Title
A molecular genetic evaluation of familial dilated cardiomyopathy

Author
Spudich, Serena S.

Publication Date
1996

Peer reviewed|Thesis/dissertation

A Molecular Genetic Evaluation of Familial Dilated Cardiomyopathy

by

Serena S. Spudich

A Thesis

Submitted in partial satisfaction of the
requirements for the M.D. with Thesis Program
of the
University of California, San Francisco

List of Figures

Table of Contents

Abstract

A form of familial dilated cardiomyopathy associated with conduction system defects seen in a large midwestern kindred, family OSU, has been previously mapped to an approximately 450 kilobase region in the centromeric region of chromosome 1 (*CDDC-1* locus). We have endeavored to combine physical and genetic mapping methods in order to positionally clone the gene responsible for this disease phenotype. Genome sequence sampling was used to search for candidate genes, candidate gene fragments, and polymorphic markers within the interval. Approximately 300 kb of random sequence was obtained from digestion, subcloning, and automated sequencing of the P1 bacteriophage clones spanning the interval. Over ten candidate genes were identified within our interval using these methods; future mutation analysis should yield a single disease gene. We have also performed genetic analysis on a newly identified kindred, family MAQ, with a phenotype of conduction disease and dilated cardiomyopathy. Clinical evaluation of this family indicates a similar but not identical phenotype as compared to that seen in family OSU, indicating some clinical heterogeneity of this disorder. Data from this family was used to perform linkage analysis using short-tandem repeat (STR) sequences. The dilated cardiomyopathy in family MAQ was demonstrated to be linked to the known *CDDC-1* locus on chromosome 1 (maximum lod score of 2.27). Thus, despite clinical heterogeneity a common gene defect causes disease in each family. Further polymorphism analysis at markers in the interval revealed a recombinant event in the genome of a single MAQ family member at *GBA*, an STR marker internal to the original flanking marker of the interval. These analyses led to the refinement of the disease gene interval at the *CDDC-1* locus.

INTRODUCTION

Familial disease presents a particularly ominous and frustrating challenge to the clinician and patient, since many such disorders are of unknown etiology, inexorable in their progression, and refractory to current methods of treatment. However, familial diseases are of particular interest to scientific investigators since their inherited nature permits examination through molecular genetic approaches. In order to elucidate the molecular basis of an illness, researchers can study families which manifest the inherited form of the disease and search for the gene which is mutated in their condition.

In recent years, specific mutations in protein-encoding genes have been identified in association with the development of numerous cardiovascular diseases, including hypertrophic cardiomyopathy, the Holt-Oram syndrome, and hereditary hemorrhagic telangectasia (1-3). Such studies have to varying degrees elucidated the molecular basis for the structural defects seen in these disorders; ongoing research should reveal more information about the mechanisms and pathophysiology of these defects. Our increased understanding of such disorders ultimately should lead to new treatment options for these conditions, allow for mutation detection in persons of unknown disease status, as well as provide insight into the pathogenesis underlying a variety of forms of heart disease.

While much progress has been made in the molecular genetic evaluation of the above disorders, these are each rare conditions causing

morbidity and mortality in a small subset of the worldwide population. In contrast, a gene defect underlying idiopathic dilated cardiomyopathy, the most common form of primary myocardial disorder, has yet to be identified. The dilatation of the ventricles leading to heart failure and rhythm disturbances in primary idiopathic dilated cardiomyopathy is indistinguishable pathophysiologically from that seen in secondary heart failure due to more common systemic and cardiac conditions (4). Furthermore, the hallmark condition of cardiac dilatation seems to be the final common pathway for a range of cardiac disorders which, taken together, are the main cause of death in developed nations.

Given the above, researchers have been led to investigate the genetic basis of idiopathic dilated cardiomyopathy. The inherited form of the disease is manifested differently between families, and studies suggest that they are a range of conditions, determined in some cases as monogenic, others as multifactorial traits. The fact that as compared to familial hypertrophic cardiomyopathy, so little is known of the genetic basis of dilated cardiomyopathy likely reflects the broad and varied nature of the disorder in terms of both inheritance patterns and subtle clinical expression across families. In addition, whereas pathologic and biochemical studies of familial hypertrophic cardiomyopathy yielded information which suggested potential disease gene candidates(1), the pathophysiology and histopathology which are seen in idiopathic dilated cardiomyopathy are nonspecific (4). Thus, a promising candidate gene underlying this condition has yet to be suggested. Furthermore, cytogenetic abnormalities which are rare but which may be helpful in the localization and identification of gene defects have never been detected in individuals affected with dilated cardiomyopathy.

We have set out to investigate the precise location and nature of the gene defect which leads to a phenotype of familial dilated cardiomyopathy associated with conduction system disease as seen in a large midwestern kindred, family OSU (5). Since no clear candidate genes for this disorder have been previously suggested by pathologic or physiologic studies, we have used a variety of molecular genetic techniques in an effort to positionally clone the gene mutated in this disorder. Our efforts have also included recruiting and evaluating a new family affected by familial dilated cardiomyopathy with conduction disease in efforts to expand our database for the gene search as well as evaluate clinical and genetic heterogeneity in this disorder.

Background and Relevance

Dilated cardiomyopathy describes heart failure characterized by ventricular dilatation and systolic dysfunction in the absence of identifiable causes such as coronary disease, alcohol or infection. In the past this disorder has been referred to as congestive cardiomyopathy but recently the term dilated has been preferred due to the significant cardiac dilatation which can be assessed earlier than the signs and symptoms of congestive failure seen late in the natural history of this condition(6).

The most universal definition of dilated cardiomyopathy is the World Health Organization criteria, described as an ejection fraction of less than 55% and an end-diastolic volume of greater than 110 cubic mm (7). The diagnosis of dilated cardiomyopathy in contrast to ischemic dilatation

(sometimes confusingly referred to ischemic cardiomyopathy), requires some evidence of normal coronary arteries in patient at risk for coronary atherosclerosis. In stringent studies of dilated cardiomyopathy, cardiac catheterization studies are performed on all patients over age 50 years (those at high risk for coronary artery disease); greater than 50% stenosis of any single coronary vessel is considered consistent with ischemic disease (8).

The relevance of pursuing the genetic basis of dilated cardiomyopathy is based on its prevalence and the morbidity and mortality associated with the disorder. The estimated prevalence of dilated cardiomyopathy is 36.5/100,000 in the US population (4). Men are slightly more often affected than females, and middle-aged people more than the young or elderly. In terms of the morbidity of the disorder, it has a severely malignant course. Dilated cardiomyopathy is the primary indication for heart transplantation in Western nations, and is the cause of 35% of cases of NYHA designated Class III and Class IV heart failure in the USA (4).

A clinical history in fully developed dilated cardiomyopathy often reflects symptoms of left ventricular dysfunction including fatigue, dyspnea on exertion, paroxysmal nocturnal dyspnea, orthopnea, and chronic cough. Either or both ventricles may be involved, however, and symptoms of right ventricular failure are often present, including ankle or abdominal edema. Signs seen on physical exam reflect these, and include evidence of cardiomegaly, jugular venous distention, and often a third and/or fourth heart sound. Atrioventricular (AV) valve disease may develop as a result of congestion and systolic regurgitant murmurs may be heard at the mitral and tricuspid areas.

The natural history of dilated cardiomyopathy is quite variable, sometimes reflecting an apparently acute onset disorder following a vague viral prodromal illness, but often suggesting a gradual cardiac dysfunction that has been developing for a period of years before presentation (9). In general, this condition is characterized by a progressive decline in systolic function ultimately leading to death from pump failure or arrhythmia. The mortality associated with this disease has consistently been calculated at less than 50% survival at two years after diagnosis (6).

Studies reveal evidence of abnormal conduction and cardiac architecture as seen on the electrocardiogram (ECG), including left bundle branch block, AV node conduction delay, atrial enlargement, ventricular hypertrophy, and tachycardia in cases of advanced disease. The echocardiogram often defines the increased ventricular dimensions, impaired contraction, and ventricular hypertrophy seen in this disorder. Lab tests are nonspecific and reflect the end organ damage of heart failure when present.

Diffuse cardiac dilatation is a common endpoint of a variety of known systemic conditions, including thyrotoxicosis, acute viral myocarditis, and muscular dystrophy. In addition, a dilatation process which is phenotypically indistinguishable from that seen in the disorders above is seen in the setting of several specific heart muscle diseases. These include infiltrative disorders such as amyloid disease, connective tissue disorders, granulomatous disease, toxicity due to alcohol, and metabolic deficiencies such as dietary selenium deficiency or congenital carnitine deficiency (9).

Approximately half of those with dilated cardiomyopathy have a primary disorder known as idiopathic dilated cardiomyopathy (IDC) (10). To define IDC, histological studies must exclude a number of specific heart muscle diseases including those noted above. Myocardial tissue in IDC displays non-specific histology, with general signs of myocyte dystrophy: fibrosis, myocyte dropout with scattered myocyte hypertrophy, degeneration or vacuolization of individual myocytes, and fatty replacement of tissues (5, 11).

Numerous conditions have been previously suggested as etiologic factors in IDC. These include an undetected viral myocarditis, post-partum dilated cardiomyopathy, and proposed immunologic mechanisms (6, 12, 13). However, no persuasive evidence for any of these as the mechanism in the majority of cases has yet emerged. Previous estimates of the prevalence of familial IDC have suggested that only 5-7% of cases were inherited (6). However, recent studies indicate a far higher proportion of inherited disease. In an evaluation by Michels et al., 20% of apparently sporadic cases of IDC in the US were found to be familial based on the thorough clinical evaluation of all first degree relatives of index patients with IDC (8). A study with a similar design by Keeling & McKenna in the UK found that 25% of all of their index patients with IDC had an inherited form of the illness when relatives who were previously thought to be unaffected were shown to have IDC by clinical studies (14).

The phenotype of the familial form of IDC is similar but not identical to that of sporadic idiopathic dilated cardiomyopathy. Overall survival does not differ between these two forms of the disorder; however the age of onset of inherited IDC is usually in the second or third decade compared with the

fourth and fifth decades for pure sporadic IDC (8). In retrospective analysis, cases of post-partum, presumed post-viral, or alcoholic dilated cardiomyopathy have occurred in families which manifest inherited IDC (6, 8). This suggests that a familial predisposition toward dilated cardiomyopathy may be unmasked in these patients by conditions such as pregnancy or the presence of toxins, possibly via neurohormonal mechanisms, or states of increased myocardial strain such as elevated blood volume or impaired cardiac contractility due to the presence of toxins. In addition, since autoimmune disease has previously been hypothesized to be a causative factor in IDC, HLA types and angiotensin converting enzyme polymorphisms have been investigated in familial IDC. Association studies between these genetic variables and development of IDC have shown no clear evidence of a connection (12, 13).

Inherited IDC

Genetic studies have revealed a number of IDC kindreds with segregation patterns consistent with X-linked modes of inheritance. Most of these families show dilated cardiomyopathy as one aspect of a more global muscular pathology. Forms of muscular dystrophy, for example, are accompanied in most patients by a profound dilated cardiomyopathy which often leads to heart failure as the primary cause of death. Mutations in the dystrophin gene have been identified with Becker muscular dystrophy, while absence of dystrophin characterizes the more severe Duchenne's muscular dystrophy.

A small number of individuals with X-linked cardiomyopathy without clinical evidence of skeletal muscle disease were found to have mutations in the dystrophin gene (14, 15). However, these patients showed evidence of dystrophy on biopsy of their skeletal muscles. Furthermore, dystrophin mutations do not lead to all cases of X-linked IDC. A recent investigation of male index patients with dilated cardiomyopathy demonstrated that none of the affected individuals had detectable dystrophin gene mutations (16). In other studies, families with dominant cardiac phenotypes were found to have identical dystrophin mutations as those with dominant skeletal phenotypes (15). Controversy over whether there is a cardiac specific promoter region in the dystrophin gene which leads to cardiac dysfunction which is relatively worse than the skeletal dysfunction has yet to be resolved (14, 17, 18).

X-linked Emery-Dreifuss dystrophy involves muscles diffusely but specifically leads to a significant dilated cardiomyopathy with conduction abnormalities. However, the disease gene has yielded little information regarding pathophysiology or molecular mechanisms thus far. Barth's syndrome, a constellation of cardiomyopathy, short stature, and neutropenia, is due to mutations in a novel protein called tazaffin. The functions of this protein in the heart or other tissues are unknown.

Mitochondrial mutations have been shown to cause a phenotype of dilated cardiomyopathy in families which manifest the disorder in a maternal pattern of inheritance. Specifically, defects in genes for mitochondrial oxidative phosphorylation have been identified in these families (15).

A small number of families have been identified which develop IDC in an autosomally inherited pattern. As noted above, the familial form of isolated dilated cardiomyopathy is indistinguishable in survival from sporadic IDC. Thus, isolated familial IDC is a malignant illness which results in 50% mortality within two years of onset. In the familial form of the disease, death usually occurs in the second or third decade of life. As a result, many families which carry the disorder are only 1-3 generations in size since individuals die before reaching reproductive age (11). These small families are difficult to evaluate using standard genetic analysis, since the power of studies is limited by size.

In light of this, intermediate phenotypes, such as dilated cardiomyopathy associated with cardiac conduction abnormalities, become useful subjects of study. Overall, 20-30% of familial IDC has been associated with conduction disease. In this setting, associated phenotypic traits may allow for identification of these affected with disease before heart failure and sudden death occur. Furthermore, the kindreds under study tend to be larger since the fatal aspects of the disorder do not develop until the 5th or 6th decades. An evident problem with this approach is the fact that in studying the genetic basis of an intermediate phenotype, we will not necessarily identify a gene which is also responsible for a phenotype of pure IDC. However, since dilated cardiomyopathy arises in a variety of contexts with such a range of manifestations, it is clear that there are multiple mechanisms which lead to a final common pathway of cardiac dilatation and heart failure. Identifying one mutated protein underlying a phenotype which involves dilated cardiomyopathy will likely suggest a previously

unknown collection of related or interactive proteins which may all be involved in generalized phenomenon of cardiac dilatation and failure.

Positional Cloning

One specific large midwestern family (family OSU) which has a form of familial dilated cardiomyopathy accompanied by conduction system defects has been previously identified and longitudinally studied by Wooley and colleagues at Ohio State University (see Figure 1) (5). In this family, although the progressive AV conduction abnormalities manifest early and allow for identification of affected individuals, dilated cardiomyopathy, sudden death, and heart failure are of late onset. Disease is inherited in an autosomal dominant fashion, such that a heterozygous parent has a fifty-percent chance of passing the condition to his or her offspring (5). Based on the fact that all affected individuals are the offspring of affected parents, there appears to be complete penetrance of this disease gene, manifest in an age-dependent manner.

Using the genetic and clinical information from seven generations of the extensive family OSU, a random genome search using short tandem repeat polymorphisms was performed by members of the Seidman laboratory which linked the disease gene to a locus at the centromeric region of chromosome 1, denoted *CDDC-1* (11). The peak multipoint logarithm of the odds (lod) score at the 7-10 cM interval between flanking markers at this locus was 13.5. The lod score indicates the statistical likelihood that two genetic loci (such as a marker and a disease gene) and linked; a lod score of more than three in the context of a random genome search indicates a

significant likelihood of linkage. Given linkage of disease to chromosome 1q1-1p1, we have employed positional cloning methods in order to identify the disease gene in the severely affected family OSU. Since *CDDC-1* was identified, several novel disease loci for families with an intermediate phenotype of dilated cardiomyopathy with or without conduction disease have been identified by linkage studies, including 3p22-p25, 1q32, and 9q13 (10, 19, 20). This genetic heterogeneity suggests that either completely disparate genes, or related genes with interactive or similar functions may lead to similar end points of cardiac dilatation.

In the absence of knowledge about the biochemical or functional defect leading to disease, an approach termed positional cloning has been developed which can be applied to the identification of new disease genes (21, 22). Positional cloning refers to the process of cloning a gene of interest based primarily on information on the position of the disease gene within the genome, as opposed to functional cloning which identifies a gene based on a protein product's presumed biologic function. In essence, an interval defined by linkage studies is refined and narrowed using sequence data and location of polymorphisms until the disease gene is found (see Figure 2).

Variants of the positional cloning approach have been used to identify the disease genes responsible for a variety of genetic disorders, the first of these being cystic fibrosis. Since the histopathology and pathophysiology seen in dilated cardiomyopathy are entirely non-specific, no single protein has been implicated as mutated in disease. As a result, positional cloning techniques are ideal in the investigation of the molecular mechanisms underlying this condition.

I have used various strategies to refine and narrow the disease gene region between the original flanking markers on chromosome 1. These strategies include sequence analysis of the DNA contained within the disease gene interval to identify candidate genes and find new markers (23), and PCR-based polymorphism analysis to look for cross-over events and narrow our interval based on the DNA of members of a newly identified family (24).

Previous Work on Positional Cloning Project

As noted above, members of the Seidman laboratory previously performed a random genome search using short tandem repeat sequences and found linkage of disease in the OSU family to the pericentromeric region of chromosome 1-- locus *CDDC-1*. The maximum multipoint lod score was found to be 13.5, and the interval was defined between markers *D1S305* & *D1S176* (11).

A YAC contig -- a set of large-insert vectors containing DNA which spans the region of interest on Chromosome 1 -- was then created by pulling out a series of contiguous regions of DNA with known short marker sequences (short tandem repeat sequences, or STRs). Assembly of a YAC contig was facilitated by the fact that the approximately 7-10 CM disease gene interval is at a junction of two regions of close homology with mouse chromosome 1 & 3 (25). Furthermore, the relative physical locations of genes in this region are highly conserved between species. As a result, an STR-based human map was created by anchoring STRs in the gene-based mouse map, which allowed for the development of a YAC contig based on

STS content with minimal genome walking (26). This revealed 30 YAC clones that covered the interval with significant redundancy. A P1-bacteriophage contig consisting of twelve P1's and two cosmids was then rapidly assembled using STRs found in both the YAC contig and the gene-based markers. This P1/cosmid contig yielded almost complete coverage in the interval spanned by the YAC contig (27).

METHODS

I. Genome Sequence Sampling: Development of a Physical Map

A. P1 Subcloning

A partial digestion of P1 DNA was performed with the restriction enzyme *Sau* 3A in an effort to create overlapping DNA fragments (22). The resulting digestion products were size fractionated on a 1.2% agarose gel. Fragments ranging from 1-4 kb in size were cut from the gel and purified using the QAIEX small molecular weight gel extraction protocol from QIAGEN. A plasmid library was developed by ligating the resulting elution product into Bam H1 cut pBlueScript vectors using the standard ligation protocol and DNA ligase (see Figure 3). DH5 alpha *E. coli* cells were transformed with the resulting plasmids using a heat shock introduction of the plasmid DNA under standard conditions, and plated onto ampicillin containing growth media. After growth overnight, colonies were stabbed and introduced into 4 ml preps of Luria broth in the presence of ampicillin, then incubated with aeration for 12-16 hours at 37 degrees (21, 22, 28, 29).

DNA was purified from the plasmids using an alkaline lysis with ethanol washing protocol (PRIMM miniprep kits), and eluted with 75 uL distilled water for storage at 4 degrees. 2-5 uL of each sample was cut with *Eco* R1 and *Not* 1 to liberate the insert DNA then the product was run on a 1% agarose gel to assess for the presence, size, and quantity of insert DNA.

B. Subclone Sequencing

Approximately 0.5 ug of template DNA was used in a 20 uL cycle sequencing reaction using 100 ng of either the T3 or T7 oligonucleotide primers and the *Taq* polymerase-containing ABI Prism Dye Terminator Cycle Sequencing Ready Reaction Mix from Perkin Elmer Corporation (30). Cycle sequencing was performed in a GeneAmp PCR System 9600 machine using the following thermal cycle: 96 degrees C for 10 seconds, 50 degrees C for 5 seconds, and 60 degrees C for four minutes, for a total of 25 cycles, followed by a hold at 4 degrees C indefinitely. The primers homologous to the pBlueScript plasmid adjacent to the polylinker are T3 (5'CCTCACTAAGGGAACAAAAGCTGG3') and T7 (5'ACTCACTATAGGGCGAATTGGGTA3') and are oriented in opposite directions (30, 31) (see Figure 4).

Extension products were purified using spin-column purification according to the Centri-Sep protocol in order to remove unincorporated dye terminators and primer fragments. Purified products were dried in a vacuum centrifuge at room temperature for storage in the dark or sequencing analysis.

Automated sequencing was accomplished according to the Sanger dye-terminator method with an Applied Biosystems Inc. Prism 377 fluorescence sequencer from Perkin Elmer Corporation. Samples were resuspended in a formamide-containing loading buffer then loaded onto a prepared and heated 6% polyacrylamide 36 cm well-to-read gel made with reagent grade acrylamide and urea. Electrophoresis was then performed with laser excitation of the dye-terminators as they passed through the base of the gel, with recording of the sequence data in a Power Macintosh folder for further processing.

C. Processing and Analysis of Sequence Data

Vector and primer sequences were edited from the sequence data using the Factura processing program from ABI, and sequence was translated into a gcg format interpretable by on-line databases using ABI Sequence Navigator software.

Sequence data was transferred to a VAX for comparisons to known sequences in the on-line databases of the human genome project (22, 23, 32). The BLAST program was first used for identification of alu sequence, and areas of high identity to alu were edited from the sequence data (33). MULTIBLAST commands which allow for processing of multiple sequences at one time were then used to search for identity of procured sequences with known genes which might serve as potential disease gene candidates. EST's with identity to our sequence were also identified as known transcripts coding for unknown proteins in our region. Areas of high homology or

identity to *E. coli* sequence were also edited out or disregarded as likely contamination (32).

DNAStar Lasergene Navigator software was also employed to create contigs of overlapping sequences and then analyze the resulting expanded sequences for homology with known genes at amino acid and protein levels. Amino acid comparisons were performed by translating the sequence into all six possible reading frames and comparing them to protein sequences in the NCBI non-redundant database using the BLASTX command. This data was all compiled in an effort to identify sequence structures of interest as well as to an*Not*ate the developing physical map of our region (23, 29).

All raw sequences were also analyzed for repetitive elements by comparison with previously identified sequences as well as with a general screen for di-, tri- and tetra- nucleotide repeats (32). Once repeat sequences were identified, they were evaluated for polymorphic potential according to size and regularity of the repetitive element. PCR-based polymorphism analysis was performed using oligonucleotide primers synthesized to flank the repeats against DNA from family members under study (see below for methods). Further haplotype analysis was performed when the repeat sequence was ordered with respect to the other markers in the interval.

II. Evaluation of new IDC (family MAQ)

A. Assignment of clinical diagnoses

All studies were performed in accordance with the Brigham and Women's Hospital Human Subject Committee Guidelines. Members of the family MAQ were clinically evaluated. Medical and family histories were obtained from all living family members and medical or post-mortem records were obtained when available from affected or deceased family members. All surviving family members were studied by physical examination, twelve lead electrocardiogram (ECG), and two-dimensional echocardiogram. ECG's and echocardiograms were interpreted using standard criteria and without knowledge of genotype.

Disease status was assigned based on data collected and medical records. Surviving individuals were considered affected if they manifested evidence of ventricular dysfunction on echocardiogram, or fixed conduction disease on ECG. Deceased individuals were considered affected based on medical records or transmission of disease to their offspring. Individuals less than 25 years were considered of indeterminate clinical status unless they were clearly affected with disease.

B. Linkage Analysis

At the time of clinical evaluation venous blood samples were obtained. Genomic DNA was prepared from frozen whole blood using the standard SDS-proteinase K extraction method. DNA samples from family members were genotyped for intragenic short-tandem-repeat (STR) polymorphisms. Analysis of STRs was performed using synthetic oligonucleotide primers (25-mers). STRs used in linkage analysis were selected based on their known chromosomal location in our interval in

published linkage maps, their linkage to the *CDDC-1* locus in our analyses, and their polymorphic information content (11, 25).

The polymorphisms used to evaluate linkage to the *CDDC-1* locus were two STRs present in the interval, *MUCIN* and *D1S305*. Alleles were typed by PCR amplification of genomic DNA samples. Polymorphic genomic sequences were amplified using standard conditions for 25 temperature cycles in a reaction volume of 10 uL (containing 200ng genomic DNA, 0.4mM each labeled and unlabeled primers, 200uM each dATP, dCTP, dGTP, dTTP, 1X PCR buffer, and 0.625 units Taq DNA polymerase). Forward primers were end-labeled for 30 minutes at 37 degrees in a 25 uL final volume containing primer, 5uL P32 labeled ATP, 1X polynucleotide kinase buffer, and 10 uL of polynucleotide kinase. Products were combined with 15uL formamide loading buffer, followed by denaturing gel electrophoresis in a 6% sequencing gel. The gel was then dried and exposed to film for 10 to 16 hours. Genotypes were scored by two independent observers, without knowledge of clinical data.

Two point linkage analyses were performed using the MLINK program. Logarithm of the odds (lod) scores were calculated with a penetrance set at 0.95 and at various recombination fractions. Because a genome-wide analysis was not performed, a lod score of more than +1.2 was taken as evidence that the odds in support of linkage to our markers are greater than 20 to 1 (24, 34).

Analysis in support of linkage would prompt further efforts at definition of the *CDDC-1* locus through linkage analysis at other markers in

this region. In addition, when appropriate, all nine polymorphic loci from the interval could be used to construct the chromosomal haplotype segregating with disease in the family MAQ (24).

RESULTS

II. Genome Sequence Sampling

A. Physical Map

Approximately 300 kb of random shotgun sequence was obtained from eight of the P1 clones spanning the *CDDC-1* interval. Analysis of this sequence data led to the identification of numerous candidate gene fragments in the disease gene region. Several genes known to map to our region were specifically localized to P1 clones or oriented with regard to other known genes or markers in the interval. Finally, several previously identified genes were localized for the first time to this chromosome centromeric region of the human genome based on finding coding sequence for these proteins through sequence sampling.

Sequence data acquired using the Perkin Elmer dye terminator method and ABI 377 Prism automated sequencer was of an average length of 500 bp of unambiguous sequence. Of the total shotgun sequence obtained, approximately 1/6 of the sequence data obtained was found to be that of P1 vector itself, a quantity consistent with the ratio of vector/insert size (27). Contamination with *E.coli* was entirely prep dependent; most P1 preps yielded about one sequence of *E. coli* in ten of human genomic sequence

(human sequence was identified based on the presence of EST's or *alu* sequences). A single P1 yielded approximately 90% *E. coli* sequence when studied so this P1 prep was re-grown using original glycerol stocks; this repeat preparation resulted in two-thirds human sequence. Ubiquitous Alu sequences were edited out manually with the most conservative margins (33).

Multiple candidate gene fragments (approximately 20 distinct ESTs) were identified in the disease gene region (Figure 5a). These were considered candidate gene fragments based on their identity with sequences in the dbEST on-line database. Most candidate gene fragments were identified originally as clones ranging from 300-500 bp in length. Each of these stretches of sequence was thought to be a small fragment of the candidate gene from which it arose. Many more sequences showed some evidence of homology to known genes but their similarity was not significant. Candidate gene fragments were assembled when possible into contigs using DNAStar Lasergene Navigator software. Most candidate gene fragments were independent, that is, they did not overlap with other candidate gene fragments sequenced from the same P1 clone. However, sequences which were found to lie in ESTs were blasted against the dbEST and dbNR to identify overlapping ESTs and genes. These were assembled into contigs and evaluated for intron/exon boundaries, unrecognized coding sequence and homology to known genes.

A number of genes which were previously known to map to our interval were identified in the process of genome sequence sampling. These include genes encoding glucocerebrosidase (GBA), a protein kinase (PKLR),

glucocerebrosidase pseudogene (GBAPS), thrombospondin 3 (TSP3), a tumor MUCIN (MUCIN), synaptotagmin, metaxin, and a protein of unknown function termed 'YL' (Figure 5b). These were known to lie near or within the interval by identification of markers and arrangement of YAC or P1 clones; however they were oriented or precisely positioned within the interval by sequence analysis.

Several genes were newly localized to our disease gene interval based on our sequence sampling efforts. These include the acetylcholine beta receptor-2 gene (ACRB-2), adenosine deaminase co-factor (ADA), farnesyl pyrophosphate synthase (FPPS), DAP-3, and a human form of notch-2 (NOTCH) (Figure 5c). All of these are potential candidate genes for conduction disease with dilated cardiomyopathy based on their known or presumed function and ubiquitous expression.

B. Repeat sequences

Our sequence sampling efforts also identified polymorphic repeat sequences in the genome. Through genome sequence sampling methods, a previously unidentified CATT repeat was discovered on P1 100. Oligonucleotide primers were synthesized to flank this repeat region, and alleles in OSU family members were amplified using standard PCR conditions. This marker, denoted *SS100t318*, was found to be polymorphic in OSU family members. One allele was concordant with disease except for in a single affected individual in the family. This finding suggested that there was a recombinant at this new P1 100 marker in the genotype of a member of family OSU, indicating that the disease gene interval could be

narrowed to less than 100 kb (from the existing interval of approximately 450 kb). Furthermore, the alleles amplified at this new marker fit into assembled haplotypes when this marker was ordered with respect to the other known markers in the region.

Using primers designed for marker *SS100t318*, the product was easily amplified from P1 100 DNA in the polymerase chain reaction, using both radiolabeled primer and non-radioactive conditions. However, when control experiments were run to ensure that this P1 100 sequence reflected DNA sequence in our interval, we found that we were unable to amplify the *SS100t318* product from any of the YAC clones spanning this region. We used chromosome 1 X-derived hybrid DNA to assess for the presence of this sequence in chromosome 1 sequence, and found that we were unable to amplify a product of the *SS100t318* primers while we could amplify control sequences.

Further genomic sequencing of the P1 100 DNA revealed sequence identity to an EST previously mapped to chromosome 20. Subsequently, a chromosome 20 control hybrid DNA sample was used in a PCR reaction with *SS100t318* primers. A product of identical size to that seen in the genomic DNA control sample was observed in the agarose gel. This finding apparently reflects chimeric nature of P1 100 insert (chimerism between chromosome 1 & 20), and suggests that the *SS100t318* tetranucleotide repeat sequence is actually located on chromosome 20. We concluded that we could not exclude P1's and narrow our interval located on chromosome 1 based on the findings with this marker which is of indeterminate location in the genome. However the fact that the genotypes generated by amplification

with these primers are so consistent with the haplotypes constructed for this region strongly suggests that the *SS100t318* marker is near a region of importance in IDC with conduction disease. Possibly, a similar marker sequence is present on both chromosome 1 and 20 which can be amplified from chromosome 1 hybrid DNA only in select conditions.

II. MAQ family studies

Clinical Evaluation

Family MAQ is a North American kindred which we identified as carrying IDC with conduction disease through a 35 year old proband, subject 221 in MAQ pedigree, who manifested conduction defects alone. This family, now residing in the Ohio River Valley, traces its ancestors to the Rhine Valley region in Germany. Of 62 family members in four generations at risk of inheriting the disorder, clinical evaluations (see methods) defined13 as affected. The inheritance of IDC with conduction disease was consistent with autosomal dominant transmission (Figure 6). Evaluation reflected high penetrance of the disease gene, since each affected individual was the offspring of an affected parent. During these studies, one individual, 2341, expired at age 31 from sudden death). Based on the natural history of progression of conduction disease to dilated cardiomyopathy, subjects were scored as affected if clinical studies demonstrated evidence of fixed conduction disease, an arrhythmia with or without cardiomyopathy, or cardiac dilatation and heart failure in the presence or absence of conduction abnormalities.

Death from heart failure or sudden death in this family ranged from 31 years of age to 59 years of age; the average age at death was 46 years (43.5 years if subject 2341 was included). Onset of CHF in two family members was at age 30 and 40 years (subject 2341 and 233, respectively). The oldest surviving affected individual was 63 years old at the time of evaluation. This subject had a permanent pacemaker placed at the age of 53 for atrial arrhythmias). Because pedigree analysis suggested that disease penetrance was age-related, individuals less than 25 years of age were considered indeterminate (14 subjects, Figure 6). Six more were scored as indeterminate due to borderline clinical criteria, such as a strongly positive clinical history without other abnormal findings, or a single borderline conduction abnormality as assessed by ECG (individuals 212, 222, 223, 341, 344, 355). Thirty-four persons were scored as unaffected.

A broad range of symptoms were present in family members at risk of inheriting IDC and conduction disease. Several at risk family members gave clinical histories of palpitations, fatigue, and dyspnea on exertion. One individual with these symptoms (individual 31) had documented two-vessel coronary artery disease by cardiac catheterization, and evidence of a remote infarct on ECG. Another individual who reported progressively worsening fatigue and shortness of breath with effort was noted to have previously undiagnosed atrial fibrillation during this evaluation (234).

One male family member, aged 57 (individual 233) first suffered the onset of CHF at age 40 years, and required heart transplant at age 45 years. Individual 34, a 63 year old female, reported a history of symptomatic

arrhythmia extending more than 20 years and had a permanent pacemaker placed 10 years prior to our evaluation.

Clinical examination revealed that multiple members of family MAQ had cardiac conduction abnormalities on ECG evaluation (Figure 7). As evidenced by the ECG, one individual(221) had sinus bradycardia with first degree AV block, two had variants of bundle branch block (LAFB and RBBB, in subjects 213 and 351, respectively), several had nonspecific T wave abnormalities (213, 221, 2341) and two had evidence of LA enlargement (213 and 351). One individual was found to have atrial fibrillation on evaluation (234). Another individual was permanently paced (34, see above).

Two dimensional echocardiograms were normal in most individuals. Two at risk family members had evidence of ventricular involvement: individual 234 had an enlarged left ventricle and her son, individual 2341, had diffuse enlargement of both the left atrium and ventricle. Incidental findings of fixed wall abnormalities and valve disease were noted in several spouses unrelated to the family.

Genetic Analysis

DNA was extracted from the blood of all 62 MAQ family members. Amplification by PCR was successful for approximately 95% of the reactions. Linkage analysis was performed in the MAQ family using two highly polymorphic STR markers. *MUCIN* and *D1S305*, markers in the *CDDC-1* interval, were selected to assess for linkage to the chromosome 1q1-1p1 locus previously linked to family OSU.

Initial data showed that disease in family MAQ was coinherited with a marker in the *CDDC-1* interval. The gene encoding the *MUCIN* protein has been previously mapped to chromosome 1q1-1p1. Analysis of a tetranucleotide repeat polymorphism in this gene revealed three alleles in the MAQ family, one of which was co-inherited with disease status without recombinants. A maximum two point lod score of 2.27 was calculated at a recombination fraction of zero, signifying 1:100 odds that the conduction disease and dilated cardiomyopathy seen in this family is due to a gene defect located near this chromosome 1 locus. The pairwise lod scores reflecting coinheritance of disease in family MAQ and polymorphic loci have been calculated at various recombination fractions (Figure 8).

To further evaluate the linkage between *CDDC-1* and the MAQ family disease we performed linkage analysis at the short tandem repeat marker *D1S305*. Amplification of MAQ family DNA with primers for this marker resulted in 8 alleles, one of which was exactly concordant with disease with the exception of a single recombinant (in subject 351). Analysis at *D1S305* also provided evidence in support of linkage, with a maximum lod score of 1.24 at a recombination fraction of 0.1 (Figure 8).

In order to refine the disease gene interval, we analyzed recombination events in the MAQ family. Additional short tandem repeat DNA probes were selected for polymorphism analysis in family MAQ, based on their heterozygosity and location in our interval. Seven polymorphic loci located in the centromeric region of chromosome 1 were amplified in MAQ family DNA to assemble the chromosomal haplotype which segregates

with disease in family MAQ. Analyses were performed using PCR amplification and gel electrophoresis, with primers for markers *D1S252, GBA, RSACA, D1S303, GATA85, LAMIN, and CATCA.*

Analysis of alleles that segregated with affection status in family OSU identified recombination events between the disease gene and polymorphic loci in one individual. Subject 351 was found to have a different allele than others affected with disease at the polymorphic marker *GBA*, a dinucleotide repeat sequence within the glucocerebrosidase gene. This marker amplified two alleles in the MAQ family and subject 351 clearly carried the unaffected allele despite clinical evidence of the disease phenotype. This indicates that a cross over event has occurred and this defines a new disease boundary at one end of the interval at *GBA* in comparison to the distal marker *D1S305.* Haplotype analysis using the entire panel of markers in unaffected as well as affected members of the family MAQ yielded no further information than did polymorphism analysis regarding recombination events (data not shown).

DISCUSSION

We have localized a disease gene for dilated cardiomyopathy with conduction disease to a small interval on chromosome 1 through a combination of genetic and physical mapping. Genome sequence sampling was performed in order to create a refined physical map of a region previously identified through genome-wide linkage studies. P1 clones containing DNA sequence whcih spanned the interval were digested and each resulting fragment was subjected to one-pass end sequencing.

Sequences were analyzed for putative exons, repetive DNA's and known genes using a variety of computer algorithms. These efforts led to the identification and localization of genes and candidate gene fragments in this region of chromosome 1. This high resolution physical map allows for the assembly of a detailed sequence-based map of this area and provides an opportunity for further structural and functional analysis of the genome.

Random genome sequence sampling was used to identify and characterize candidate genes, nucleotide repeats, and other sequences of interest in our disease gene interval. We used a method of genomic sequencing followed by database comparison to create a refined physical map and positionally clone the gene for conduction disease with dilated cardiomyopathy at the *CDDC-1* locus.

Our sequencing efforts led to the identification of multiple candidate gene fragments, on the basis of their identity with known expressed sequences, or cDNA's, in the human genome. The database comparisons have also yielded specific positional information on several genes which were previously known to lie within or near our interval. Finally, I have found that several previously unmapped genes which encode for known proteins are present in our interval. Interestingly, this has turned out to be an extremely gene-rich region, with an estimated gene density of approximately one gene per 10 kb.

We have also used short tandem repeat-based linkage analysis to demonstrate here that dilated cardiomyopathy with conduction disease in a recently identified family maps to the previously identified locus *CDDC-1* on

chromosome 1p1-1q1. Disease in the MAQ family is linked to this locus, with a maximum lod score at *MUCIN*, one of two markers tested in our interval, of 2.27 with the most conservative diagnosis assignments.

DNA from MAQ family members was amplified at markers flanking the *CDDC-1* locus on Chr 1 to assess for linkage at our existing disease locus. These experiments indicate conclusively that this family's disease is linked to the *CDDC-1* locus on Chr 1. Since this family gives a significant lod score indicating linkage, we hypothesize that a mutation in this family is due to a mutation at the same locus. Thus, they almost certainly carry mutations in the same disease gene, and there is no clear evidence of genetic heterogeneity among these two familes OSU and MAQ. Whether these families carry the same allele or whether disease is caused by distinct alleles in the same gene is presently undetermined. The identification and evaluation of new markers in this region and ultimately the disease gene itself will define the nature of the mutations fully.

This study does provide evidence that there is clinical, if not genetic, heterogenity of dilated cardiomyopathy with conduction disease as seen in the families OSU and MAQ. MAQ family disease is clinically similar but not identical to the dilated cardiomyopathy with conduction disease seen in other families whose disease links to chromosome 1 centromeric region. For example, the clinical features of patients whose disease maps to chromosome 1p1-1q1 varies subtly in the nature of conduction defects. Family OSU displays sinus bradycardia and progressive AV block, while MAQ patients develop bundle branch blocks, atrial enlargement, non-specific T-wave

changes and atrial arrythmias. The age of onset of dilitation and CHF and age at death in these two families follow a similar pattern, however.

Could these two families be distantly related? Historical data from each family suggests a common origin in the German Rhine Valley. Although each later settled in the Ohio River Valley in the United States, no direct family or historical connection has yet been made. Yet, the similarity of their phenotype and linkage to the same locus suggest that they are likely to be distantly related. Certainly, genetic heterogenity has been previously demonstrated between distinct families with dilated cardiomyopathy with conduction disase; loci at chromosome 3, distal 1, and 9 have been recently identified (10, 19, 20). Whether the genes responsible will prove to be a collection of related or homologous genes or genes encoding a set of interacting proteins has yet to be elucidated. To date no disease gene in a family manifesting this phenotype has been discovered. Analysis of individuals from further familial dilated cardiomyopathy families and of more loci will be useful in determining the precise location of the disease gene within this interval.

Several mechanisms could account for the clinical heterogeneity seen in this disorder. As we have indicated, IDC with conduction disease in these two families are both caused by mutations at the *CDDC-1* locus. However, the contiguous gene syndrome could explain the slightly different manifestation of conduction disease between these families. For example, a gene responsible for cell signaling may be contiguous to a gene specifically related to the efficacy of the atrioventricular conduction system at our disease gene locus. A partial deletion of one of these genes could have led to

the phenotype in family MAQ, whereas a larger deletion involving a portion of the contiguous gene could lead to the disease manifestations seen in the family OSU. Another explanation for the clinical heterogeneity observed in this study is that there are different mutations in a single disease gene which lead to disparate clinical syndromes; however this would be less likely if the families MAQ and OSU are actually related. Finally, it is possible that in fact these families harbor mutations in two distinct, but closely linked disease genes. However, this scenario is extremely unlikely, since it would require that two genes leading to such similiar clinical syndromes were both located at the *CDDC-1* locus.

Along with finding multiple candidate genes and candidate gene fragments within our interval, we have also been able to identify recombinant alleles carried by a MAQ family member at a marker in our interval. Individual 351 appears to be a definitive recombinant at STRs *D1S305* and *GBA*, though her DNA is uninformative at other markers based on evaluation of genotypes at eight markers in the region. Finding a marker with a recombinant allele, defined as a different allele at a marker than others with the same disease status, suggests that the marker is located outside of a disease interval. Moving one boundary of the interval to the marker *GBA* which we believe to be proximal to the previous boundary *D1S305* allows us to reduce the area potentially containing the mutated gene.

Thus, we have been able to use genetic information combined with physical mapping data to both narrow and characterize our disease gene interval. Diagnosis has proved difficult for a few family members who have borderline criteria for disease; further clinical information such as the

progression of clinical findings and data from several as yet unstudied family members will be useful in future assessment. Due to the small size of the MAQ family and the ambiguity of a few diagnoses in upper generations, however, the above may be the extent of the information that can be gleaned from family MAQ.

Our small interval still contains at least ten known genes as well as multiple candidate gene fragments as previously defined. Ongoing efforts on this project include Northern blot analysis of the candidate gene fragments and genes found through genome sequence sampling to assess for expression in cardiac tissue. Once certain transcripts have been identified as expressed in atrial, ventricular, or conduction system tissue, they will be evaluated as candidates by screening the known sequences for mutations in the DNA of affected family members. This may be accomplished by detection of large deletions or mutations using Southern blotting of genomic DNA. If no mutations are detectable in a candidate gene using this method, the precise nucleotide sequence of this gene in a patient's genomic DNA can be compared to that in a normal individual to evaluate for base-pair substitutions.

After the disease gene has been identified, the function of the protein which it encodes may be investigated through a variety of techniques. If it encodes a known protein, previous studies of this molecule should lead to a rapidly increased understanding of the mechanisms underlying cardiac dilitation and conduction abnormalities. If the disease gene is a novel protein, structural and functional studies must be undertaken to reveal the means through which the mutated version leads to disease. For example,

open reading frames in the sequence can be identified, and these stretches may be assessed for homology to other genes with known functions. Based on primary and secondary structure a predicted protein structure can be developed and analyzed for potential interactions with other proteins, and ultimately for function in human cardiac tissue.

Certain conclusions regarding the nature of the disrupted protein can be drawn from the clinical and genetic observations of this study. For example, given the fact that the disease is clinically manifest in the heterozygous state, the function of the responsible protein must be related to the dose of the allele product. Due to the lack of consanguinity in the families under study, we have no clear evidence of homozygous offspring; however, the severity of the heterozygous state suggests that the homozygous state might be lethal. In additin, the fact that the effect of the mutated protein is related to the quantity of the product indicates that its function is not entirely determined by transcriptional regulation. Presumably upregulation of gene expression is insufficient in this context for the full function of the protein which is mutated in this disorder.

Identifying the type and function of the disrupted protein in this dilated cardiomyopathy phenotype will be useful in further screening of potentially affected family members as well as in understanding the proteins involved in the phenomenon of cardiac dilatation, a significant cause of heart failure. Genetic studies of cardiovascular diseases are gaining importance primarily because new tools of genetic investigation have recently become available which can help to pinpoint and closely characterize the mutations associated with familial cardiac diseases. Many

'multifactorial' cardiovascular diseases such as hypertension and premature atherosclerosis appear to run in families. Characterization of the genetic bases of these disorders may ultimately be a powerful tool in their diagnosis, prevention, and treatment.

Acknowledgments

I am very grateful to Drs. Christine and Jonathan Seidman for inviting me into their laboratory, for encouraging and guiding me in my interest in genetics, and for serving as role models in the world of academic science and medicine. I appreciate the time, effort, and teaching that was offered to me by all of the members of the Seidman laboratory, especially Diane Fatkin with whom I learned the basics of molecular genetics during my tenure in the lab. In particular I thank my mentor on this project, Dr. Calum MacRae, for sharing with me his enthusiasm and knowledge of genetics.

The author thanks the Sarnoff Endowment for Cardiovascular Science, Inc. for generous support during the research tenure, as well as guidance in selecting a laboratory, and providing me with the opportunity to present my work at the Annual Meeting for the Society of Fellows in 1996.

Figure 1. Pedigree of Family OSU

A form of familial dilated cardiomyopathy accompanied by conduction system defects has been previously identified in a large midwestern kindred (Family OSU). Seven generations of this family have been evaluated over a twenty-year period. In this family, although progressive AV block and sinus bradycardia manifest early and allow for the identification of affected individuals, cardiac dilatation, heart failure, and sudden death have their onset in the fourth and fifth decades. Circles denote females, squares males, and filled shapes indicate individuals affected with disease. Inheritance is consistent with autosomal dominant transmission: each heterozygous affected parent has a 50% chance of passing disease to his or her offspring.

Figure 2. Positional Cloning Approach

In the absence of knowledge about the biochemical or functional defect leading to a disease, an approach termed positional cloning has been developed which can be applied to the identification of new disease genes. Positional cloning refers to the process of cloning a gene of interest based on its position in the genome as initially defined by genome-wide linkage studies. This is contrasted to functional cloning which identifies a gene based on a protein's presumed biologic function. In this method, an interval is defined by linkage studies and is then refined and narrowed using a combination of genetic and physical mapping techniques. We used genome sequence sampling to gain physical information about our disease gene region. Physical information was combined with genetic information

gleaned from evaluating a newly identified family with a phenotype of dilated cardiomyopathy and conduction disease in an effort to narrow our interval and identify the disease gene of interest.

Figure 3. Genomic Sequencing I

Given linkage data, a P1 bacteriophage contig was developed which contained a genomic DNA representation of the entire region covered by the disease gene interval. Multiple fragments of DNA from each P1 were generated through partial digestion with the frequently cutting restriction enzyme *Sau*3A. This process yeilded overlapping fragments which were gel isolated, and bands of 1-4 kb in size were cut from the agarose and extracted. These fragments were then ligated into pBlueScript plasmid vectors containing known polylinkers and flanking primer sites.

Figure 4. Genomic Sequencing II

After transforming DH5-alpha strain *E. coli* cells with plasmid and using ampicillin resistance to select for insert-containing plasmid clones, the cells were grown overnight in LB broth at 37 degrees. The plasmids were then extracted and DNA was purified using alkaline lysis and ethanol washing. Cycle sequencing in a thermal heat cycler was performed using the Sanger dye terminator method to amplify each fragment in the presence of dye-labelled dideoxy-chain terminators. These samples were sequenced using an ABI Prism 377 automated sequencing machine. In this method, a specific dye associated with each fo the four bases is excited by a laser beam as the samples pass through a polyacrilamide gel. The laser sends the signals to

computer software which then generates peaks corresponding to each of the four bases in real time, 'reading' the sequence in each sample.

Figure 5. Physical Map -- Distribution of candidate gene fragments and genes across the *CDDC-1* region

a. Sequencing has led to the identification of multiple candidate gene fragments on the basis of their identity to expressed sequences, or cDNA's, in the human genome (represented by the short, narrow lines below contig). The location of candidate gene fragments on this map reflects the P1 clone from which it was sequenced, and does not necessarily reflect its relative position in the genome.

b. Genes which were known to lie near or within the interval were oriented or precisely positioned within the interval by sequence analysis (represented by the arrows associated with gene titles).

c. Finally, several unmapped genes have b*Eco*me disease gene candidates after their coding sequences were discovered in this region (represented by straight lines with gene titles).

Figure 6. Pedigree of Family MAQ

We have recently identified a new sibship with a phenotype of dilated cardiomyopathy with conduction disease similar to that seen in the family OSU. Of 62 family members evaluated in four generations, 13 were scored as affected (filled shapes), 20 as unknown based on age < 25 years or indeterminate affection status (shaded shapes), and 34 were considered unaffected.

Figure 7. Clinical characteristics of affected individuals in family MAQ

This presents in a tabular format the clinical features of affected members of family MAQ as assessed by medical histories and studies including the ECG and echocardiogram.

Figure 8. MAQ Family disease linkage to chromosome 1 markers *MUCIN* and *D1S305*

Chromosome 1 short tandem repeat markers were used to establish linkage to the 1p1-1q1 locus previously identified as *CDDC-1*. Linkage studies were performed using primers flanking a dinucleotide repeat sequence located within the human tumor *MUCIN* gene, denoted *MUCIN*. lod scores were calculated at a variety of recombination fractions (thetas). The maximum two-point lod score of 2.27 was shown at a recombination fraction of zero. Pairwise lod scores were also calculated at a*Not*her short tandem repeat marker in the interval, *D1S305*. A maximum two point lod score of 1.24 was calculated at a theta set at 0.1 also reflected co-inheritance of disease with this chromosome 1 locus.

Figure 9. Disease haplotypes in affected members of family MAQ

Alleles were amplified at a total of nine markers in MAQ family DNA to identify the disease alleles and affected haplotype in this family. All alleles were exactly concordant between affected individuals except for at two markers in one individual, subject 351. With primers flanking markers at *D1S305* and *GBA*, alleles other than those seen in other affected individuals were amplified from the DNA of subject 351, an affected MAQ family

member. Thus, a definitive recombination event has taken place, altering the affected haplotype in this individual, and suggesting that the disease gene interval is bounded by *GBA* at one end rather than the previous distal marker *D1S305*.

Bibliography

1. Geisterfer-Lowrance, A., S. Kass, G. Tanigawa and e. al. 1990. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy-chain gene missense mutation. *Cell* 62:999-1006.
2. Basson, C., G. Cowley, S. Solomon and e. al. 1994. The clinical and genetic spectrum of the Holt-Oram syndrome (heart-hand syndrome). *N Engl J Med* 330:885-891.
3. Shovlin, C., J. Hughes, E. Tuddenham and e. al. 1994. A gene for hereditary haemmorhagic telangectasia maps to chromosome 9q3. *Nat Genet* 6:205-9.
4. Goodwin, J. and E. Olsen (1980) *Cardiomyopathies: realisations and expectations* (Springer-Verlag, Berlin).
5. Graber, H., D. Unverferth, P. Baker and e. al. 1986. Evolution of a hereditary cardiac conduction and muscle disorder: a study involving a family with six generations affected. *Circulation* 74:21-34.
6. Fuster, V., B. Gersh, E. Giuliani and e. al. 1981. The Natural History of Idiopathic Dilated Cardiomyopathy. *Am J Cardiol* 47:525-531.
7. 1978 in *Task Force on Cardiomyopathies*, London).
8. Michels, V., P. Moll, F. Miller and e. al. 1992. The frequency of familial dilated cardiomyopathy in a series of patients with idiopathic dilated cardiomyopathy. *N Engl J Med* 326:77-82.
9. Goodwin, J. 1970. Congestive and hypertrophic cardiomyopathies. *Lancet* I:731-739.
10. Olson, T. and M. Keating. 1996. Mapping a cardiomyopathy locus to chromosome 3p22-p25. *J Clin Invest* :528-532.

11. Kass, S., C. MacRae, H. Graber and e. al. 1994. A gene defect that causes conduction system disease and dilated cardiomyopathy maps to chromosome 1p1-1q1. *Nature Genetics* 7:546-551.

12. Limas, C., C. Limas, H. Boudoulas and e. al. 1994. Anti b-receptor antibodies in familial cardiomyopathy: Correlation with HLA-DR and HLA-DQ gene polymorphisms. *Am Heart J* 127:382-386.

13. Olson, T., S. Thibodeau, P. Lundquist and e. al. 1995. Exclusion of a primary defect at the HLA locus in familial idiopathic dilated cardiomyopathy. *J Med Genet* 32:876-880.

14. Mutoni, F., L. Wilson and G. Marrosu. 1995. A mutation in the dystrophin gene selectively affecting dystrophin expression in the heart. *J Clin Invest* 96:693-9.

15. Kelly, D. and A. Strauss. 1994. Inherited Cardiomyopathies. *N Engl J Med* 330:913-919.

16. Michels, V., G. Pastores, M. PP and e. al. 1993. Dystrophin analysis in idiopathic dilated cardiomyopathy. *J Med Genet* 30:955-957.

17. Mutoni, F., M. Cau and A. Ganau. 1993. Deletion of the dystrophin muscle-promoter region associated with X-linked dilated cardiomyopathy. *N Enlg J Med* 329:921.

18. Saito, M., H. Kawai and M. Akaike. 1996. Cardiac dysfunction with Becker muscular dystrophy. *Amer Heart J* 132:642-7.

19. Krajinovic, M., B. Pinamonti, G. Sinagra and e. al. 1995. Linkage of familial dilated cardiomyopathy to chromosome 9. *Am J Hum Genet* 57:846-852.

20. Durand, J., L. Bachinski, L. Bieling and e. al. 1995. Localization of a gene responsible for familial dilated cardiomyopathy to chromosome 1q32. *Circulation* 92:3387-3389.

21. Collins, F. 1992. Positional cloning: Let's not call it reverse anymore. *Nature Gen* 1:3-6.

22. Harshman, K., R. Bell, J. Rosenthal and e. al. 1995. Comparison of the positional cloning methods used to isolate the BRCA1 gene. *Hum Mol Genet* 4:1259-1266.

23. Kupfer, K., M. Smith, J. Quackenbush and e. al. 1995. Physical mapping of complex genomes by sampled sequencing: A theoretical analysis. *Genomics* 27:90-100.

24. Ott, J. (1994) *Analysis of human genetic linkage* (The Johns Hopkins University Press, Baltimore).

25. Weith, A., G. Brodeur, G. Bruns and e. al. 1995. Report of the second international workshop on human chromosome 1 mapping 1995. .

26. Hudson, T., L. Stein, S. Gerety and e. al. 1995. An STS-based map of the human genome. *Science* 270:1945-1954.

27. Pierce, J. and N. Sternberg. in press. Using the bacteriophage P1 system to clone high molecular weight (HMW) genomic DNA. *Methods in Enzymology* .

28. Smith, M., A. Holmsen, Y. Wei and e. al. 1994. Genomic sequence sampling: a strategy for high resolution sequence-based physical mapping of complex genomes. *Nature Gen* 7:40-47.

29. Selleri, L., M. Smith, A. Holmsen and e. al. 1995. High-resolution physical mapping of a 250-kb region of hnman chromosome 11q24 by genomic sequence sampling (GSS). *Genomics* 26:489-501.

30. Nurminsky, D. and D. Hartl. 1996. Sequence scanning: A method for rapid sequence acquisition from large-fragment DNA clones. *Proc Natl Acad Sci, USA* 93:1694-1698.

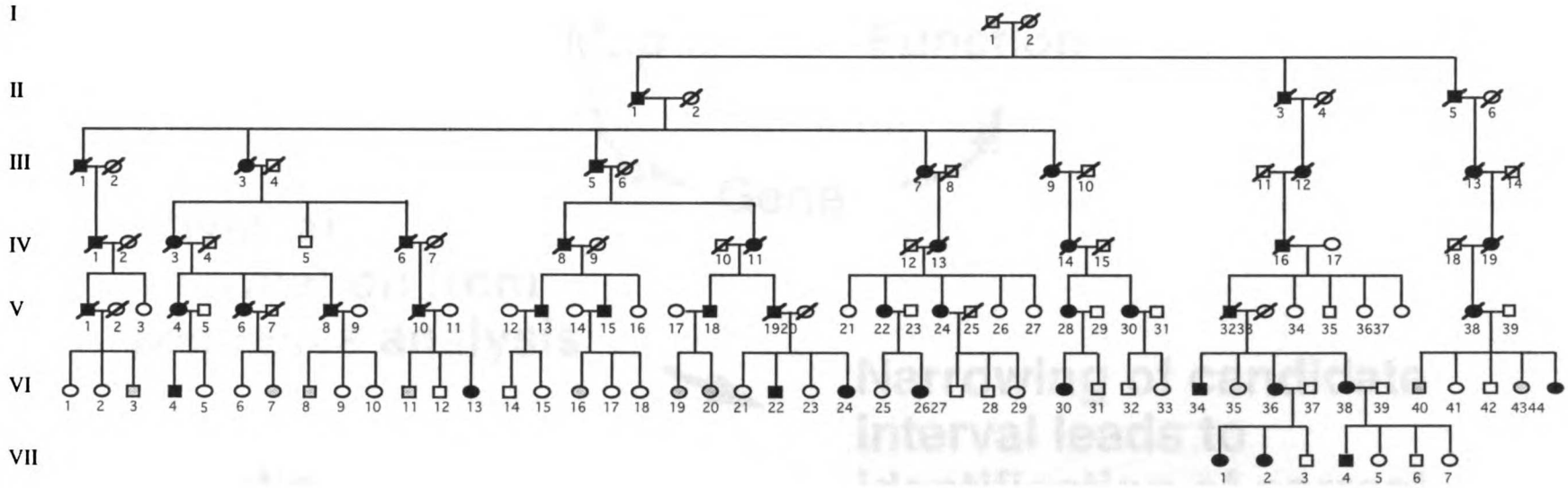
31. Pollak, M., E. Brown, Y. Chou and e. al. 1993. Mutations in the Human Ca2+-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell* 75:1297-1303.

32. Altschul, S., M. Boguski, W. Gish and e. al. 1994. Issues in searching molecular sequence databases. *Nature Gen* 6:119-129.

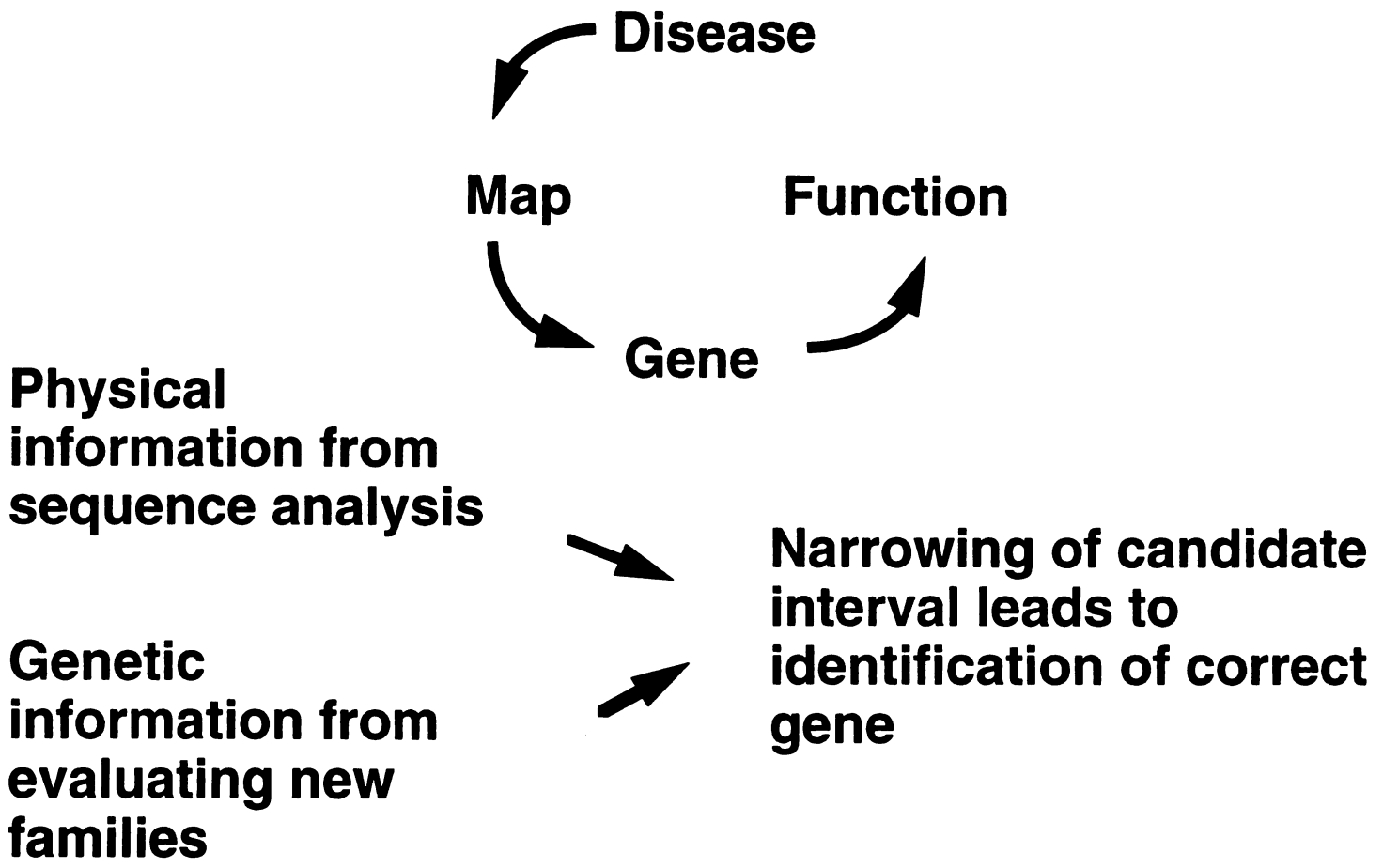
33. Yulug, I., A. Yulug and E. Fisher. 1995. The frequecy and position of *Alu* repeats in cDNA's, as determined by database searching. *Genomics* 27:544-548.

34. Lathrop, G., J. Lalouel, C. Julier and J. Ott. 1984. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci, USA* 81:3443-3446.

Pedigree of Family OSU



Positional Cloning Approach



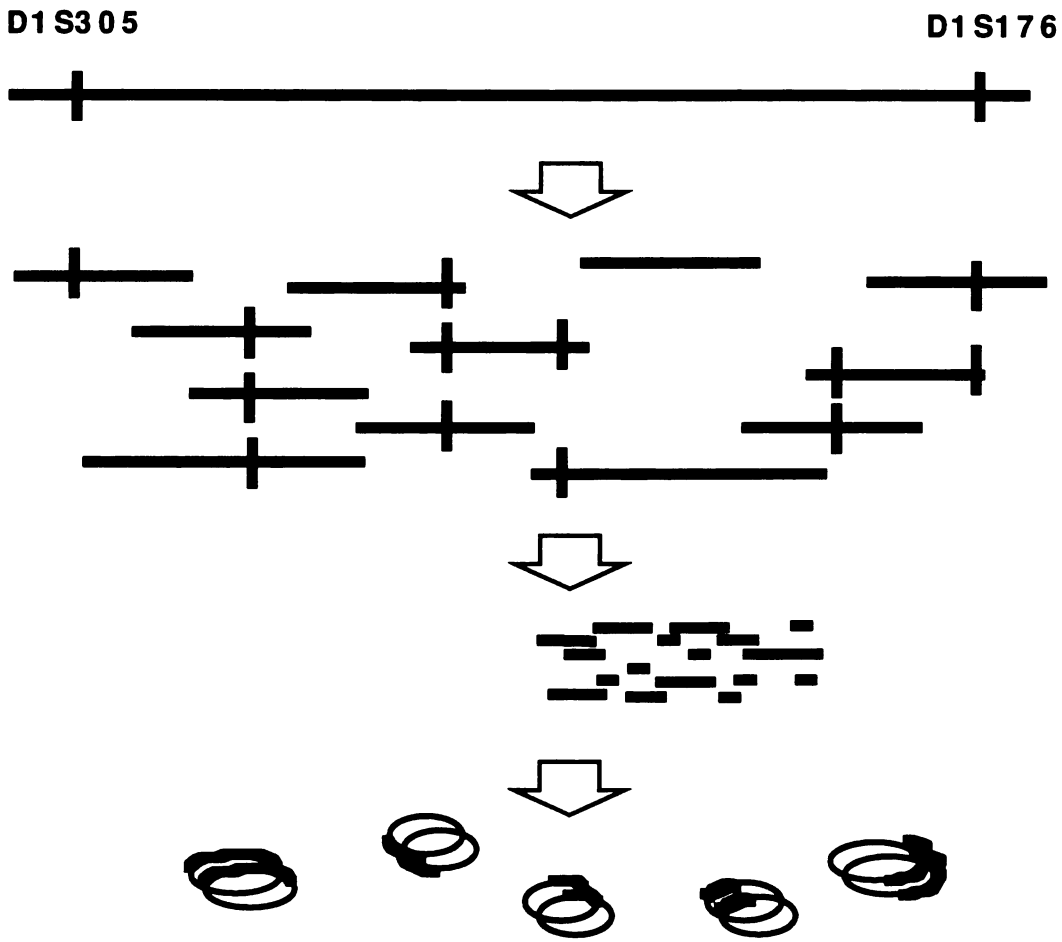
Genomic Sequencing I

Disease gene interval

P1 large-fragment DNA clones

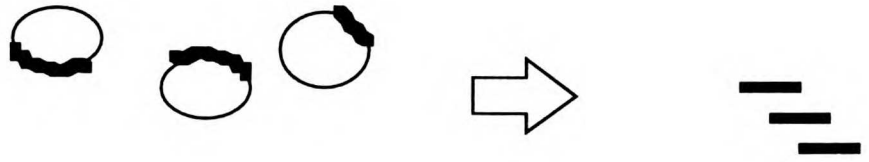
Overlapping fragments of P1 inserts

Fragments cloned into plasmid vectors

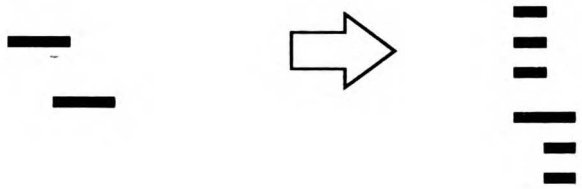


Genomic Sequencing II

Copy plasmid,
purify DNA



Cycle sequencing
of ends in presence
of dye terminators

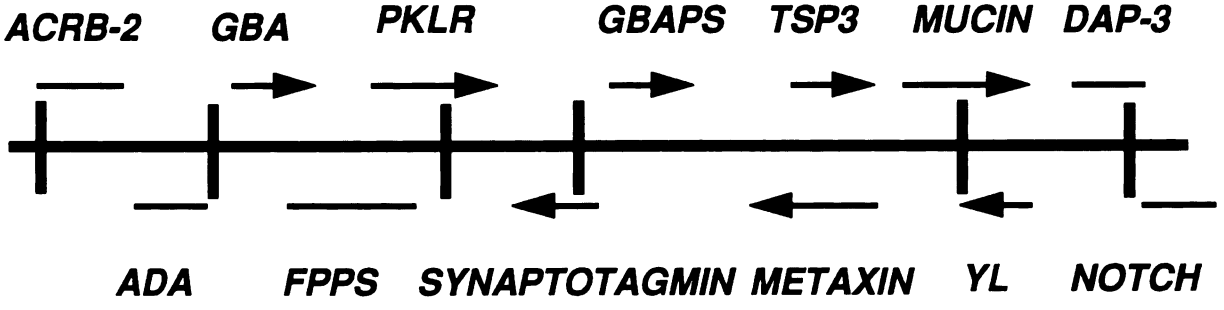


ABI
sequencing
of fragments

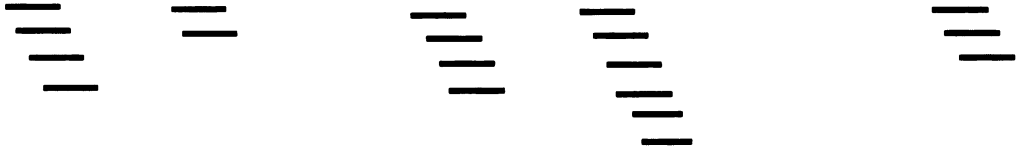
ATTCAGAACCTTGCGTTCATAG

Physical Map

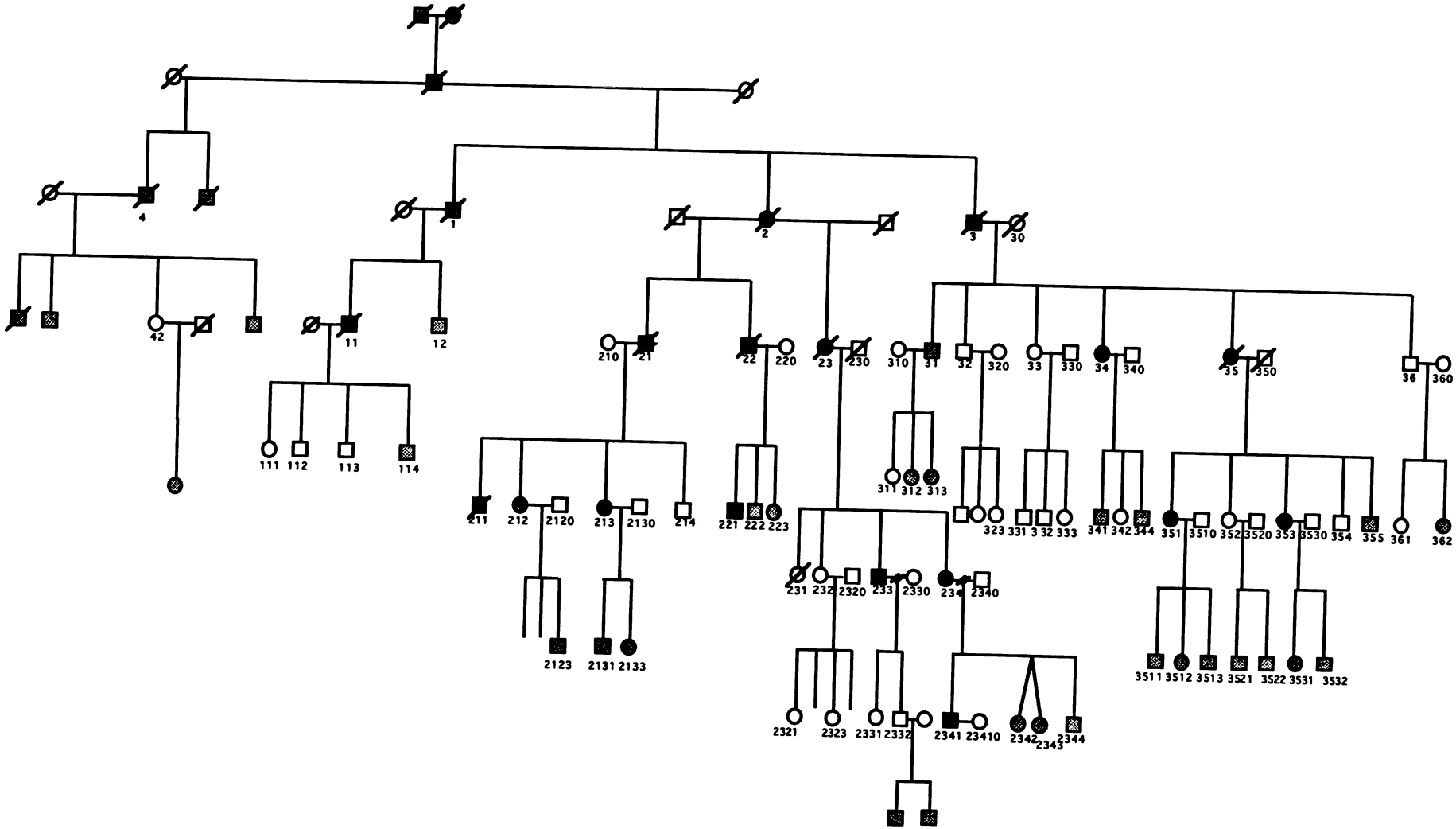
Genes found or oriented in interval



Candidate gene fragments found in interval



Pedigree of Family MAQ



Clinical characteristics of affected individuals in family MAQ

ID number	Age/Gender	Clinical manifestations/ECG/ECHO findings	Therapy
21	d.59 / M	N/A	unknown
211	d.31 / M	N/A	unknown
213	43 / F	on ECG: L anterior fascicular block, LA enlargment, T wave abnormalities	none
22	d.50 / M	N/A	unknown
221	35 / M	on ECG: Sinus bradycardia, 1st degree AV block, T wave abnormalities	medical Rx
23	d.40 / F	N/A	unknown
233	57 / M	CHF onset at age 40, S/P heart transplantation at age 45	transplant, medical Rx
234	50 / F	on ECG: Atrial fibrillation on echo: Left ventricle upper limit of nl size	medical Rx
2341	30 / M	CHF onset at age 30, on ECG: T wave abnl, on echo: LV & LA mildly enlarged	medical Rx
34	63 / F	Sx secondary to arrythmias x 20 years, pacemaker placed at age 53	pacemaker medical Rx
35	d.50 / F	N/A	unknown
351	39 / F	on ECG: LA enlargement, RBBB	none
353	36 / F	on ECG: ventricular conduction delay	none

MAQ Family Disease Linkage to Chromosome 1 Markers MUCIN and D1S305

Pairwise LOD scores of chromosome 1 loci and disease in Family MAQ

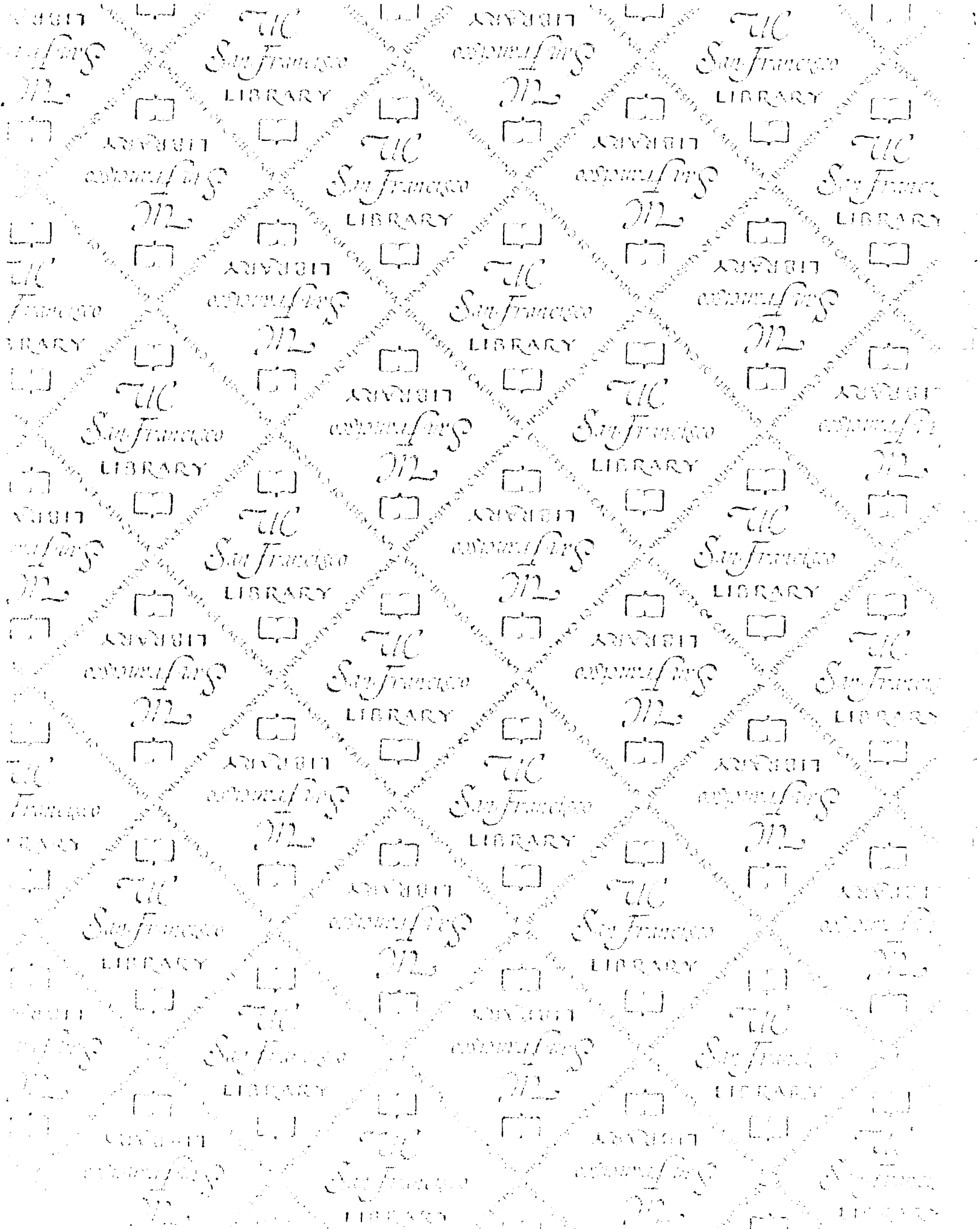
Locus	Recombination Fraction (theta)					
	0.00	0.01	0.05	0.1	0.2	0.3
MUCIN	2.27	2.26	2.17	2.00	1.55	0.99
D1S305	-3.24	0.58	1.17	1.24	0.94	0.49

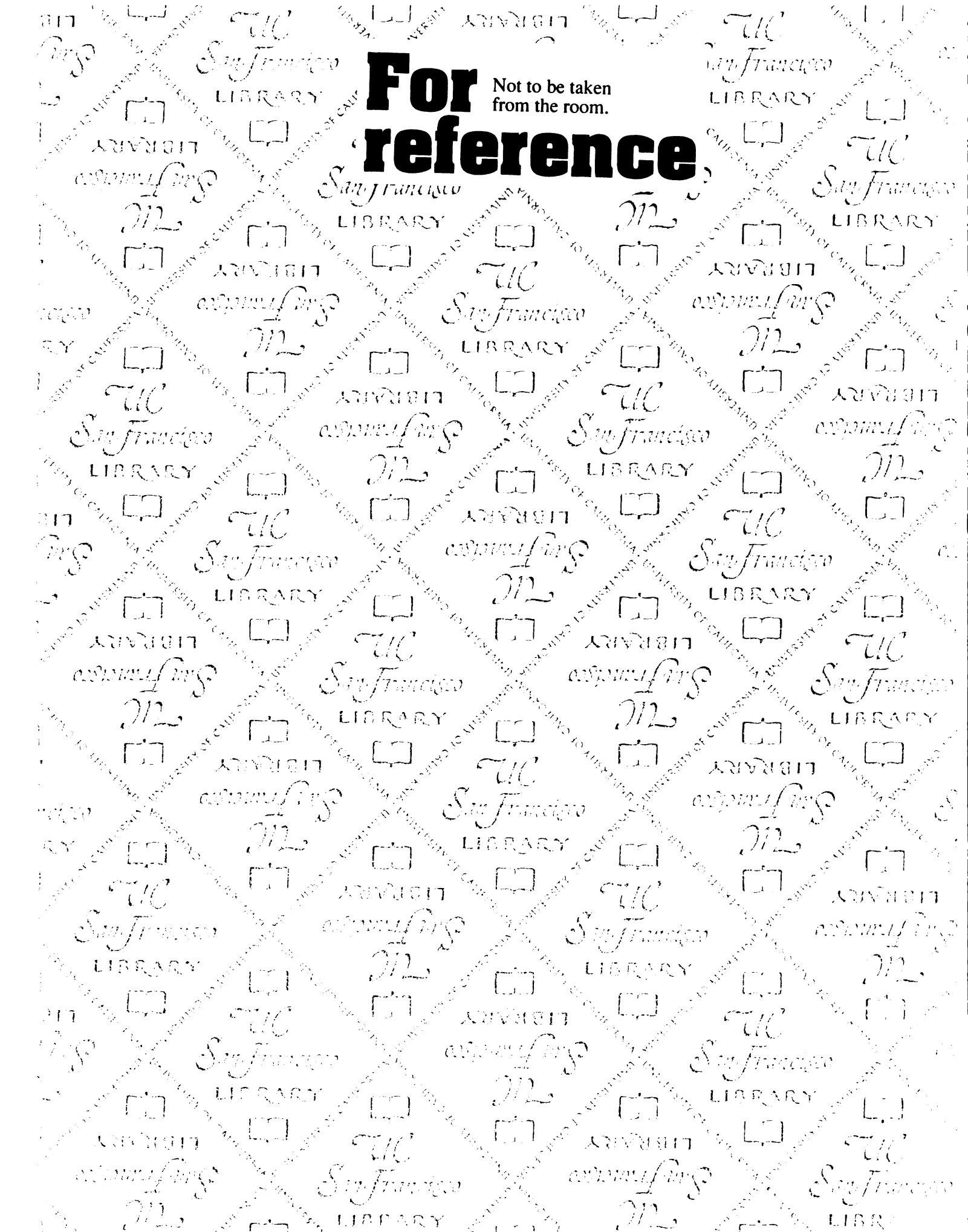
Disease Haplotypes in Affected Members of Family MAQ

Ordered STR markers spanning interval

Pedigree ID	D1S252	D1S305	GBA	RSACA	D1S303	GATA 85	LAMIN	CATCA
21	X							
211								
213								
22								
221								
23								
233	X							
234	X							
2341	X							
34								
35								
351	X	X	X					
353								

X In the table above indicates a crossover event, reflected in the fact that an allele other than the common disease allele was amplified from the subject's DNA using STR primers


The San Francisco LIBRARY


For reference
Not to be taken from the room.
UC San Francisco LIBRARY