
- Haldane, J.B.S. and C.A.B. Smith. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. *Ann. Eugen.* **14**: 10.
- Jaspers, N.G.J., J. De Wit, M.R. Regulski, and D. Bootsma. 1982. Abnormal regulation of DNA replication and increased lethality in AT cells exposed to carcinogenic agents. *Cancer Res.* **42**: 335.
- Jaspers, N.G.J., R.B. Painter, M.C. Paterson, C. Kidson, and T. Inoue. 1985. Complementation analysis of ataxia-telangiectasia. In *Ataxia-telangiectasia: Genetics, neuropathology, and immunology of a degenerative disease of childhood* (ed. R.A. Gatt and M. Swift), p. 147. A.R. Liss, New York.
- Keijzer, W., N.G.J. Jaspers, P.J. Abrahams, A.M.R. Taylor, C.F. Arlett, H. Takebe, P.D.S. Kimmont, and D. Bootsma. 1979. A seventh complementation group in xeroderma pigmentosum. *Mutat. Res.* **62**: 183.
- Knowlton, R.G., O. Cohen-Haguenaer, N. Van Cong, J. Frezal, V.A. Brown, J.C. Braman, J.W. Schumm, L.C. Tsue, M. Buchwald, and H. Donis-Keller. 1985. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature* **318**: 380.
- Lander, E.S. and D. Botstein. 1986. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proc. Natl. Acad. Sci.* **83**: 7353.
- Macklin, M.T. 1959. Inheritance of retinoblastoma in Ohio. *Arch. Ophthalmol.* **62**: 842.
- McKusick, V.A. 1971. *Mendelian inheritance in man: Cata-*

- logs of autosomal dominant, autosomal recessive, and X-linked phenotypes*. 5th edition. Johns Hopkins University Press, Baltimore.
- Morton, N. 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**: 277.
- . 1956. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J. Hum. Genet.* **8**: 80.
- Norio, R., H.R. Nevanlinna, and J. Perheentupa. 1973. Hereditary diseases in Finland. *Ann. Clin. Res.* **5**: 109.
- Ott, J. 1985. *Analysis of human genetic linkage*. Johns Hopkins University Press, Baltimore.
- Reeder, S.T., M.H. Breuning, K.E. Davies, R.D. Nicolls, A.P. Jarman, D.R. Higgs, P.L. Pearson, and D.J. Weatherall. 1985. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317**: 542.
- Romeo, G., M. Bianco, M. Devoto, P. Menozzi, G. Mastella, A.M. Giunta, C. Micalizzi, M. Antonelli, A. Battistini, F. Santamaria, D. Castello, A. Marianelli, A.G. Marchi, A. Manca, and A. Miano. 1985. Incidence in Italy, genetic heterogeneity, and segregation analysis of cystic fibrosis. *Am. J. Hum. Genet.* **37**: 338.
- Stanbury, J.B., J.B. Wyngaarden, D.S. Fredrickson, J.L. Goldstein, and M.S. Brown. 1983. *The metabolic basis of inherited disease*. McGraw Hill, New York.
- Trevor-Roper, P.D. 1952. Marriage of two complete albinos with normally pigmented offspring. *Br. J. Ophthalmol.* **36**: 107.
- Tsui, L.C., M. Buchwald, D. Barker, J.C. Braman, R.G. Knolton, J.W. Schumm, H. Eiberg, J. Mohr, D. Kennedy, N. Plesvic, M. Zsiga, D. Markiewica, G. Akots, V. Brown, C. Helms, T. Gravius, C. Parker, K. Rediker, and H. Donis-Keller. 1985. Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**: 1054.
- Wainwright, B.J., P.J. Scambler, J. Schmidtke, E.A. Watson, H.Y. Law, M. Farrall, H.J. Cooke, H. Eiberg, and R. Williamson. 1985. Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**: 384.
- White, R., S. Woodward, M. Leppert, P. O'Connell, M. Hoff, J. Herbst, J.M. Lalouel, M. Dean, and G. Vande Woude. 1985. A closely linked genetic marker for cystic fibrosis. *Nature* **318**: 382.
- Willard, H.F., M.H. Skolnick, P.L. Pearson, and J.-L. Mandel. 1985. Report of the committee on human gene mapping by recombinant DNA techniques. *Cytogenet. Cell Genet.* **40**: 360.