Linkage Analysis and Long QT Syndrome
Using Genetics to Study Cardiovascular Disease

Mark Keating, MD

**Background.** Recombinant DNA technologies have facilitated the development of a set of polymorphic DNA markers covering the human genome. General linkage analysis in families predisposed to inherited disease is now feasible. Linkage analysis can help identify a disease gene even when relatively little is known about the disorder.

**Methods and Results.** Using this approach, we have identified linkage between a gene that causes the long QT syndrome and DNA markers on chromosome 11.

**Conclusions.** The identification of the chromosomal location of the long QT locus is the first step in defining the specific mutations that cause this disease. (*Circulation* 1992;85:1973–1986)

**KEY WORDS** • DNA markers • restriction fragment length polymorphisms • recombination, genetic • arrhythmias

To discover genes that play an important role in cardiovascular disease, investigators have studied proteins that might be important in a disorder. If biochemical or physiological abnormalities were found, these investigators could work back to the gene that encoded the protein, thereby identifying the genetic mutations that caused the disorder. This approach has helped to advance our understanding of cardiovascular biology and disease. In particular, investigators have used this approach to make great progress in our understanding of abnormalities in lipoprotein and lipid metabolism. Recently, investigators have taken a different and complementary approach to these problems, one that involves molecular genetics. Often referred to as “reverse genetics” or “positional cloning,” this strategy begins with the chromosomal localization of a gene that is responsible for an inherited disorder. (Neither of these terms accurately describes the actual process of mapping a disease gene and ultimately identifying disease-causing mutations. The term “reverse genetics” does not reflect the process but rather the change in experimental direction from starting with proteins to beginning with genes. The term “positional cloning” is more accurate in that it implies the cloning of a gene on the basis of its chromosomal location. Merely cloning a gene located near a disease locus, however, does not prove that it is the disease gene. We prefer to use the term “mapping and mutational analysis” to describe the process of disease gene identification outlined here.) The ultimate goal of the approach is to define abnormalities in the protein product of the gene and explain the disease pathogenesis. The purpose of this article is to review the theory and practice of the first step in this genetic strategy, linkage analysis of families with inherited disease.

**Genetic Basis of Linkage Analysis**

In the late 19th century, Gregor Mendel demonstrated that traits were inherited as independent units; that is, the inheritance of one trait did not influence the likelihood of inheriting a second.2 This fundamental rule of inheritance is known as Mendel’s law of independent assortment. In the early 1900’s, Thomas Hunt Morgan and his coworkers proposed the existence of an exception to this law. In breeding experiments performed with *Drosophila*, these investigators noted that certain traits, or genes, were inherited together (coinherited) more frequently than would be predicted by chance.3,4 These coinherited genes were said to be linked. It later became apparent that genes were coinherited or linked because they were physically located on the same chromosome.5 Linkage analysis is a technique that can be used to identify genes that are coinherited and therefore are located on the same chromosomal segment.

Human cells contain 23 pairs of homologous chromosomes, for a total of 46 diploid chromosomes. An individual inherits one set of chromosomes from his mother and the other set from his father. Each chromosome consists of a linear array of double-stranded deoxyribonucleic acid (DNA); the entire genome consists of approximately 3 billion base pairs of DNA. Some of this DNA is organized as genes (approximately 100,000 of them) that encode proteins, but most of the DNA has no known function. Each chromosome contains a different complement of genes and intervening sequences. For example, the genes that encode the human lymphocyte antigens (HLAs) are always located

From the Division of Cardiology, Department of Human Genetics and Eccles Program in Human Molecular Biology and Genetics, University of Utah Health Sciences Center, Salt Lake City, Utah.

Supported by the National Institutes of Health, an American Heart Association Established Investigator Award, and a Syntex Scholars Award.

Address for correspondence: Mark Keating, Eccles Institute for Human Genetics, Room 2100, University of Utah, Salt Lake City, UT 84112.

Received January 9, 1992; revision accepted March 11, 1992.
on the short arm of chromosome 6 (chromosome 6p), whereas the gene for apolipoprotein B is always located on the short arm of chromosome 2 (chromosome 2p). The order of genes in the human genome is invariant, and the relative position of one gene is said to be its locus. Homologous chromosomes are similar but not identical. Each cell contains two copies, or alleles, of the apolipoprotein B gene, each located on its respective homologue of chromosome 2. DNA sequences that separate genes, however, may vary considerably, and even allelic forms of the same gene contain subtle DNA sequence differences. The DNA sequence differences between homologous chromosomes are what make linkage analysis possible.

A well-known example of linkage in human families is sex or X-linked inheritance (Figure 1). An inherited disease that affects only sons of apparently unaffected mothers is caused by the inheritance of a recessive gene located on the sex-determining X chromosome. Sex linkage is relatively easy to identify as a mode of inheritance because of an obvious variation in the human genome; males inherit one X chromosome (from their mothers), whereas females inherit two (one from each parent). A mother with the recessive gene carries another, presumably normal, copy of that gene on her second X chromosome, so she would be phenotypically normal. Her sons, however, have a 50% likelihood of inheriting the X chromosome containing the aberrant gene and manifesting the disease phenotype.

Because the identification of X linkage was relatively simple, much of the early progress using linkage analysis involved genes located on the sex chromosomes. An example of an X-linked gene that affects the cardiovascular system is dystrophin, the gene responsible for both Duchenne and Becker muscular dystrophies. Most inherited disorders, however, are not linked to the X chromosomes but are caused by genes located on one of the 22 pairs of homologous chromosomes (the autosomes). To identify linkage with genes located on the autosomes, it is necessary to identify genetic variations that are specific for those autosomes. Identifiable genetic variations are called polymorphic markers.

### DNA-Based Markers

The ideal genetic marker would be easily detectable and would exist in a number of different, or polymorphic, forms. Linkage between a marker and a disease gene is detectable only if the individual with the disease gene carries two different forms of that marker. Like genes, a marker has a specific chromosomal location or locus, and different forms of the same marker are also referred to as alleles. Thus, allelic forms of the same marker share the same chromosomal location. When an individual carries two different alleles of one marker locus, he is said to be heterozygous for that marker. If the alleles are identical, he is said to be homozygous.

Until recently, only a limited number of markers were available. Most of these were protein polymorphisms like blood group antigens or HLA. Only about 30 of these protein polymorphisms were useful, and these represented only a few chromosomes and an even smaller portion of the human genome. For lack of genetic markers, most human inherited disorders could not be approached by use of linkage analysis.

The situation improved dramatically with the advent of recombinant DNA technology in the 1970s. In 1978, the University of Utah sponsored a genetics retreat at a ski resort in the Wasatch mountains. At that meeting, David Botstein, Ronald Davis, and Mark Skolnick first suggested that variations in DNA sequence, which were known to exist in other organisms, might also exist in humans. They further suggested that these variations could be detected by use of the tools of molecular biology and that they could be used as polymorphic markers in linkage studies. One year later, Kan and Dozy demonstrated the existence of DNA sequence variation at the β-globin locus associated with sickle cell anemia. Shortly thereafter, DNA variation was also observed at the same locus in association with thalassemias. Thus, the β-globin gene was the first DNA-based polymorphic marker. In 1980, Ray White and his colleagues identified the first highly polymorphic locus in an arbitrary, noncoding stretch of DNA (Wyman and White).

### Restriction Fragment Length Polymorphisms

The new strategy has proved to be very powerful. Human DNA sequences contain a high level of variation; for example, differences in sequence exist between homologous chromosomes at every 200–300 base pairs. The initial problem was how to detect them. One way to detect sequence polymorphisms is with restriction enzymes. A restriction enzyme, produced by and purified from certain bacteria, recognizes a specific sequence in a linear piece of double-stranded DNA called a restriction site, where it leaves the DNA. Variations in DNA sequence that create or eliminate a restriction site will result in a change in the length of the resultant restriction fragment (Figure 2A). These sequence variations became known as restriction fragment length polymorphisms (RFLPs).
To detect an RFLP, the restriction fragments are first separated by agarose gel electrophoresis (Figure 3). DNA molecules are drawn through a gel by an electric field and separated on the basis of size. Over the same period, smaller fragments travel farther than larger ones. These double-stranded DNA fragments are then denatured (converted from the double-stranded to the single-stranded form), and the resultant single-stranded molecules are transferred to a nylon or a nitrocellulose membrane, a technique known as Southern blotting. Once denatured, these single-stranded pieces of DNA will bind to (hybridize with) a complementary strand of DNA with great specificity.

Human DNA contains multiple copies of any given restriction site, so the filter generated by Southern blotting will contain multiple DNA fragments from all 46 chromosomes. Most of these DNA fragments will not be polymorphic. To identify the RFLP of interest, a radiolabeled piece of DNA (probe DNA) is incubated with the filter. Probe DNA must be complementary to sequences adjacent to the specific restriction site polymorphism so that it will bind to specific sequences on the filter. If the filter is then exposed to x-ray film (a process known as autoradiography), the size of the complementary DNA fragments will be revealed.

Many RFLP's are caused by a single base pair change, resulting in the presence or absence of the restriction site. The RFLP, therefore, exists in only two forms. Because each individual has two copies of the RFLP, three different combinations are possible (Figure 2A); the individual may be homozygous for one form of the RFLP (a genotype of 1,1), homozygous for the second form (a genotype of 2,2), or heterozygous (a genotype of 1,2). With luck, such a polymorphic marker may be informative for linkage in a large, well-characterized family. If a polymorphic marker and the disease locus are located on the same chromosome, one would expect the disease phenotype to be coinherited (to cosegregate) with one of the polymorphic alleles more than 50% of the time. If, on the other hand, the DNA marker and the disease gene were on different chromosomes, then one would expect them to coinherited approximately 50% of the time, or to segregate independently. In Figure 4A, individuals thought to be affected by an inherited disorder are indicated by filled circles or filled squares. The genotypes for these individuals are also shown by an RFLP marker that shows two different alleles. The affected father in generation 1 is heterozygous at this hypothetical marker locus (a genotype of 1,2), so his children should be informative for linkage; that is, each chromosome is marked by a different allele. Each child should inherit either the 1 or the 2 allele from the affected father. In this example, all affected children inherit the 1 allele, and all nonaffected children inherit the 2 allele. If this pattern persists in subsequent generations, one can conclude that the polymorphic marker and the disease gene are likely to be located on the same chromosomal segment, or linked.

Although very useful, RFLP markers may have reduced power because of inadequate heterozygosity in the population. If, for example, the affected father in generation 1 of the pedigree in Figure 4 had a homozygous genotype (a genotype of 1,1) at this marker locus, all of his children would be expected to inherit one copy of that allele (Figure 4B). In this example, both affected and normal children would inherit one copy of the 1 allele, and the marker would be uninformative for linkage.

Markers for Variable Number of Tandem Repeats

Fortunately, another kind of DNA marker exists, one that detects many different polymorphic forms in the general population. Human DNA contains sequences that are tandemly repeated throughout the genome (Figure 2B; References 14 and 15). The function of these tandem repeats is uncertain, but the number of repeats is highly variable, and these variations can be detected. The vast majority of these sequence variations are stably inherited (that is, they do not mutate with high frequency), so they can be used for linkage analysis. Probes for markers based on the variable number of tandem repeats (VNTRs) will detect a large number of variations in the general population, greatly increasing the likelihood that individuals in a family will be heterozygous, thereby increasing the informativeness of the marker for linkage. Like RFLPs, VNTRs can be detected by use of restriction enzymes in a technique similar to the one described above (Figure 3).

Polymerase Chain Reaction–Based Markers

Recently, a number of different types of repeated units have been identified in human genomic sequences (Figure 2C; References 16–18). As with VNTRs, the function of these repeated sequences is uncertain, but they can be used for linkage analyses if the polymorphic forms can be detected. In general, however, these repeated sequences are too small (for example, two base-pair, or dinucleotide, repeats like CA repeats) to be detected by a strategy that uses restriction enzymes. However, these small repeated sequences can be detected by use of a recent technical advance, the polymerase chain reaction (PCR; Figure 5). In PCR, a double-stranded piece of DNA is denatured and then enzymatically copied, or amplified. After one round of amplification, the result is a net doubling of DNA. If the process is repeated through multiple cycles of amplification, multiple copies of DNA will result. Thus, PCR is an enzymatic DNA copying machine.

PCR specificity is in part dependent on the starting material, or template DNA. In the experiments described here, template DNA is usually human genomic DNA that contains two double-stranded copies of every locus in the genome. A second critical factor for PCR specificity results from the sequence of small synthetic pieces of DNA of approximately 20 nucleotides called oligonucleotide primers. Like probe DNA, a primer is designed to be complementary to a specific DNA sequence in the template DNA. In general, two different primers (a primer pair) that flank a specific locus of interest are used in each PCR reaction. When the template and primer pair are denatured and then cooled, the primers will bind to complementary sequences on the template. The enzyme, or polymerase, that copies DNA in PCR reactions is primer-dependent; that is, the polymerase will copy only the specific sequences that are adjacent to the primer. Furthermore, the polymerase is unidirectional; that is, it will copy adjacent sequences only from the 5' to the 3' direction.
A. RFLP

INDIVIDUAL A

RESTRICTION ENZYME CUT SITE

INDIVIDUAL B

B. VNTR

INDIVIDUAL C

TANDEM REPEAT

INDIVIDUAL D

C. Sequence-based polymorphisms

INDIVIDUAL E

DINUCLEOTIDE REPEAT

INDIVIDUAL F
In this way, the specific sequences that separate the primer pair may be amplified from a heterogeneous population of template DNA. If the template DNA used in the reaction is human genomic DNA, each locus is represented twice, and both will be copied. If the two loci are identical, the PCR products will also be identical and, therefore, indistinguishable. If, however, the amplified locus contains a sequence-based polymorphism (sequence-tagged site, or STS) and if the individual to be tested is heterozygous for that polymorphic locus, both copies will be amplified, resulting in two PCR products of slightly different size (Figure 2C). These products can be distinguished by acrylamide gel electrophoresis, a technique that is very sensitive and can distinguish between DNA fragments differing in size by only one or two nucleotides. If the oligonucleotides used in the PCR reaction are radiolabeled, the polymorphic products can again be visualized by autoradiography.

PCR can also be used to identify nonrepetitive sequence-based polymorphisms like single base pair substitutions, deletions, or insertions. Two commonly used techniques that can detect single base pair variations are PCR–single-strand conformation polymorphism analysis (PCR-SSCP; References 19 and 20) and PCR–denaturing gradient gel electrophoresis (PCR-DGGE; Reference 21). The first step in both techniques is to amplify a small (approximately 400-base pair) section of DNA from an individual’s genomic DNA. If that individual has two different versions of the DNA segment of interest, PCR will amplify both, and two different amplification products will result. These two different products can then be distinguished by SSCP or DGGE. The principle underlying SSCP is that during electrophoresis, a single strand of DNA will migrate through a non-denaturing gel at a rate that depends on the size and the specific sequence of the strand. Another strand of the same size containing a single nucleotide substitution or other sequence polymorphism will, under certain electrophoretic conditions, travel through the same gel at a slightly different rate, presumably because of conformational differences. In DGGE, by contrast, double-stranded DNA fragments are separated by electrophoresis through a gel that contains an increasing gradient of denaturant. As the DNA passes through the gel, it eventually reaches a concentration of denaturant that will convert the DNA from the double-stranded to the single-stranded form, dramatically changing the mobility of the fragments. The point at which the fragment of double-stranded DNA begins to denature depends in part on the sequence, and a single base pair substitution may make a significant difference. If the PCR products are radiolabeled, separated by SSCP or DGGE, and subjected to autoradiography, different polymorphic forms of homologous segments of DNA can be identified and used as DNA markers for
linkage analysis. Because PCR-SSCP and PCR-DGGE can be used to examine many different DNA samples simultaneously, these techniques have been particularly useful in screening for disease-associated mutations.

**Genetic Recombination**

Linkage analysis would be simple (albeit much less powerful) if there were no recombination; in the absence of genetic recombination, only 22 markers would be required to determine the autosomal location of a genetic locus. However, parental chromosomes are not transmitted to offspring in their original form; during the generation of germ cells, recombination, or crossing over, occurs between loci on homologous chromosomes (Figure 6). Genetic recombination events that occur during the first phase of meiosis (meiotic recombination) occur with a frequency roughly proportional to the distance between the loci. Thus, two genetic loci that are separated by a few thousand base pairs would rarely recombine. Conversely, two loci on the same chromosome that are separated by 20 million base pairs would recombine frequently, probably in approximately 20% of meioses. Two loci on the same chromosome that are so far apart that they recombine in 50% of meioses will appear to be unlinked, as the same pattern of coinheritance (50%) would appear if the markers were located on separate chromosomes. From these data, the genetic distance between two linked loci can be estimated by calculating the extent of recombination that has occurred between them. In this way, ordered linkage maps of polymorphic DNA markers can be generated.

Because of recombination, it may be necessary to test hundreds of different polymorphic markers with different chromosomal locations before linkage is found, and even then the identification of linkage is not guaranteed. For example, if a disease gene is located on a chromosomal segment that carries no defined markers, linkage would not be obtained. Thus, recombination has complicated linkage analysis in families and necessitated the development of many genetic markers. Once linkage
between a polymorphic marker and a disease locus has been identified, however, the identification of recombination between the disease locus and additional flanking markers is a powerful method for refining the location of the disease locus.

**Linkage Maps of DNA Markers**

Technical advances in molecular biology have facilitated the development of extensive banks of DNA-based polymorphic markers. Like genes, these markers have specific and consistent chromosomal locations. By linkage analysis, the position of one marker relative to other markers can be determined. In this way, linkage maps of genes and DNA markers have been generated for every human chromosome. An example of a linkage map for markers on chromosome 11 is shown in Figure 7.

The task of generating a complete linkage map of the human genome is far from completion; by examining Figure 7, for example, one can see that large gaps in the map for chromosome 11 exist. One goal of the Human Genome Project is to refine linkage maps so that continuous maps exist with markers separated by no more than 1 or 2 million base pairs. Nevertheless, the currently available DNA markers and linkage maps have greatly facilitated linkage studies involving human genetic disorders, problems that appeared to be unapproachable only a few years ago.

**Linkage Analysis in Practice**

Recently, we undertook a linkage study that may serve as an example of the usefulness of this approach. The long QT (LQT) syndrome is an inherited disorder that causes syncope and sudden death from cardiac arrhythmias, particularly malignant arrhythmias like torsade de pointes and ventricular fibrillation. Presymptomatic diagnosis of LQT has been based on the identification of a prolonged QT interval on ECG, a finding that is often associated with this disorder. Unfortunately, this diagnostic test is imprecise, and many cases are not detected until symptoms arise. This problem is particularly unfortunate because treatment options exist for LQT.

It is reasonable to propose that improved diagnosis and treatment of LQT might result from a better understanding of the mechanisms that underlie this disorder. Unfortunately, we know very little about the pathogenesis of LQT. Mutations in sodium, potassium, chloride, or calcium channel genes or in genes that regulate these genes or gene products could lead to delayed myocardial repolarization in affected patients. This, in turn, could cause secondary depolarizations (afterdepolarizations) and predisposition to ventricular arrhythmias. There is little evidence, however, to support or refute any of these hypotheses, and testing each of them would be a formidable task, particularly because many of these genes have not been identified.

LQT appears to be inherited as an autosomal dominant trait, so it is reasonable to hypothesize that this disorder is caused by the inheritance of a single gene. This hypothesis can be tested by linkage analysis. An advantage of this approach is that no knowledge of the biochemistry or physiology of the disorder is required for success.

**Families**

The first step in linkage analysis is to identify a suitable family. We were fortunate in having access to the largest single family with LQT that has been described (Figure 8). This family has been followed and carefully characterized for many years by Michael Vincent and Katherine Timothe of the LDS Hospital in Salt Lake City, Utah. The story of this family begins in the mid 1800s, when the two brothers shown in the second generation of Figure 8 emigrated from Denmark to a small town in Utah. We do not know much about their health except that they were excluded from the family profession of fishing for health reasons. Nevertheless, they fathered the offspring shown in Figure 8. As one can see in the figure, the LQT phenotype is inherited as an autosomal dominant trait; every generation contains affected individuals, offspring of affected parents are affected approximately 50% of the time, and the sexes are approximately equally represented. This pattern of inheritance suggests that the LQT phenotype...
in this family results from the inheritance of a single gene.

**Phenotypic Characterization**

The observable properties, such as the physical appearance, of an organism is said to be its phenotype. By contrast, the genetic constitution of an organism is said to be its genotype. The phenotype of an organism is produced by its genotype in conjunction with the environment. The second step in linkage analysis is to accurately determine the phenotype with respect to LQT of every individual in the family. As linkage analysis is based on a comparison of genotype and phenotype, phenotypic misclassifications may lead to false conclusions about linkage, so this step is an important one. Figure 9 shows the spectrum of corrected QT interval (QTc; Reference 38) for family members at risk for inheriting the LQT gene. As one can see, two peaks exist, one with a mean that approximates the mean for a control group shown at the bottom of Figure 9. The second peak is much longer, with a mean of about 0.48 second. It is apparent, however, that these two curves overlap. To solve this problem, we chose a conservative approach to phenotyping. Asymptomatic individuals with a QTc of 0.41 second or less were classified as normal, and symptomatic individuals with a QTc of 0.45 second or greater and asymptomatic individuals with a QTc of 0.47 second or greater were classified as affected. All family members that did not meet these diagnostic criteria were classified as uncertain. As a result, the family was divided into three approximately equal groups.

**Genotypic Characterization**

The third step in linkage analysis is genotyping, a process that is outlined in Figure 3. DNA extracted from the leukocytes of each family member is incubated with a restriction enzyme. The resultant restriction fragments are separated by agarose gel electrophoresis and transferred to nylon filters. These filters are then incubated with a DNA marker that detects RFLPs in the general population. The pattern of RFLPs for each individual in the study is entered into a computer relational data base that interfaces with the Linkage programs. These programs analyze the phenotypes and genotypes of individuals in the family and calculate the odds that a DNA marker and the disease gene are linked. The odds are presented in a logarithmic form called the LOD score (LOD is an acronym for the logarithm of the odds). A LOD score of +3 (odds of 1,000/1) is considered to be good evidence for linkage. A LOD score of −2 (odds of 1/100) effectively excludes linkage at a given locus.
As we had few other physiological clues and no data about the chromosomal location of the LQT locus, we began our study with highly polymorphic markers on chromosome 1 and worked our way to chromosome 22. In the process, LOD scores of more than +2 (odds favoring linkage of more than 100/1) were identified with two different markers. Although tantalizing, these scores proved to be the result of chance, because the scores did not improve when adjacent markers were tested. As linkage was not identified in our first series of experiments, we began to test additional markers from untested chromosomal regions until linkage was finally identified (Figure 7). In the process, we successfully scored more than 245 markers and excluded approximately 60% of the human genome. Linkage was identified using the Harvey ras-1 (H-ras-1) gene as a marker; this marker detects a VNTR polymorphism located on the short arm of chromosome 11 (chromosome 11p; References 44 and 45). The LOD score was +16.43, indicating that the odds favoring linkage were greater than $10^{10}$/1.

**Genetic Distance**

As mentioned above, genetic recombination complicates linkage analysis and certainly increased the complexity of our study. Recombination frequencies, however, can be converted to genetic map distances using specific mapping functions that take into account undetected double-crossover events. For example, if recombination between two loci occurs with a frequency of 1%, the loci are said to be separated by a genetic distance of 1 centimorgan (cM). If, on the other hand, the recombination frequency is 20%, the map distance is approximately 26 cM (the relation between map distance and recombination fraction is not linear at high recombination frequencies because of the frequency of double recombinants). On average, 1 cM corresponds to a physical distance of about 1 million base pairs of DNA. In our study, no recombination was observed between the disease locus and H-ras-1. Thus, the maximum LOD score (Table 1) was identified at a recombination fraction of zero. If, however, the data showed recombination between the two loci, the computer would have calculated a maximum LOD score at a specific recombination fraction, or theta ($\theta$). As no recombination was observed between LQT and H-ras-1, these loci are tightly linked in genetic terms and are likely to be in close physical approximation (probably within 3 million to 6 million base pairs).

On the basis of the complete linkage between H-ras-1 and LQT, this proto-oncogene became a candidate gene for the disorder. This hypothesis was especially exciting because a physiological rationale for the involvement of ras in LQT exists. Recent experiments done in a cell-free system have shown that ras proteins and GAP, the GTPase activating protein, can regulate atrial potassium channels. As abnormal myocardial potassium currents could cause LQT, it was not difficult to imagine that nononcogenic mutations in H-ras-1 might cause this disorder. Further evidence that a candidate gene like H-ras-1 is the disease gene can be obtained by identifying mutations in sequences derived from affected individuals. To date, no mutations have been identified in the coding sequence of the H-ras-1 gene in LQT pa-
patients (unpublished observations). On the basis of these data, we cannot be definitive about the candidacy of H-ras-1 in LQT. As will be described in more detail below, many other candidate genes for LQT are likely to reside in this chromosomal subunit, and each must be considered until the disease gene is identified.

**Locus Homogeneity**

As a result of these linkage studies, genetic diagnosis of LQT for members of the family shown in Figure 8 is secure. Can these markers be used to diagnose LQT in other families? In the six additional and unrelated LQT families studied to date, the answer is yes. One of these secondary families is shown in Figure 11. The genotypes for each individual using the H-ras-1 gene as a VNTR marker are also shown. In the families studied, nine alleles were observed at this locus, and they are numbered 1–9. Focusing on the lower left section of the pedigree in Figure 11, one can see that the unaffected father is homozygous at the H-ras-1 locus with a 4,4 genotype. The phenotypically affected mother, on the other hand, is heterozygous at ras with alleles 3 and 8. All of the children should inherit the 4 allele from their father and either the 3 or the 8 allele from the affected mother. As one can see, all the affected children inherited the 8 allele from their mother, whereas normal children inherited the 3 allele. In the rest of the family, every affected individual inherited the 8 allele, indicating that in this family, like the original family studied, LQT is probably linked to loci on the short arm of chromosome 11. The combined LOD score for the six additional families was 5.25, indicating that the odds favoring linkage were greater than $10^7/1$. Again, the maximum LOD score was at a recombination fraction of zero, indicating that no recombinants would exist in any of these families. Thus, the H-ras-1 marker can be used in these families for presymptomatic diagnosis of LQT.

**Limitations of Linkage Analysis**

Although linkage analysis using recombinant DNA technology is quite powerful, it has limitations. It is possible, for example, that mutations at several different genetic loci cause a phenotype that is similar to LQT. Locus heterogeneity for other disorders, such as hypertrophic cardiomyopathy and the Marfan syndrome, has been described, and it would not be surprising to learn that LQT is similar. In that event, markers on chromosome 11p15.5 would not be useful for genotypic diagnosis in unlinked families.

Even if the majority of LQT proves to be a result of mutations in one gene, linkage analysis cannot be used to determine the gene-carrier status of individuals in very small families or in sporadic cases. In Figure 11, linkage between the marker and the disease gene was suggested because the disease phenotype always cosegregated with a specific genotype (the 8 allele) in this large family. Multiple meioses were scored in several generations, and marker informativeness was good (affected parents were heterozygous). In a smaller family (Figure 12A) or in a sporadic case (Figure 12B), it may not be possible to establish linkage, even if the marker is highly polymorphic and very tightly linked to the disorder. In the small family shown in Figure 12A, identification of linkage is not possible because there are not enough meiotic events available for a statistically significant LOD score to be established. In the

![Figure 8. Long QT syndrome (LQT) kindred 1532. Individuals having the characteristic symptoms of fainting or sudden death due to ventricular arrhythmias, with prolongation of the QT interval on ECG, are represented by filled circles or squares. Unaffected individuals are represented as empty circles and squares; individuals with an equivocal or unknown phenotype are stippled. The pedigree structure has been altered to protect confidentiality. Informed consent was obtained from each individual before the study began. In this family, LQT is inherited as an autosomal dominant trait.](http://circ.ahajournals.org/doi/abs/10.1161/01.CIR.85.6.1982)
sporadic case shown in Figure 12B, on the other hand, no meiotic event can be scored because neither parent is known to be phenotypically affected. The only exception to this limitation occurs when the genetic variation that causes the disorder is detectable as a RFLP or other polymorphism. For example, the mutation that causes sickle cell anemia causes a specific restriction enzyme site change that is detectable as a RFLP. The polymorphism detected by the H-ras-1 gene is a VNTR and is not the cause of LQT, so improved diagnosis of LQT in sporadic cases will not be available until we have identified and characterized the mutations that cause this disorder. Sporadic cases of LQT may be a tremendous help, however, because they may lead us to

**Figure 9.** Graphs showing distribution of corrected QT (QTc) intervals for long QT syndrome (LQT) kindred 1532 (panel A) and for a control group (panel B). The number of individuals with a given QTc is indicated on the Y axis, and the actual QTc interval in seconds is indicated on the X axis.

**Figure 10.** Variable number of tandem repeats (VNTR) marker shows misinheritances. Genotypes are shown for individuals in a family by use of a VNTR marker located on chromosome 2. The genotypes for the children are flanked by genotypes for the mother at left and for the father at right. Each child inherited one allele from the mother and one allele from the father. One child (arrow) fails to inherit an allele from the father. If this pattern is confirmed after resampling and repeated testing using multiple markers, it is suggestive of false paternity, and this child would not be included in a linkage study.
new, easily detectable, mutations in the LQT gene, thereby defining that gene. For that reason, we are eager to examine these individuals.

Identification of Disease Genes by Mapping and Mutational Analysis

The obvious concern that one has when initiating a linkage study is that linkage will not be identified. As DNA markers and linkage maps continue to improve, such concerns will be minimized, and continued improvement will be limited by the availability of families and phenotyping difficulties. Once linkage between a DNA marker of known chromosomal location and an inherited disease has been clearly identified, the focus turns to identification of the disease gene, and a number of strategies become possible. With luck, a gene that was previously mapped to the same chromosomal segment will become a candidate for the disease gene because of its chromosomal location and the feasibility of its involvement in the disorder. Frequently, however, no reasonable candidate gene is apparent, and a different, much more labor-intensive approach must be taken. In general, primary linkage analysis localizes a disease gene to a DNA segment of approximately 4 million to 10 million bases on either side of the linked marker. This DNA segment may contain hundreds of genes, most of which have yet to be identified and characterized. To identify the disease gene in such a large sea of nucleotides can be daunting, but it can be done. The general strategy that can be used has been referred to as “reverse genetics” or “positional cloning,” but we generally use the phrase “mapping and mutational analysis” because it more clearly defines the process. The first step is to confine the disease gene to a chromosomal segment of about 1 million base pairs. This can be done by a combination of refined linkage and physical mapping. To improve the linkage map, one must identify new polymorphic markers that are very tightly linked to, but are recombinant with, the disease phenotype in families. These markers are called flanking markers and can be identified from cloned DNA segments from the region of interest. As the goal is to identify linked markers that recombine with the disease phenotype, it is very helpful to identify and characterize new extended families with the disorder. Another localizing method that has proved very helpful for disease gene identification is the identification of chromosomal anomalies like translocation and small deletions that are associated with the disease. If such an anomaly exists in a patient’s DNA, its identification may help localize the disease gene to a segment of several hundred bases, greatly facilitating the process.

Once the disease gene has been confined between flanking markers to a chromosomal segment of about 1 million base pairs, the identification of genes that will be candidates for the disease gene solely on the basis of their location becomes feasible. First, the DNA that

<table>
<thead>
<tr>
<th>Recombination fraction</th>
<th>0.00</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOD scores</td>
<td>16.44</td>
<td>15.18</td>
<td>13.81</td>
<td>10.79</td>
<td>7.42</td>
<td>3.76</td>
</tr>
</tbody>
</table>

Linkage analysis was performed with the computer program LINKAGE. A DNA marker defining a restriction site polymorphism at the H-ras-1 locus, pUC6J6.6, was found to be tightly linked to the long QT syndrome (LQT) locus, with a maximum LOD score of 16.44 at a recombination fraction of zero. This score corresponds to odds of \( >10^{14}/1 \) in favor of linkage. LOD scores have been calculated with penetrance of 0.90. As no estimates of LQT gene frequency exist, for purposes of this study the frequency of this rare disease gene was assumed to be 0.001. LOD, logarithm of the odds.

FIGURE 11. Long QT syndrome (LQT) pedigree showing H-ras-1 genotypes. Individuals having the characteristic pattern of fainting or sudden death due to ventricular arrhythmias, with prolongation of the QT interval on ECG, are represented by filled circles or squares. Above each symbol, individual alleles are listed for the marker pUC6J6.6 (Harvey ras-1). At this variable number of tandem repeats marker locus, the restriction enzyme Msp I revealed nine distinct alleles within the families, ranging in size from 1 to 5.1 kilobases (kb). The disease gene cosegregates with the 1.2-kb allele (No. 8) in this kindred. The pedigree structure has been altered to protect confidentiality.

FIGURE 12. Pedigrees showing limitations of linkage analysis. Linkage analysis is not informative for genetic diagnosis in small families (panel A) or in sporadic cases (panel B) because a pattern of phenotype/genotype cosegregation cannot be established. Individuals thought to be affected by a disorder are indicated by filled squares, and individuals thought to be phenotypically normal are indicated by empty circles or squares. The mutant gene causing the disease is indicated by a filled chromosomal segment. Genotypes for an adjacent variable number of tandem repeats marker are indicated numerically.
spans the flanking markers must be isolated and cloned in overlapping segments of 30,000-400,000 bases. DNA markers from these cloned segments must be developed so that the orientation of the overlapping clones can be determined. This process of physical mapping, sometimes called chromosome walking, is laborious, but improvements in the size and specificity of genomic DNA libraries and technical advances continue to facilitate the process. Next, the genes that reside in the physically mapped DNA segments must be identified, generally by using genomic DNA clones to screen complementary DNA libraries. Any complementary DNA that maps back to the correct chromosomal segment immediately becomes a candidate for the disease gene and must be characterized. At this point, the goal is to identify disease-associated mutations in the candidate gene, a process that may involve many different techniques, often including PCR-SSCP or PCR-DGGE and direct sequencing. Evidence that a candidate gene is the disease gene includes the cosegregation of a mutation with the disease in one family, the identification of mutations in the same gene in different families, and the identification of germ line mutations in sporadic cases of the disease. Once the gene has been identified, the structure and function of the protein product of the gene can be examined by use of biochemical and physiological techniques. These studies may support the hypothesis that specific mutations are important in disease pathogenesis and should facilitate insight into the molecular mechanisms of the disease. The process of disease gene identification by mapping and mutational analysis is challenging, but the ends justify the effort.

Summary

Linkage analysis can be used to determine the chromosomal location of a disease gene in a setting in which little or no physiological information is available. Recombinant DNA technologies have greatly facilitated the development of new, highly polymorphic DNA markers, which, in turn, have made general linkage analysis feasible in humans. Most of our progress has been in studies of disorders that are caused primarily by the inheritance of one gene, so-called single-gene disorders, like LQ1, which we have linked to chromosome 11p15.5. However, better mapping techniques and detailed genetic maps also offer hope for an improved understanding of the genetic variations that are important in complex disorders like hypertension and atherosclerosis.

Acknowledgments

I would like to thank M. Leppert, J. Mason, S. Prescott, and W. Jin for critically reviewing this manuscript; D. Atkinson for creating the figures; and R. White, J.-M. Lalouel, G.M. Vincent, M. Curran, A. Ewart, L. Piatek, K. Timothy, and V. Turner for their help and suggestions.

References

Linkage analysis and long QT syndrome. Using genetics to study cardiovascular disease.

M Keating