# Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis

Roderick H. J. Houwen[1], Siamak Baharloo[2], Kathleen Blankenship[2], Peter Raeymaekers[3], Jenneke Juyn[1], Lodewijk A. Sandkuijl[4] & Nelson B. Freimer[2]

It is now feasible to map disease genes by screening the genome for linkage disequilibrium between the disease and marker alleles. This report presents the first application of this approach for a previously unmapped locus. A gene for benign recurrent intrahepatic cholestasis (*BRIC*) was mapped to chromosome 18 by searching for chromosome segments shared by only three distantly related patients. The screening results were confirmed by identyifying an extended haplotype conserved between the patients. Probability calculations indicate that such segment sharing is unlikely to arise by chance. Searching the genome for segments shared by patients is a powerful empirical method for mapping disease genes. Computer simulations suggest that, in appropriate populations, the approach may be used to localize genes for common diseases.

[1]*Wilhelmina Children's Hospital, Utrecht, The Netherlands*
[2]*Neurogenetics Laboratory and Center for Neurobiology and Psychiatry, Department of Psychiatry, University of California, San Francisco, Box F-0984, San Francisco, California, 94143, USA*
[3]*Center for Human Genetics, University of Leuven, Campus Gasthuisberg, Herestraat 49, B-3000, Leuven, Belgium*
[4]*Institute of Clinical Genetics, Erasmus University, Rotterdam and Department of Human Genetics, Leiden University, The Netherlands*

*Correspondence should be addressed to N.B.F.*

The lod score method of linkage analysis has become a standard tool in human genetics; with highly polymorphic markers it is possible to roughly map genes for simple mendelian disorders with only a single moderately sized family[1]. Applying this approach to common diseases is complicated because one must specify parameters describing the transmission of the trait, which may be unknown for such diseases[2]. Although several robust non-parametric linkage methods have been proposed[3,4], the sample size required for these methods may be so large that collection of adequate numbers of subjects may be difficult even for common diseases[5]. Thus, there is a need for parameter free gene mapping methods that are powerful when relatively few informative individuals are studied. Genetic association studies have been appealing to investigators because they can fulfill both of these criteria. However, until the development of complete genetic maps, it was not practical to screen the genome using association methods, and they were limited to evaluation of small numbers of candidate loci.

The availability of a set of mapped markers now permits genetic association studies to be undertaken across the entire genome, based on linkage disequilibrium (LD) between a disease locus and particular marker alleles or haplotypes. Under conditions of LD, associations are preserved between alleles at tightly linked loci[2]. In LD methods, as in standard linkage analysis, the low rate of recombination between tightly linked disease and marker loci is the basis for mapping disease genes. Both loci are inherited identical by descent (IBD) from a founding ancestor who brought the disease mutation into the population; recombination or random segregation will have reduced the probability of allele sharing except in the region containing the disease mutation. LD approaches are based on the population as the unit of investigation, as opposed to standard linkage methods which focus on families; for example, in lod score analysis, information from all family members is used, while several non-parametric methods focus on information from affected pairs of relatives[3,4]. Identifying LD between disease and marker alleles has been instrumental in positional cloning of disease genes[6,7], but it has been used only after linkage studies had reached the limits of their resolution, that is, no additional recombinants could be detected. In such cases LD was evaluated by identifying conserved haplotypes in patients, rather than by statistical tests of association, as in standard population genetic analyses[8].

LD approaches are most powerful when there is a clear delineation between the marker alleles of affected individuals and those of the general population, that is, when most patients are descended from only one or a few ancestors who transmitted the disease mutation. Therefore, populations that have grown as a result of recent immigration, such as the United States, are less suitable than those in which growth has occurred primarily via reproduction. For example, the Finnish population is characterized by a relatively small number of founders and little recent immigration, and LD mapping, in Finland, has been used to pinpoint the location of several disease genes after they had first been mapped by conventional linkage analysis[9,10].

In 1986, Lander and Botstein proposed that LD mapping could be employed to screen the entire human genome for disease loci, without conventional linkage analyses[11]. This proposed approach has not previously been implemented, however, as it was not feasible until a set of

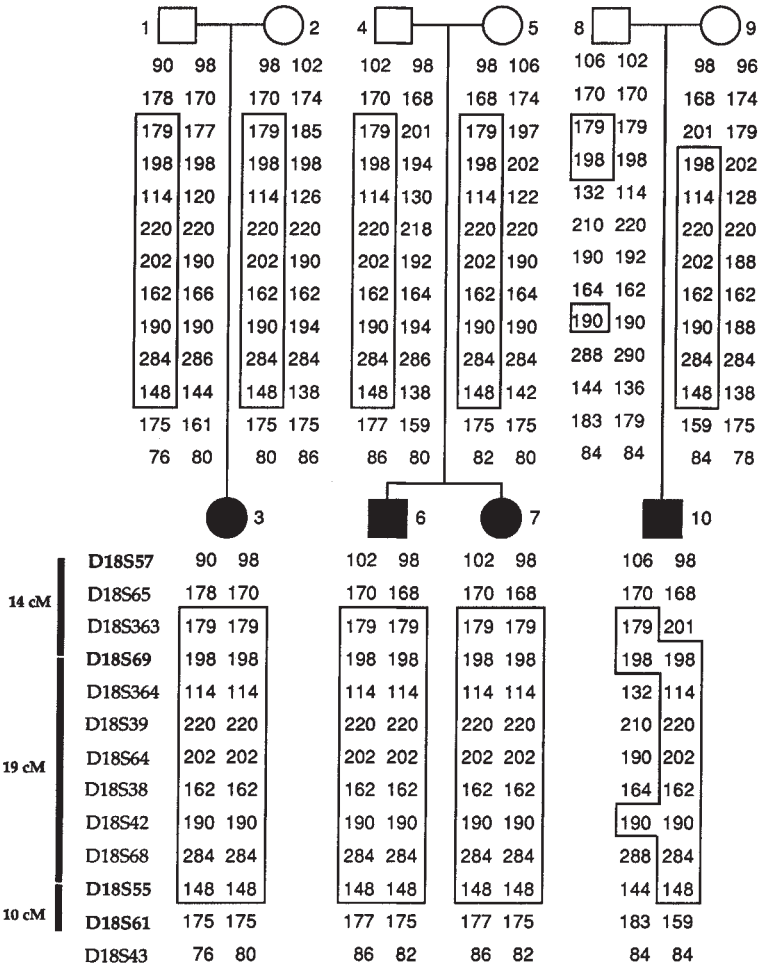


Fig. 1 Haplotypes of *BRIC* patients and their parents using 13 microsatellite markers over the putative candidate region on chromosome 18. The markers from this region that were used in the original genome screening are in bold. The "disease" haplotype is outlined, with two areas possibly shared by six out of six disease chromosomes indicated. The marker distances are approximate and were derived from Genethon[12,16] and the CHLC database.

mapped markers covering the genome became available[12]. We report here the first use of LD for the initial mapping of a human disease gene, having used 256 markers to locate the gene responsible for benign recurrent intrahepatic cholestasis (*BRIC*). We also describe an empirical approach, using reconstruction of conserved haplotypes, to identify regions shared identical-by-descent (IBD) by affected individuals and provide calculations to evaluate the approximate probability that such sharing occurs by chance.

## Benign recurrent intrahepatic cholestasis

We screened the genome for LD by identifying chromosomal segments shared between only three individuals affected with *BRIC*, a rare autosomal recessive disease. The clinical manifestations of *BRIC* include intermittent episodes of cholestasis without extrahepatic bile duct obstruction, with initial elevation of serum bile acids, followed by cholestatic jaundice which generally spontaneously resolves after periods of weeks to months[13–15]. Because the exact relationships between the patients could not be determined, standard lod score analysis was not suitable for mapping the *BRIC* gene. *BRIC* is a good test case for LD mapping for three reasons: (i) It is rare, so there was a reasonable likelihood that most cases in a given population arose from one or a few ancestral mutations; (ii) An isolated study population was available, making a single disease haplotype even more likely; (iii) Affected individuals in this population were distantly separated from each other; we thus hypothesized that these individuals would share few segments across the genome, and that the segment containing the *BRIC* gene could be differentiated from the other segments by typing enough additional markers to detect a conserved haplotype.

## The study population and patient relationships

We identified three *BRIC* patients originating from a somewhat isolated fishing community, of several thousand individuals, in The Netherlands. Although this population is not a geographic isolate, most of its members descend from individuals who lived in the vicinity by the 17th century. Until recently, the majority of marriages were between individuals born in the community. Dramatic population growth occurred in the 19th and early 20th centuries, before exogamous marriages became common. Because endogamy occurred in this population for several generations, it is not surprising that the *BRIC* patients are distantly related; all were born from consanguineous relationships. Nevertheless, no single ancestor could be identified through genealogic studies as the possible source of all six disease chromosomes of the *BRIC* patients. The known relationships between the patients are probably fewer than the unknown relationships which preceded available written records.

To estimate the minimum number of generations between the three patients and their unknown common ancestor, we determined the shortest distance for all known relationships. We were able to establish nine of the fifteen possible relationships (that is, patient A via father to patient A via mother, patient A via father to patient B via father, and so on). For those relationships the connecting ancestor lived, on average, 5.3 generations ago. Assuming that the remaining, unknown connections are at least one generation prior to the earliest known connection (seven generations ago) we concluded that the patients' chromosomes had undergone an average of at least six meioses since their shared origin. Given the dramatic rise in the population of their community beginning about the year 1800, it is likely that 10–12 generations represents an upper bound on their separation from a common ancestor.

## Genotyping results

DNA samples were obtained from the affected individuals, their parents (to enable construction of marker haplotypes) and from an affected sibling of one of the patients (Fig. 1); because these siblings should share 50% of their genomes IBD, only one of them was used in searching for shared segments. However, we genotyped both individuals because regions that they do not share IBD would be extremely unlikely to contain the *BRIC* gene. On these samples, 256 microsatellite markers were tested, spaced at

Table 1 Marker segments shared by at least three of six BRIC disease chromosomes

| Marker segment | Length (cM) | Number shared | Resolution of haplotype |
|---|---|---|---|
| *D1S167–D1S248* | 19 | 3/6 | none |
| *D1S248–D1S187* | 07 | 3/6 | none |
| *APOA2–D1S194* | 05 | 3/6 | none |
| *D4S403–D4S418* | 19 | 3/6 | none |
| *D4S396–D4S428* | 04 | 3/6 | none |
| *D5S406–D5S432* | 11 | 3/6 | none |
| *D10S464–D10S109* | 20 | 3/6 | none |
| *D10S109–D10S185* | 18 | 3/6 | none |
| *D17S806–D17S790* | 08 | 3/6 | none |
| *D18S57–D18S69* | 14 | 3/6 | none |
| *D18S69–D18S55* | 19 | 5/6 | shared |
| *D18S55–D18S61* | 10 | 3/6 | shared |
| *D19S216–D19S225* | 40 | 3/6 | none |
| *D21S268–D21S266* | 08 | 3/6 | none |

Distances are based on the 1993-1994 Généthon map or CHLC database (Ver. 2). A total of 60 additional microsatellite markers were tested to verify the possible haplotypes in the regions containing these segments.

about 10–20 centiMorgan (cM) intervals across the 22 autosomes[12]. These markers provided 235 inter-marker segments which cover approximately 3,260 cM, or an estimated 90% of the autosomal genome[16]. For each pair of adjacent markers the haplotypes transmitted to patients were constructed; for 205 segments all six transmitted haplotypes could be scored, while for the remaining segments one of the flanking markers could only be read for two of the patients. For the segments with six scored haplotypes, the average number of unique haplotypes was 5.50, remarkably close to the maximum of 6.0 expected for completely informative markers. No region was found for which all six transmitted haplotypes were identical, but one area bounded by *D18S69* and *D18S55* (located in chromosome 18q21–18q22 and covering approximately 19 cM) showed sharing by five out of six haplotypes. No additional segments showed sharing by more than three haplotypes. Thirteen segments, dispersed on eight chromosomes, displayed sharing of three of the six haplotypes (Table 1).

The observed sharing in these 14 segments could be due to inheritance of the segment from a common ancestor (that is IBD) or, if the segment was not inherited from a single ancestor, to a high frequency of the alleles in the relatively isolated study population, that is the segments are "shared" identical-by-state (IBS) rather than IBD. We anticipated that at least some IBS segment sharing would be observed, and therefore, we typed additional markers in each of the shared segments (a total of 60 markers). In most of the tested segments, haplotypes that had appeared identical, using the original markers, were observed to diverge when the results of the new marker typings were added. The exceptions were two adjacent regions on chromosome 18, including the one that was shared by five out of the six chromosomes (Table 1). A complete haplotype for the shared chromosome 18 region is shown in Fig. 1.

Assuming that the genetic map is approximately correct, we have screened about 90% of the genome and have not found other regions shared IBD by the affected individuals. This coverage further supports the mapping of *BRIC* to chromosome 18. After the second stage of marker typing (to detect haplotypes), there were no unscreened regions larger than about 20 cM; thus, if undetected IBD segments exist in these regions they must be shorter than the IBD region on chromosome 18.

Of the markers tested so far in the shared segment on 18q, the sixth disease chromosome has the same sized allele as the other five chromosomes only for *D18S42* and *D18S69*. This chromosome may carry a mutation of a different origin, or, alternatively, may provide crucial information that could narrow down the *BRIC* candidate region.

### Probability calculations

The identification of a haplotype, covering about 20 cM, shared by five out of six chromosomes in distantly related patients provides strong empirical evidence for the mapping of the *BRIC* locus. Further proof is provided by calculations that estimate the probability of patients' sharing a given chromosomal segment IBD by chance rather than because of the presence of a disease gene, that is, of a "false positive" finding. Calculations (see Methodology) were performed to estimate the maximal probability of random sharing in the *BRIC* analysis, with three patients separated by six generations. The approximate probability of at least one of the 235 tested segments being shared IBD on three or more of the six patients' chromosomes is only 0.011. Similarly, the probability to find five or more of the six chromosomes sharing a segment IBD by chance is less than 0.0000005 ($5\times 10^{-7}$). For both cases, corrections were performed for multiple tests.

### Applicability of a search for shared segments

An important question for other mapping studies in isolated populations is whether the observed haplotype sharing for *BRIC* was merely fortuitous or whether one might readily expect to see such significant results. We carried out computer simulations to assess the power of a search for shared segments in a situation comparable to the *BRIC* analysis: six chromosomes in patients, six generations removed from a common ancestor, assuming homogeneity in the origin of the disease chromosomes as well as of the disease locus. In this scenario, the probability of finding IBD sharing on three or more of six chromosomes is more than 98% with a 6 cM map, close to 90% with a 10 cM map, and almost 70% with a 14 cM map (Fig. 2).

For many inherited diseases, complete homogeneity is unlikely, even in an isolated population. Accordingly, we investigated the power of a search for shared segments under 50% heterogeneity (Fig. 2). In these simulations, 'heterogeneity' merely indicates that not all of the individuals are affected due to inheritance of a single genotype from a common ancestor. That is, heterogeneity may be due to two or more different mutations at the same disease locus (allelic heterogeneity), existence of another locus elsewhere in the genome (locus heterogeneity), or existence of phenocopies (aetiologic heterogeneity), or any combination of these three. As expected, in this scenario power to detect IBD sharing is reduced, compared to the simulations assuming homogeneity. However, the loss of power under heterogeneity could be compensated for by investigating more patients or by testing a denser map of markers; for instance, the probability is more than 93% that a segment shared by four out of 16 chromosomes can be detected using a 6 cM marker map.

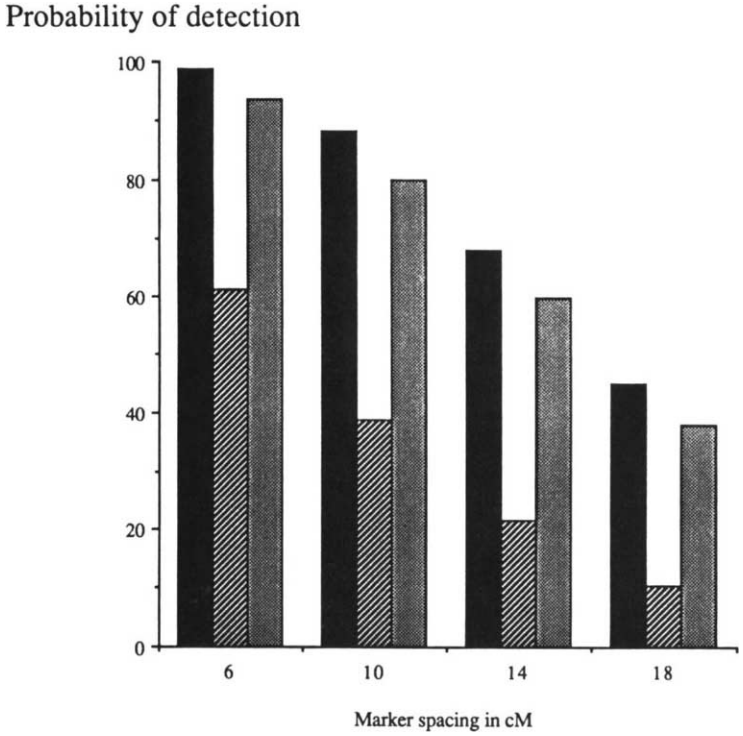


Fig. 2 The results of computer simulations showing the probability of detecting shared segments under three different situations, with varying densities of marker maps. The first bar in each series shows predictions for a situation similar to that of BRIC (searching for segments shared by at least three out of six chromosomes, under homogeneity). The other bars show the effect of increasing sample size on the probability to detect segment sharing under conditions of moderate heterogeneity; sharing is by at least four chromosomes out of a total of either ten or 16, with 50% heterogeneity.
■, homogeneity, 3/6; ▨, heterogeneity, 4/10; ■, heterogeneity, 4/16.

## Discussion

In mapping *BRIC*, we report the first use of LD to identify the chromosomal location of a human disease gene through whole genome screening. Our approach is primarily empirical; the results were obtained by searching for chromosomal segments that are shared by patients, and then demonstrating that sharing is IBD. The initial screening steps led to finding 14 segments that are shared IBS, including contiguous segments on chromosome 18. The proof that any of the segments is shared IBD rests on construction of haplotypes conserved over several polymorphic markers in this region. The *BRIC* haplotype on chromosome 18 consists of eight markers for which at least five out of six chromosomes demonstrate sharing; with a haplotype of eight markers over more than 20 cM, even if the shared alleles are frequent in the study population, it is extremely unlikely that they are shared by chance, rather than by descent. The fact that the non-transmitted chromosomes in this region do not contain the haplotype shared by patients is further evidence that IBD sharing here is restricted to the disease chromosomes. For example, for *D18S364* the six non-transmitted chromosomes bear six different alleles.

Our results demonstrate that a search for shared segments can be used to map a disease gene, with a very small sample of patients, for a rare trait with a simple inheritance pattern. Because we genotyped just ten individuals, over 250 markers were screened with very few experiments. This sample (with a total of four patients) is too small to attain a significant finding via regular linkage analysis (that is, a lod score of three or more). Consideration of all known relationships in calculating lod scores can enhance the informativeness of consanguineous families for linkage analysis, but leads to impractically long calculation times. The intractability of the lod score method is particularly evident when multiple markers are analysed simultaneously. In contrast, the evaluation of shared segments derives its strength from the information provided by simultaneous consideration of several adjacent markers.

The search for shared segments is similar to other LD methods in that it focuses on a sample of patients rather than on individual families or pairs of relatives. However, standard statistical tests for association implicitly assume that affected individuals are virtually unrelated, and therefore ignore the probability of chance IBD sharing of alleles. Complete genome screening for LD has generally been considered impractical, based on the assumption that it would require typing thousands of markers at very narrow intervals[17]. This assumption does not apply in isolated populations which have formed within the past several centuries; as demonstrated for *BRIC*, in such populations only a few hundred markers may be needed to localize a disease gene. Detecting shared segments among patients provides evidence for localization of a disease gene, not because they display marker allele frequencies that are significantly different from those in the general population, but because the conservation of haplotypes over large chromosomal regions indicates IBD, regardless of the population.

In recently isolated populations it is likely that patients will be distantly related; this is the basis for finding shared segments. However, the existence of such relationships could also lead to false positive findings. As shown here, by estimating the closest possible degree of relationship between affected individuals, it is feasible to calculate an approximate probability that observed segment sharing would occur by chance. For this case, even the conservative probability that we calculated is sufficiently low to provide support for the localization of *BRIC*. With less strong empirical evidence it may be useful to develop more precise probability calculations. Methods for calculating such exact probabilities have previously been proposed, although only for evaluating random sharing of two chromosomes[18]. Also, for isolated populations in which extensive genealogical data are available, it may be useful to quantify the extent of consanguinity between affected individuals in calculating exact probabilities.

LD approaches are distinct from other methods that have recently been proposed for mapping rare mendelian traits, notably homozygosity mapping (HM) [19] (see Box). HM has been used to map loci in selected candidate regions[20,21], but its efficiency for full genome screening is still unknown. By definition, HM is focused on identifying homozygosity, that is for mapping recessive diseases. In contrast, the search for shared segments treats each chromosome separately, and thus is feasible with dominant

## Comparison of gene mapping methods based on IBD

Several strategies have been developed to map disease genes via linkage, that is by demonstrating a small recombination frequency and thus close proximity between disease and marker loci. These methods are all based on identifying, in a set of affected individuals, marker alleles or haplotypes inherited **identical by descent** (**IBD**) from a common ancestor who transmitted the disease gene. If we compare two such patients at a given marker, their alleles would not merely be the same size in base pairs (**identical by state,** or **IBS**), but can be shown, statistically or empirically, almost certainly to have been inherited without change from the common ancestor. The various linkage methods mainly differ in the number of meioses separating this ancestor and the members of the study sample (Fig. 3). These differences determine the composition of the ideal sample for each method, and also the type of analysis used to recognize linkage.

The differences between linkage methods can be graphically depicted (Fig. 3) by considering the change in size, over time, of chromosome segments inherited IBD. The X axis shows the median length of a segment containing a particular disease gene, shared IBD by two affected individuals. The Y axis shows the number of meiotic steps separating the individuals (via the common ancestor).

Individuals separated by two meiotic events share very large segments IBD. This corresponds to the situation of **affected sib-pair** analysis. Because sibs share so much of the genome IBD, and the common ancestor (a transmitting parent) gives rise to only one set of affected sibs, a sizable sample is required; only by identifying a significant deviation from expected segregation patterns in allele or haplotype sharing, can the region containing the disease gene be differentiated from the numerous other regions inherited IBD.

Extended pedigrees are most often analysed using the **lod score method**, in which marker data are used to evaluate the likelihood of linkage in comparison to the likelihood of no linkage. Here, closely related affected individuals are tested, usually separated by between two to six meiotic steps, and a single individual may produce enough descendants to permit rough localization of a disease gene. However, as the IBD fragments are still quite large after six meiotic steps, it is unlikely that single pedigrees will be useful for fine mapping; the few recombinations rarely occur exactly where one wishes.

**Homozygosity mapping** is based on identification of regions of homozygosity in patients born from consanguinous matings. In this case the common ancestor carried a single copy of the disease chromosome which was transmitted to the patient via both parents. When the patient is the product of a first or second cousin marriage, which is not uncommon in some populations, only around the disease gene will there likely be IBD segments of sufficient size to be detected using currently available marker maps.

**Linkage disequilibrium** (LD) around a disease gene signifies that a given marker allele (or haplotypes of several marker loci) is over-represented on chromosomes containing the disease mutation. LD usually arises due to founder effects, occurring when one or a few individuals bring a chromosome containing a disease mutation into a small population; if the population then expands in isolation from further immigration, a high proportion of patients will carry this chromosome. Over time, recombination ensures that progressively smaller fragments of the founder chromosome remain intact. Therefore, populations which were founded and expanded in the distant past (such as 100 generations ago) are useful for pinpointing the exact location of disease genes, as one expects that fragments inherited IBD will be less than 1 cM in length. In such populations, between 1,000 and 2,000 markers would probably be needed to identify the IBD regions through a complete genome screen. With fewer markers it is extremely unlikely that associations would be detected. However, utilizing more recently founded populations, such as the one described here, it is expected that IBD segments of between five and 20 cM would be conserved around a disease gene; therefore, a **search for shared segments** is practical with only a few hundred markers.

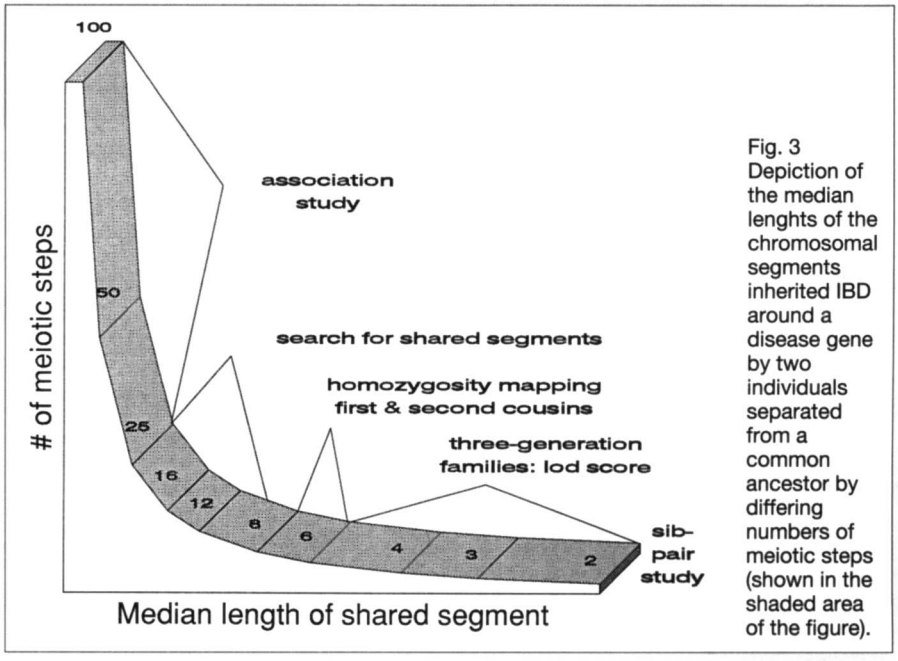


Fig. 3 Depiction of the median lenghts of the chromosomal segments inherited IBD around a disease gene by two individuals separated from a common ancestor by differing numbers of meiotic steps (shown in the shaded area of the figure).

as well as recessive disorders; for example, if *BRIC* were an autosomal dominant disorder, with three patients we would expect to detect, at most, sharing of three out of six chromosomes. Thus, for a simple dominant trait one needs a sample size which is about double what would be needed for a recessive trait in an equivalent population.

Exclusion of false positive regions using lod score analysis may be difficult because, within a family, when one identifies linkage with a given marker, there will likely be some evidence of linkage at adjacent marker loci. By contrast, in LD analyses, affected individuals in most cases will not share the same alleles at adjacent loci, in the absence of true linkage, unless the markers are in strong LD in the population as a whole. As demonstrated for *BRIC*, segments that initially appear identical can be distinguished from those that are truly IBD by typing more markers in the region and evaluating the presence or absence of shared haplotypes. In addition, in contrast to lod score analyses, the calculation used to evaluate the probability of false positive findings for *BRIC* was free of assumptions regarding genetic parameters.

All methods of gene mapping are potentially impeded by aetiological heterogeneity. Even in carefully chosen populations, the homogeneity evident for *BRIC* may not apply for a particular disease. For example, for the relatively uncommon autosomal recessive form of retinitis pigmentosa, genetic locus heterogeneity has recently been identified within a single isolate, in the same part of The Netherlands as the community from which the *BRIC* patients originated[22]. It is thus reassuring that, as evidenced by our power calculations, the search for shared segments provides a high probability of detecting candidate regions even under a moderate degree of aetiologic heterogeneity. The improvement in power achieved by increasing the number of individuals tested suggests that this approach may be feasible for studying common diseases. The possibility of identifying shared segments in patients with common diseases is supported by a recent report of a conserved haplotype for hereditary non-polyposis colon cancer (HNPCC), one of the most frequent inherited cancer syndromes; in this case, a 10 cM segment on chromosome 3 was shared by affected individuals in seven of 18 families investigated in Finland[23], and an identical germline mutation in the causative gene was subsequently discovered in each of these patients. Locus heterogeneity has been proven for HNPCC by the cloning of two responsible genes, one on chromosome 2 and the other on chromosome 3(ref. 24).

Despite its demonstrated utility for simple disorders and potential application for complex ones, an important caveat must be considered before implementing a search for shared segments. Because the assumption is made that the majority of affected chromosomes are descended from one or a few transmitting founders, highly cosmopolitan populations, such as those of North American or most European countries are probably not well suited to this approach, as the underlying presumption may be violated even for rare mendelian disorders. However, it is likely that, within these countries, populations drawn from certain rural regions will be equivalent to our Dutch sample used to map *BRIC*. For mapping common diseases, samples drawn from relatively homogenous national populations would be most amenable to using a search for shared segments; examples of such populations include those of Finland, Sardinia, and Costa Rica.

As originally proposed[11], LD mapping would have required a more detailed genetic map than the one used to map *BRIC*, because it was envisaged that the relationships between the affected individuals would be more distant than in our study sample; that is, segments inherited IBD surrounding the disease gene would be shorter, on average, than the roughly 20 cM observed around the *BRIC* gene. Although detecting IBD requires a larger number of markers in a population of more distantly related patients, there would also be less probability of detecting segments that are IBD by chance. As genetic marker maps become increasingly dense, it will be possible to tailor sampling strategies for specific disorders and populations.

## Methodology

**Study population.** Church and civil records were reviewed to establish genealogies of the affected individuals. All relationships are known back to the grandparents; beyond that point there is some missing information. A complete description of the study population and the clinical findings is given in a manuscript submitted elsewhere by Th. de Koning *et al.*

**Marker genotyping.** Généthon markers[11,16] were supplemented by markers from the Cooperative Human Linkage Center (CHLC) public database[25]. Genotyping procedures were previously described[26]. Marker order on chromosome 18 (Fig. 1) was derived from the CHLC database (v2), from Généthon, and from the *BRIC* data. Marker distances were mostly derived from the Généthon map. For others, the position relative to flanking Généthon markers was obtained from the CHLC database.

**Constructing haplotypes.** For each pair of adjacent markers, phase on the patients' chromosomes was constructed by tracing the parental origin of the alleles. The number of occurrences in the three patients was counted for each haplotype. Whenever phase could not be reconstructed with certainty (that is, the two parents and the child were all identically heterozygous at one or both markers) the various possible haplotype configurations contributed 0.5 to the haplotype count. The number of unique haplotypes was determined by summing all haplotype counts, with haplotypes with more than one occurrence contributing only as one occurrence.

**Statistical analyses.** *The probability of detecting a random region which is IBD.* First, the approximate probability was calculated for a random chromosomal segment (not containing a disease gene), bounded by two markers x cM apart, to be identical to a region of a particular ancestral chromosome g generations ago:

$$p_1(x)=0.5^g \cdot (1-0.01x)^g \qquad (1)$$

Here, the quantity 1–0.01x represents the probability of observing no crossovers in an interval with recombination fraction of 0.01x Morgan (= x cM) in a given meiosis. The factor 0.5 relates to the probability that a given chromosome of a parent is transmitted to a child. In the case of *BRIC*, we assumed that g=6. Next, the probability was calculated that k out of n individuals, connected via a common ancestor g generations ago, inherit a segment that is identical without interruption to a given chromosome of that ancestor:

$$p_2(k, x)=\binom{n}{k} \cdot (p_1(x))^k \cdot (1-p_1(x))^{n-k} \qquad (2)$$

For *BRIC*, n=6 and the calculations were performed for k=3 and k=5. When evaluating the possibility of a false positive IBD finding, that is, sharing of a random area in the genome by k or more chromosomes, one is not concerned about the possibility that patients have inherited an identical segment of a given chromosome of a common ancestor, but rather that several patients have inherited an identical segment of any of the two chromosomes of that ancestor. In fact, in studies of human disease, distant relatives are usually related via an ancestral couple rather than via a single ancestor. To account for the corresponding four ancestral chromosomes each providing an opportunity for sharing by chance, the total probability for a false positive IBD finding was approximated as follows:

$$p_3(k,x)=4 \cdot \sum_{i=k}^{n} p_2(i,x) \qquad (3)$$

The summation in this equation is related to the fact that we are interested in the probability that k *or more than k* chromosomes share a segment by chance.

The probabilities mentioned thus far relate to sharing by chance of a particular region. When evaluating the overall probability for a false positive finding anywhere in the genome, we summed the values $p_3$ over all m intervals tested to obtain a final probability:

$$p(k,m)=\sum_{j=1}^{m} p_{3(k,x_j)} \qquad (4)$$

Here, m=235 and the size of the intervals ($x_m$) was determined from the Généthon and CHLC maps. These calculations are conservative in that the probability of a false positive finding in one interval is not independent of the probability of a false positive finding in the adjacent intervals. We are currently extending the above equations to correct for this non-independence.

*Power to detect disease loci.* For approximate power calculations, computer simulations were carried out with the MEIOSIM computer program[27], which applies a simple Monte Carlo approach. In this case, a map of 30 markers which were completely informative (with no IBS sharing allowed) and equidistant (at 2 cM), was followed

through six meioses. In one series of 10,000 simulations, a disease gene was included in the middle interval (between markers 15 and 16, at 1 cM from each), and it was required that the same disease mutation be present at that location before and after the six meioses. This simulation represents the situation where a carrier ancestor transmits a disease gene to a carrier/affected descendant. By extracting from these simulations only information on four of the markers, expected results from a map of markers at 6 or 10 or 14 or 18 cM were studied. In each simulation, we counted how frequently the marker alleles and the segments at two adjacent markers in the descendant were identical to those initially present on the ancestral chromosome. The following example illustrates this: from a simulated descendant chromosome, alleles for markers at 3 cM and 9 cM on either side of the disease gene were checked for identity to the ancestral alleles (—M11-(6 cM)–M14–(3 cM)–disease gene–(3 cM)–M17–(6 cM)–M20—, that is, a map of markers at 6 cM intervals). For each of the three segments (M11–M14, M14–M17, M17–M20), the probability of sharing was high due to the presence of the disease gene in the middle interval on both the ancestral and the descendant chromosome: interval M11–M14 was shared in 5,860 of 10,000 simulations, interval M14–M17 in 6,921, and M17–M20 in 5,841 simulations. The frequency of all eight possible combinations of sharing/non-sharing for the three intervals was counted. For the calculation of the power of testing six disease chromosomes and searching for segments shared by three out of six chromosomes, all possible combinations of the eight patterns of sharing/non-sharing of the three intervals on the six chromosomes were listed. For each of those combinations it was checked whether sharing occurred on three or more chromosomes for one or more of the segments. If this was the case, the probability of that combination was computed based on the frequencies of the eight different patterns as obtained in the simulations. The sum of probabilities for all combinations that met the sharing criteria was taken as an approximation of the power of the test.

In a second series of 10,000 simulations, the ancestral and descendant chromosomes did not carry a disease gene, and therefore there was only a small probability that any segment of the ancestral chromosome was maintained in the descendant ('unlinked' situation). Sharing of the segments was now infrequent (M11–M14 and M14–M17: 117 in 10,000; M17–M20: 120 in 10,000), and the observed frequencies were close to the approximate false positive frequency (0.011) calculated in equation 4 for sharing of three or more out of six haplotypes. These frequencies were combined with the frequencies obtained in the linked simulations, to reflect frequencies for a sample of patients with 50% heterogeneity. Now, all combinations for samples of 10 and 16 chromosomes were taken, and all sharing by four or more chromosomes of one or more segments was scored.

**Acknowledgements**

*We thank S. Blower and S. Service for useful suggestions on the statistical methods and on the manuscript and J. Gitschier for comments on the manuscript. This work was supported by grants from NIH (MH00916, MH49499 and MH47563) and an award from the National Alliance for Research on Schizophrenia and Depression (N.B.F.) J.J. was supported by the Netherlands Digestive Disease Foundation and P.R. by a post-doctoral award from the National Fund for Scientific Research, Belgium. S. Barondes, PI for grant MH47566.*

Received 21 March; accepted 19 October 1994.

1. Heutink, P. *et al.* A gene subject to genomic imprinting and responsible for hereditary paragangliomas maps to chromosome 11q23-qter. *Hum. molec. Genet.* **1**, 7–10 (1992).
2. Ott, J. *Analysis of human genetic linkage* 2nd edn (Johns Hopkins University Press, Baltimore, 1991).
3. Elston, R.C. Sib pair screening tests for linkage. *Genet. Epidemiol.* **1**, 315–326 (1984).
4. Weeks, D.E. & Lange, K. The affected-pedigree-member method of linkage analysis. *Am. J. hum. Genet.* **42**, 315–326 (1988).
5. Cook, J. *et al.* Availability of type II diabetic families for detection of diabetes susceptibility genes. *Diabetes* **42**, 1536–1543 (1993).
6. Kerem, B-S. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080 (1989).
7. MacDonald, M. *et al.* The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genet.* **1**, 99–103 (1992).
8. Lewontin, R.C. On measures of gametic disequilibrium. *Genetics* **120**, 849–852 (1988).
9. Hastbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992).
10. Lehesjoki, A.E. *et al.* Localization of the *EPM1* gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum. molec. Genet.* **2**, 1229–1234 (1993).
11. Lander, E.S. & Botstein, D. Mapping complex genetic traits in humans: new methods using a complete RFLP linkage map. *Cold Spring Harbor Symp. Quant. Biol.* **51**, 49–52 (1986).
12. Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
13. Summerskill, W.H.J. & Walshe, J.M. Benign recurrent intrahepatic obstructive jaundice. *Lancet* **ii**, 686–690 (1959).
14. Shapiro, R.H. & Isselbacher, K.J. Benign recurrent intrahepatic cholestasis. *New Engl. J. Med.* **268**, 708–711 (1963).
15. Brenard, R., Geubel, A.P. & Benhamou, J.P. Benign recurrent intrahepatic cholestasis: A report of 26 cases. *J. clin. Gastroenterol.* **11**, 546 (1989).
16. Gyapay G. *et al.*, The 1993–1994 Généthon human genetic linkage map. *Nature Genet.* **7**, 246–339 (1994).
17. Plomin, R., Owen, J.O. & McGuffin, P. The genetic basis of complex human behaviors. *Science* **264**, 1733–1739 (1994).
18. Donnelly, K. The probability that related individuals share some section of genome identical by descent. *Theor. Pop. Biol.* **23**, 34–63 (1983).
19. Lander, E.S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567–1570, (1987).
20. Ben Hamida, C. *et al.* Localization of Friedreich ataxia phenotype with selective vitamin E deficiency to chromosome 8q by homozygosity mapping. *Nature Genet.* **5**, 195–200 (1993).
21. Pollak, M.R. *et al.* Homozygosity mapping of the gene for alcaptonuria to chromosome 3q2. *Nature Genet.* **5**, 201–204 (1993).
22. van Soest, S. *et al.* Assignment of a gene for autosomal recessive retinitis pigmentosa (RP12) to chromosome 1q31–1q32.1 in an inbred and genetically heterogenous disease population. *Genomics* **22**, 499–504 (1994).
23. Nystrom-Lahti, M. *et al.* Close linkage to chromosom 3p and conservation of ancestral founding haplotype in hereditary nonpolposis colorectal cancer families. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6054–6058 (1994).
24. Bodmer, W., Bishop, T., & Karran, P. Genetic steps in colorectal cancer. *Nature Genet.* **6**, 217–219 (1994).
25. Buetow, K. *et al.* High resolution genetic maps incorporating multiple classes of short tandem repeat polymorphisms. *Am. J. hum. Genet.* **53**, A176 (1993).
26. DiRienzo, A. *et al.* Mutational processes of simple-sequence repeat loci in human populations. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3166–3170 (1994).
27. Sandkuyl, L. & Ott, J. Determining informativity of marker typing for genetic counseling in a pedigree. *Hum. Genet.* **82**, 159–62 (1989).