Molecular Cloning, Chromosomal Mapping, and Characterization of the Human Cardiac–Specific Homeobox Gene *hCsx*

David Turbay,[1] Stephanie Burns Wechsler,[†] Kristina McQuate Blanchard, and Seigo Izumo[‡]

Cardiovascular Research Center and Departments of Internal Medicine, [†]Pediatrics and Communicable Diseases, and [‡]Biological Chemistry, University of Michigan Medical Center, Ann Arbor, Michigan, U.S.A.

ABSTRACT

Background: *Csx/Nkx2.5*, a murine nonclustered homeobox gene expressed primarily in the heart, has significant sequence similarity to the *Drosophila tinman* gene. *Tinman* is essential for heart and gut formation in *Drosophila*. Targeted mutation in the mouse gene, *Csx/Nkx2.5*, arrests cardiac development during early embryonic stages, suggesting an evolutionary conservation in cardiogenesis.

Materials and Methods: We have isolated and characterized a human homolog, *hCsx*, from an adult cardiac cDNA library. Northern blotting and ribonuclease protection was used to define the pattern of expression during normal development and in disease states. Chromosomal localization of the gene was determined by somatic cell hybrid analysis and fluorescent in situ hybridization.

Results: The predicted amino acid sequence of *hCsx* has 87% overall homology to the murine gene with 100% identity in the homeodomain. The homeodomain sequence of *hCsx* is 95% identical to its *Xenopus* homolog, and 65% to *tinman*. *hCsx* mRNA was detected exclusively in the heart. *hCsx* transcript was detected at 12 weeks in human embryonic heart, the earliest time point examined, and was up-regulated 5-fold between 12 and 19 weeks. There was no significant alteration of *hCsx* message level in the myocardium of 14 patients with end stage heart failure compared to a normal control. The human gene mapped to the distal portion of chromosome 5, the 5q34–q35 region. This defines a new synteny region between human chromosome 5q and the *t*-locus of mouse chromosome 17, where the mouse *Csx* gene is located.

Conclusions: *hCsx*, the human homolog of *Drosophila tinman*, is expressed in heart in a tissue restricted manner. Distal 5q trisomies produce several phenotypic abnormalities, including a high incidence of congenital heart disease. Isolation of the *hCsx* gene will allow further studies of mutations in this gene and their potential associations with some forms of congenital heart disease in humans.

INTRODUCTION

The coordinated activation of regulatory genes plays a determinant role in the developmental process. The identity of any given cell lineage is finally determined by the complement of genes activated in the correct spatiotemporal fashion. Genes encoding transcription factors are particularly important because they control the expression of downstream target genes in this process. Activation of transcription factors can initiate a genetic cascade determining the final developmental destinies of groups of cells that will become a specific organ or tissue. The discovery of the MyoD family of transcription factors was a major breakthrough for the under-

Address correspondence and reprint requests to: Seigo Izumo 7220 MSRB III, The University of Michigan Medical Center, 1150 West Medical Center Drive, Ann Arbor, MI 48109-0644, U.S.A.

[1]Present address: Division of Immunogenetics, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, U.S.A.

Molecular Medicine, Volume 2, Number 1, January 1996 86-96

standing of the genetic cascade leading to skeletal muscle development (1). These myogenic regulatory factors (MyoD, myogenin, *myf*-5 and MRF4) share a basic helix-loop-helix motif that binds to a *cis*-regulatory element called the E-box in the enhancers of different muscle specific genes (2). Skeletal and cardiac myocytes share several contractile proteins. Nevertheless, mice homozygous for null mutations in MyoD and *myf*-5 genes show a complete absence of skeletal muscle but no gross abnormality of the heart pointing to a divergent regulation between these two striated muscle cell types (3).

Homeobox genes are master regulatory genes that specify the body plan of metazoans (4). They share a common motif of 180 bp, the homeobox, which encodes the 60 amino acid homeodomain, the DNA-binding domain of this family of proteins (5). Some homeobox genes are organized in complexes (clustered), while others are randomly dispersed (non-clustered) throughout the genome (6). The clustered superfamily of homeobox genes seems to arise from a single cluster in the primitive metazoans (7). It is likely that as different organisms became genetically and anatomically more complex, duplications of the primordial genes arose. There are two clusters (*Antennapedia* and *Bithorax*) in *Drosophila* and four clusters (*HoxA* to *HoxD*) in mouse and human. These genes are organized in the same order along the chromosome as they are expressed along the antero-posterior body axis in what is known as the colinearity rule (8). Over- and mis-expression of some members of the *Hox* gene family, as well as null mutation analyses suggest that vertebrate *Hox*-type homeobox genes function in a similar way as their invertebrate homologs (8). In the non-clustered homeobox superfamily there are several mammalian homologs of each *Drosophila* gene which are named after the insect gene. They include the POU-family, paired-family, LIM-family, and NK-family, among others (6). These randomly inserted genes do not follow the colinearity rule, often contain additional conserved motifs, and may play a critical role in tissue specification during development (8).

A homeobox gene whose expression pattern is restricted to specific cell lineages during development would be of significant interest as a candidate tissue-specification gene. There has been a recent interest in the *Drosophila* NK homeobox gene family (9), as a group of genes which may specify cell fates of various tissues due to the restricted tissue expression of many of these genes. For example, the *NK-1/S59* gene is expressed in the somatic mesoderm, in anterior portions of the central nervous system, and in the midgut of *Drosophila* (10). Rodent *NK-2* homologs are expressed in a tissue restricted fashion: *mNKx2.1/TTF-1* is expressed in the thyroid anlage, the lung bud, and the developing brain; *mNKx2.2* and *mNKx2.4* are expressed in certain areas of the central nervous system; and *mNKx2.3* is expressed in the developing gut (11,12). Expression of the *NK-3/bagpipe* gene is restricted to the gut and a subset of heart progenitors in *Drosophila* (10). In the *bagpipe* mutant embryos, visceral mesoderm formation is disrupted, suggesting the requirement of *bagpipe* for gut development (13).

The NK-4/*tinman* gene is expressed in the heart precursor cells of the dorsal mesoderm (14). *Tinman* mutants do not form the dorsal vessel, the *Drosophila* equivalent of the vertebrate heart (13,15). *Tinman* has several different vertebrate relatives (16). The mouse gene, *Csx/Nkx2.5* is expressed from Day 7.5 postcoitum (p.c.) in the cardiac progenitor cells and its cardiac expression continues through the adult stage (17,18). Targeted disruption of *Csx/Nkx2.5* results in abnormal heart morphogenesis and embryonic lethality around 9–10 days p.c. (19). In null mutant embryos, the beating linear heart tube forms normally but cardiac myogenesis is arrested at 8.5 days p.c. when the looping process occurs, with subsequent death within 1–2 days (19).

It is important to point out that in addition to the homeobox genes, there are two other transcription factors, *GATA-4* and *MEF2C*, expressed in the murine precardiac mesoderm that are also candidate regulatory genes for cardiac morphogenesis (20,21). To date, the exact roles of these transcription factors in mammalian cardiac development have not been elucidated.

Due to the significant functional conservation of the *Csx*-like genes among species, it may be hypothesized that a human homolog of *tinman* exists and may be related to some forms of congenital heart disease and other cardiac pathologic states. In the present study, we report the cloning of the human homolog of the *Csx/Nkx2.5* gene, *hCsx*. Its sequence, chromosomal location, expression pattern in normal tissues and disease states are presented.

MATERIALS AND METHODS

Cardiac cDNA Library Screening and DNA Sequencing

A lambda gt10 library (a gift from Dr. Seth Alper) prepared from a single adult human left ventricle (22) was screened with a PstI fragment of mouse *Csx* containing the conserved homeobox and NK2 box (17). Low stringency hybridizations were performed with 35% formamide, 1% SDS, 2×SSC, 10% Dextran sulfate and salmon sperm DNA at 42°C. DNA sequencing reactions were carried out using the Taq DyeDeoxy Terminator or DyePrimer cycle sequencing kit (Applied Biosystems, Foster City, CA, U.S.A.). Both strands were sequenced on an automated DNA Sequenator (Applied Biosystems 370A) by the DNA Sequencing Core Facility (Ann Arbor, MI, U.S.A.). Sequence analysis was performed with MacVector and AssemblyLIGN software (Eastman Kodak, Rochester, NY, U.S.A.). Database searches were performed using the BLAST program (Whitehead Institute for Biomedical Research, Cambridge, MA, U.S.A.).

Ribonuclease Protection Analysis

We analyzed left ventricle RNA previously isolated from adult patients with end-stage heart failure of different etiologies, undergoing cardiac transplant surgery at Brigham and Women's Hospital (Boston, MA, U.S.A.) (23). The subjects included six patients with idiopathic dilated cardiomyopathy, six with ischemic cardiomyopathy, and two adults with congenital heart disease (one with L-transposition of great vessels and one with an atrial septal defect and congenital mitral stenosis, both of whom had undergone prior cardiac surgery). RNA samples of organ donors with no known cardiac pathology were used as controls. To examine developmental changes, the previously described ventricular RNA samples (23) from artificially aborted fetuses (gestation 12–19 weeks) were used. The original protocol of human cardiac tissue procurement (Dr. Paul D. Allen, Brigham and Women's Hospital) was approved by the institutional review board as previously described (23).

Ribonuclease (RNase) protection assays were performed essentially as described (24). The ^{32}P-labeled *hCsx* probe was synthesized by runoff transcription with T3 RNA polymerase from the pBluescript plasmid (Stratagene, La Jolla, CA, U.S.A.) containing the full length *hCsx* cDNA linearized with *Ava*II. The γ-actin probe was similarly synthesized with SP6 RNA polymerase from a *Hin*fI-linearized plasmid, pSP64 (a gift from Dr. Ellis J. Neufeld) containing a portion of the 3′-untranslated region of the human γ-actin gene. Five micrograms of total human ventricular RNA was used for each protection with 10 μg of yeast tRNA as a negative control in each experiment. All samples were simultaneously hybridized with *hCsx* and γ-actin probes. All hybrids were digested with RNase A and RNase T1 at 37°C for 30 min. The digested products were run on 6% denaturing gels, followed by autoradiography. Relative amounts of mRNA were determined by densitometry in the linear response range of the X-ray films for each message.

Southern and Northern Blot Analysis

Human genomic DNA was extracted from mononuclear blood cells as described previously (24), and used for Southern blotting. Ten micrograms genomic DNA samples were digested with *Bam*H1, *Eco*R1, or *Hin*dIII then electrophoresed in a 0.8% agarose gel and transferred to a Hybond-N Nylon membrane (Amersham Life Science, Arlington, IL, U.S.A.). The membranes were hybridized as previously described (25) and washed under high stringency to a final wash in 0.1×SSC, 0.5% SDS at 65°C. After this, the blots were subjected to autoradiography.

A human multiple tissue Northern blot containing approximately 2 μg of poly A$^+$ RNA per lane from different human tissues was obtained (Clontech, Palo Alto, CA, U.S.A.). The blot was hybridized and washed following the manufacturer's directions.

Chromosomal Location and Intron-Exon Mapping

Chromosomal mapping was first determined by somatic cell hybrid analysis. The Human/Rodent Somatic Cell Hybrid Panel #2 was obtained from NIGMS (Coriell Institute for Medical Research, Camden, NJ, U.S.A.). The presence of the *hCsx* gene was determined by polymerase chain reaction (PCR) performed with 200 ng hybrid genomic DNA, 500 μM each primer, 10 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM dNTPs, and 1.5 units of AmpliTaq DNA polymerase (Perkin Elmer-Cetus, Norwalk, CT, U.S.A.) in a final volume of 25 μl. The primers used were, sense (from the 3′ coding region, P1S): 5′-AACTTCGTGAACTTCGGCGTC-3′, and antisense (from the 3′ untranslated re-

gion, P1AS): 5′-GTCTCCGCAGGAGTGAATG-3′ (see Figure 2B below). After an initial denaturation step for 5 min at 95°C, the samples were amplified for 30 additional cycles at 94°C for 1 min, 60°C for 30 sec and 72°C for 30 sec. The 229-bp human specific product was visualized on a 2% agarose gel stained with ethidium bromide, subcloned into the pCR II vector (Invitrogen, San Diego, CA, U.S.A.) and sequenced.

A human genomic P1 clone containing the *hCsx* gene was obtained by screening a human foreskin fibroblast P1 library by PCR (26,27) with the primers described above. The primers used to amplify the intronic sequence of the *hCsx* P1 clone were, sense (from the 5′ coding region, P4S): 5′-TGTGCGTCTGCCTTTCC-3′ and antisense (from homeodomain, P4AS): 5′-TGCGTG GACGTGAGTTTCAG-3′ (see Fig. 2B below) following the same PCR conditions described above. The P1 plasmid containing the *hCsx* genomic clone was used as a probe and fluorescence in situ hybridization (FISH) was performed as described previously (28) by the Cancer Center Cytogenetic Core Facility (Ann Arbor, MI, U.S.A.).

RESULTS

Cloning of the Human Cardiac-Specific Homeobox (*hCsx*) cDNA

Screening of a human ventricular cDNA library using the murine *Csx* cDNA probe resulted in the isolation of five positive clones from approximately 2×10^6 recombinants. Partial sequencing and restriction mapping indicated that these clones represented a single transcript, *hCsx*. Sequence analysis (Fig. 1) revealed a single open reading frame with a putative initiation codon preceded by a strong Kozak consensus initiation sequence (29). The longest *hCsx* cDNA clone is 1585 nucleotides in length that predicts a protein of 323 amino acids with no poly(A) tail at the 3′ end. Overall, the predicted *hCsx* protein sequence has 87% identity with its mouse homolog and about 60% with the *Xenopus* homolog, *XCsx2/XNkx2.5* (Ref. 30 and Issei Komuro, Yu Chang Fu, D.T., and S.I., unpublished data). The homeodomain sequence of *hCsx* is 100% identical to its mouse homolog and 95% to its *Xenopus* homolog (Fig. 2A, a). There are two other conserved domains in the *hCsx* sequence, the TN-domain and the NK2 box, located amino- and carboxy-terminal to the homeobox, respectively. Interestingly, in the regions corresponding to the conserved

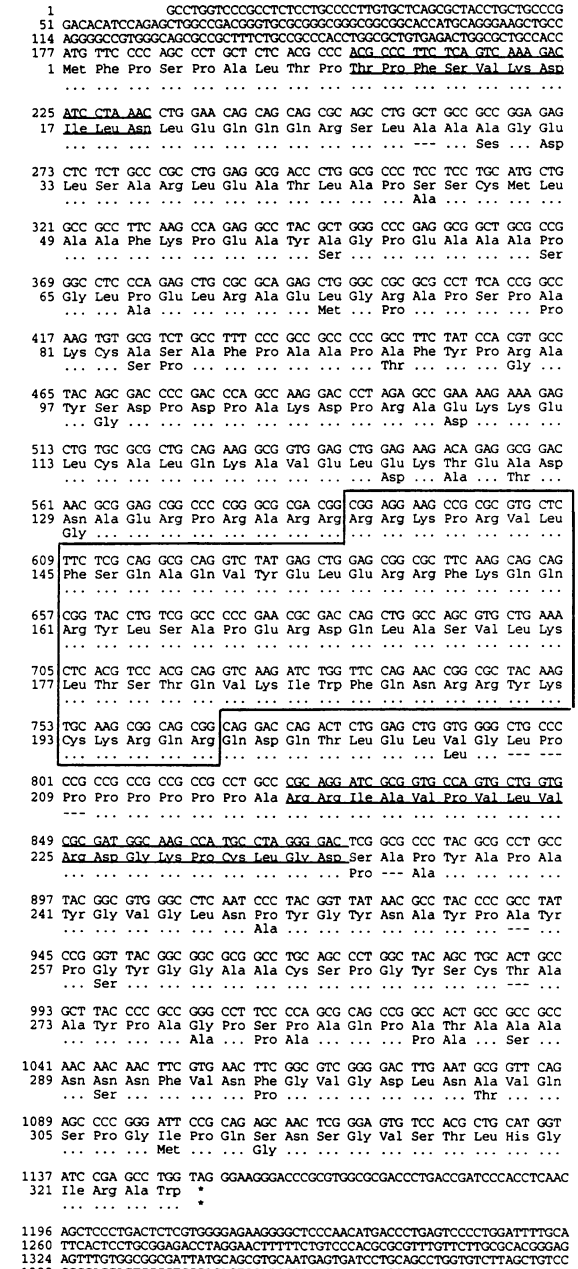

FIG. 1. Nucleotide and predicted amino acid sequence of *hCsx*

The top and middle lines indicate nucleotide and amino acid sequence of *hCsx*, respectively. The bottom line indicates identical amino acid sequence with mouse *Csx/Nkx2.5* by dots, dashes indicate gaps between the human and mouse sequences, and non-identical amino acids are noted by their three letter code. The homeodomain is boxed. The conserved amino terminus TN-domain and the NK2-box located in the carboxy terminus following the homeobox are underlined. While out-of-frame start codons are found upstream of the predicted initiation codon, no in-frame stop codons are present. This sequence has been deposited to GenBank database (Accession no. U34962).

A

a. Homeodomain

		Helix 1		*Helix 2*		*Helix 3*		**% identity**
hCsx	RRKPRVLFSQ	AQVYELERRF	KQQRYLSAPE	RDQLASVLKL	TSTQVKIWFQ	NRRYKCKRQR		
Csx/mNkx2.5	----------	----------	----------	----------	----------	----------		100%
mNkx2.3	----------	---F------	----------	-EH---S---	----------	----------		93%
Tix/mNkx2.6	Q--S------	---LA-----	-----T----	-EH---A-Q-	----------	-----S-S--		82%
TTF-1/mNkx2.1	---R------	----------	---K------	-EH---MIH-	-P--------	-H---M---A		82%
Bax/mNkx3.1	QKRS-AA--H	T--I----K-	SH-K------	-AH--KN---	-E--------	-----T--KQ		65%
XCsx2/XNkx2.5	----------	----------	---K------	--H--N----	----------	----------		95%
XCsx3/XNkx2.3	----------	---E------	----------	--H--NS---	----------	----------		93%
XCsx4/XNkx3.1	QKRS-AA--H	A--F------	NH-R---G--	-AD--AS---	-E--------	-----T--E-		63%
NK-1/S59	P-RA-TA-TY	E-LVS--NK-	-TT----VC-	-LN--LS-S-	-E--------	---T-W-K-N		55%
NK-2	K--R----TK	--T-------	R----R----	-EH---LIR-	-P--------	-H---T--AQ		72%
NK-3/bagpipe	KKRS-AA--H	---F------	A------G--	SEM--KS-R-	-E--------	-----T--KQ		67%
NK-4/tinman	K---------	---L---C--	RLKK--TGA-	EII-QK-N-S	A---------	-----S--GD		65%
Antp	-KRG-QTYTR	Y-TL---KE-	HFN---TRRR	-IEI-HA-C-	-ER-I-----	---M-W-KEN		43%
Bcd	P-RT-TT-TS	S-IA---QH-	L-G---T--R	LAD-SAK-A-	GTA------K	---RRHKI-S		45%

b. TN-domain

hCsx	TPFSVKDILN
Csx/mNKx2.5	----------
XCsx2/XNKx2.5	----------
NK4/tinman	----------
TTF-1/NKX2.1	-----S---S
NK3/bagpipe	----IN---T

c. NK2-Specific domain

hCsx	RRIAVPVLVRDGKPCLGD
Csx/mNkx2.5	----------
XCsx2/xNkx2.5	----------
Tix/mNkx2.6	--V-----L---------
XCsx3/XNkx2.3	--V----------I-G
NK2	--V-------N-------

B

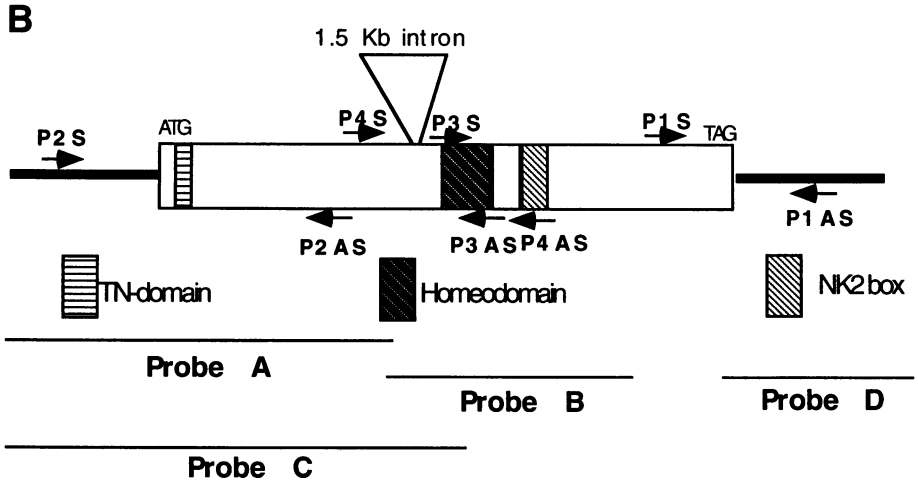

FIG. 2. *hCsx* sequence comparison and schematic diagram

(A) Interspecies sequence comparison of the *hCsx* conserved domains. a, Homeodomain; b, TN-domain; c, NK2 domain. Dashes indicate amino acid identity with *hCsx*. The three helix motifs of the homeodomain are overlined. There is 100% identity between the murine *Csx* homeodomain and the *hCsx* homeodomain. In addition, the NK-2 and TN domains are completely conserved between the human, mouse, and *Xenopus* genes. (B) Schematic diagram of the *hCsx* cDNA. Entire boxed area indicates the coding region; thick lines, untranslated regions. The three *hCsx* subdomains are indicated in the figure. The probes used in the hybridizations and ribonuclease protection assays as well as the location of PCR primers are shown in the diagram (see text for details).

TN-domain and the NK2 box, the identity is 100% among the human, mouse, and *Xenopus* genes (Fig. 2A, b and c). The *hCsx* homeodomain shows very low homology (less than 50%) to most known homeodomains with the exception of those in the NK2 gene family of *Drosophila*, *Xenopus*, and mouse.

Expression of *hCsx* in Human Tissues

To examine the tissue distribution of *hCsx* transcripts, a Northern blot containing poly A^+ RNA from several adult human tissues was examined. A band corresponding to a transcript size of approximately 1.6 kb was detected exclusively in

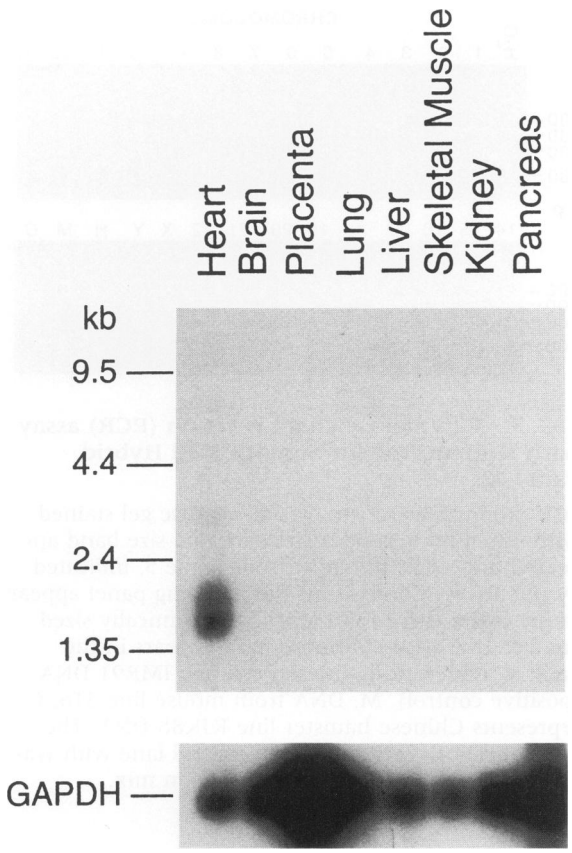

FIG. 3. Northern blot analysis of *hCsx* mRNA
The upper panel shows a human multiple tissue northern blot containing 2 μg of poly A^+ RNA from eight different adult tissues probed with a 5′ portion of the *hCsx* cDNA (nucleotides 1–608, Probe C, Fig. 2B). Hybridization signals were seen only in the heart lane. Hybridization of the blot with a PstI fragment containing the entire coding region and a part of the 3′ untranslated region of the rat glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is shown in the lower panel to verify RNA integrity (23). Unequal loading of the lanes is noted with more intense bands in the placenta and pancreas lanes.

the heart (Fig. 3). There was another lighter band in the heart lane of approximately 3 kb as well. It may correspond to unprocessed RNA of the same transcript or the product of a related gene. No signal was detected in the other adult tissues examined including brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas.

To examine the level of expression of *hCsx* in human fetal heart and in different pathologic states, RNase protection analysis was performed using a probe corresponding to the 3′ end of *hCsx* cDNA (Fig. 2B, Probe D). A human γ-actin probe corresponding to the 3′ untranslated region was used as an internal control (Fig. 4A). Densitometric analysis of *hCsx* transcript normalized by γ-actin message is shown in Fig. 4B. The *hCsx* transcript was detected at 12 weeks of embryonic development, the earliest time point examined. There was 5-fold increase in the *hCsx* message occurring between 12 and 19 weeks (Fig. 4B). To determine whether expression of *hCsx* mRNA is altered in the failing human myocardium, we analyzed left ventricular RNA samples from patients with ischemic cardiomyopathy ($n = 6$) and idiopathic dilated cardiomyopathy ($n = 6$) and two patients with congenital heart disease. No significant difference in abundance of *hCsx* message in the different pathologic states was detected. (Fig. 4 A and B).

Isolation of Genomic Clone of *hCsx*

In order to begin to characterize the genomic organization of *hCsx* gene, a human genomic P1 clone containing the *hCsx* gene was isolated by PCR as described in Materials and Methods. PCR analysis of the P1 genomic clone confirmed that it contained sequences of the cDNA from the 5′ untranslated and coding region using primer pair P2S/P2AS (Fig. 2B), and the homeodomain region using primer pair P3S/P3AS (Fig. 2B). The primers used were, P2S: 5′-CGTGGGCAGCG CCGCTTT-3′, P2AS: 5′-GTCGGGGTCGCTGTA GGCAC G-3′, P3S: 5′-CCCCGGGCGCGACGGCG GAGG-3′, P3AS: 5′-CAGAGTCTGGTCCTGCCG CTG-3′. An intronic sequence upstream of the homeobox was also amplified using primer pair P4S/P4AS (Fig. 2B). PCR reactions using primer pair P4S/P4AS with the *hCsx* cDNA template gave an expected size band (288 bp), while an approximately 1.8-kb product was detected when the *hCsx* P1 genomic clone was used as a template (data not shown).

Hybridization of a human genomic Southern blot with Probe A (containing the 5′ end of the cDNA, Probe A, Fig. 2B) and subsequently with Probe B (containing the homeobox and the NK2 box, Probe B, Fig. 2B), detected a predominant band in each lane as well as other faint bands (data not shown). These results suggest that this gene is present in a single locus in the human genome, but that there are likely other related genes (perhaps containing an *hCsx*-like homeodomain and NK2 box) that remain to be isolated in the human genome.

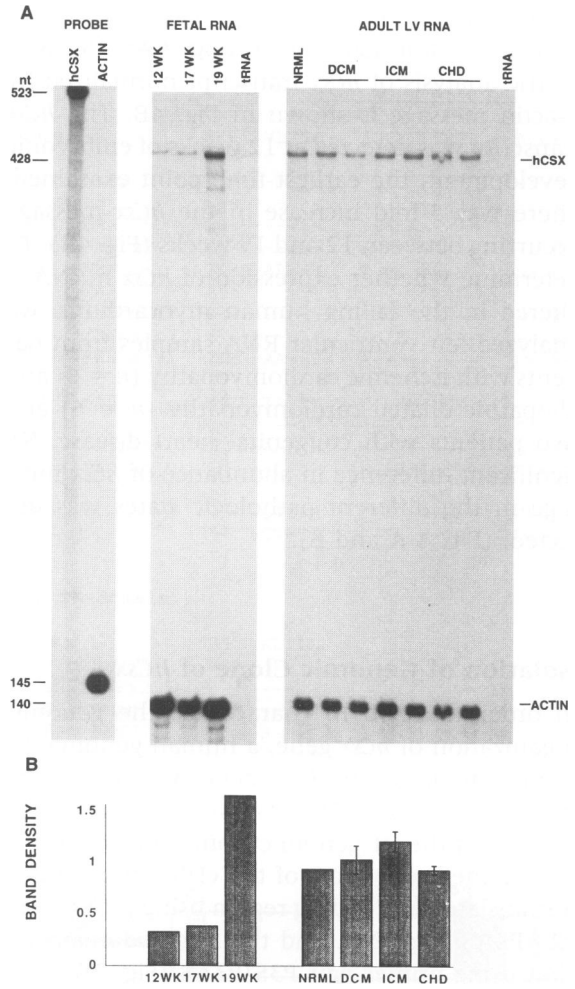

FIG. 4. Expression of *hCsx* mRNA in human fetal heart and different pathologic states

(A) Representative composite autoradiogram showing expression of *hCsx* mRNA in human myocardium from selected fetal stages and several pathologic states. Total RNA (5 μg) was used for each lane and ribonuclease protection analysis was performed using *hCsx* and gamma-actin antisense riboprobes. Yeast tRNA (10 μg) was used as a negative control. The left panel shows the undigested probes, *hCsx*, 523 nucleotides (nt) in length and γ-actin, 145 nt in length. Protected fragments of the expected sizes (*hCsx*, 428 nt; γ-actin, 140 nt) were present in all samples examined. (B) Densitometric analysis of *hCsx* transcript levels during human embryonic development and in different disease states of the LV myocardium. The relative abundance of the *hCsx* mRNA was normalized to the level of γ-actin mRNA in each sample. Densitometric score of the normal control was defined as 1.0. Data shown represents the mean ± SEM. LV, left ventricle; WK, week; NRML, normal; DCM, dilated cardiomyopathy (n = 6); ICM, ischemic cardiomyopathy (n = 6); CHD, congenital heart disease (n = 2).

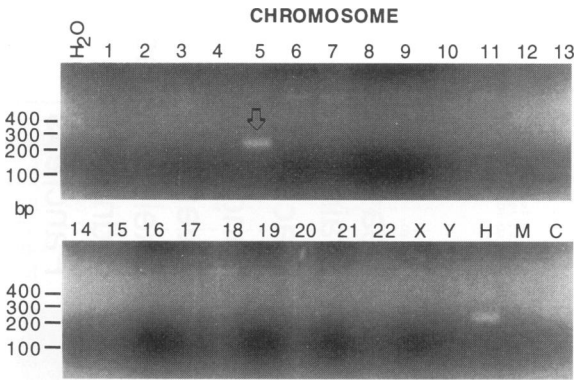

FIG. 5. Polymerase chain reaction (PCR) assay with Human/Rodent Somatic Cell Hybrid Panel #2

PCR products were run on 2% agarose gel stained with ethidium bromide. The expected size band appeared only with human chromosome 5, indicated by the arrow. Controls for the mapping panel appear in the lower three right lanes. An identically sized band to that with chromosome 5 appears in the Lane H, representing human cell line IMR91 DNA (positive control). M, DNA from mouse line 3T6, C, represents Chinese hamster line RJK88 DNA. The lane labeled H_2O is a negative control lane with water in place of template in the reaction mix.

Chromosomal Location of *hCsx* Gene

The mouse *Csx/Nkx2.5* gene maps within the *t*-locus of the mouse chromosome 17 (Ref. 31 and Issei Komuro, Neal Copeland, D.T., and S.I., unpublished data). This region of mouse chromosome 17 has been shown to have synteny homology to human chromosome 6 (32). To determine chromosomal location of the *hCsx* gene, somatic cell hybrid PCR analysis was performed using the Human/Rodent Somatic Cell Hybrid Panel #2. A band of the expected size (229 bp) was reproducibly detected in human chromosome 5 (Fig. 5). This PCR product was subcloned and sequenced, which showed 100% homology with the *hCsx* cDNA sequence.

In order to determine the chromosomal location of *hCsx* more specifically, FISH analysis was performed using the entire P1 clone as a probe. A pair of signals from both chromatids was present in the long arm of a B-group chromosome. After banding, the signal was localized in the distal portion of chromosome 5 on band q35, close to the junction with band 5q34 (Fig. 6 A–C). No signal was detected by FISH on chromosome 6, confirming the results of somatic cell hybrid PCR analysis.

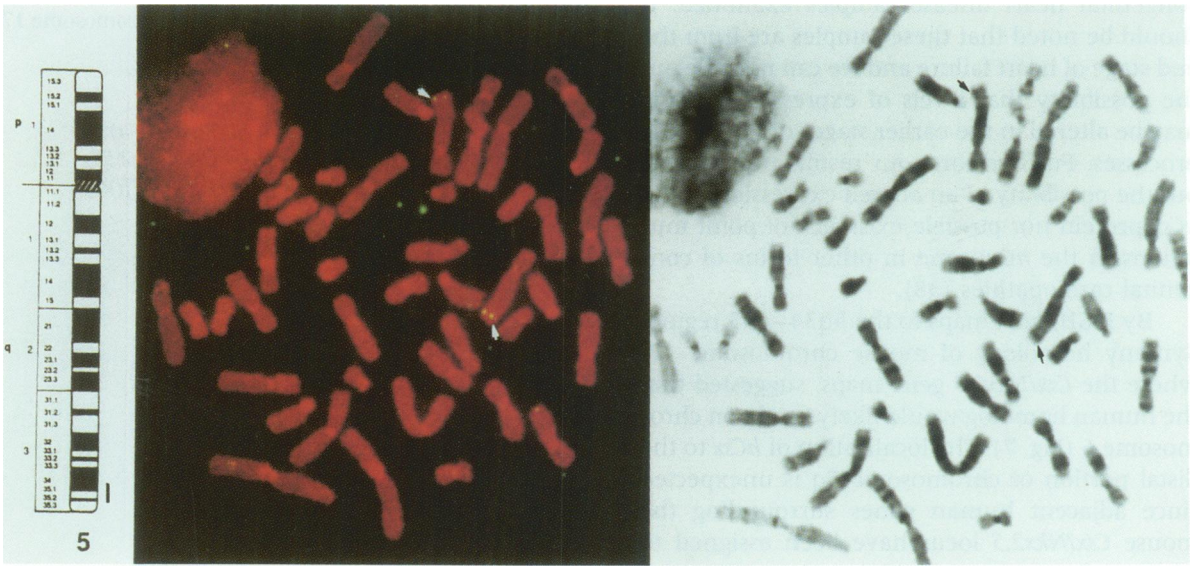

FIG. 6. Chromosomal localization of the *hCsx* gene

(A) Idiogram of human chromosome 5 showing *hCsx* localization. A vertical line shows the region where the *hCsx* gene is located, 5q34–q35. (B) Fluorescent *in situ* hybridization analysis of a metaphase diploid cell demonstrates a pair of signals (white arrows) in the distal portion of the long arm of a B-group chromosome. (C) Partial G-banding karyotype of the same slide shows that the signals (black arrows) correspond to the 5q34–35 region.

DISCUSSION

Homeobox genes play critical roles in tissue organization by regulating downstream target genes (8). As with the widely studied *Hox* genes, the NK family seems to be evolutionarily conserved and duplicates may have arisen from a single gene (6). In this report we describe the first human gene isolated to date that belongs to the NK family of homeobox genes. The homeodomain regions of the NK family members are more closely related to each other than to other homeodomains (9). While the functional significance of the additional conserved domains, the NK-2 domain and the TN domain, are not known, these domains represents another area of close relationship among these genes in several species. In adult human tissues, *hCsx* expression appears to be restricted to the heart. This tissue restricted pattern of expression is evolutionarily conserved among diverse species including *Drosophila* (15), *Xenopus* (30), mouse (17,18), and humans.

In mammals, cardiac muscle cells are derived from the anterior lateral plate mesoderm. In humans, commitment of mesodermal cells to the cardiac lineage occurs very early, when cells migrate to form the cardiogenic area at the beginning of the third week of human embryonic development (33). By the end of the 3rd week the primitive heart is formed and soon becomes functional, perfusing the different embryonic tissues and allowing normal development of the human embryo (33). We detected the *hCsx* transcript at 12 weeks, the earliest time point available. Since in the mouse, *Csx/Nkx2.5* transcript is present as early as 7.5 days p.c. (17,18), it is likely that *hCsx* transcript may have been present much earlier, perhaps as early as at the time of commitment of the cardiogenic area (around 16 to 18 days of human development) (33,34). There was a significant increase in *hCsx* transcript level between 17 and 19 weeks gestational age (Fig. 4B). At this time there is rapid growth of the heart and an associated increase in the ventricular stroke volume (35–37). In the mouse, *Csx/Nkx2.5* expression is relatively constant from late embryonic to adult stages (17). Similarly, levels of expression of *hCsx* appear relatively stable from the mid-embryonic stage to the adult; however, a more detailed study is necessary to make this a final conclusion.

Because *Drosophila tinman* and mouse *Csx/Nkx2.5* are essential for heart development, we hypothesized that expression of *hCsx* may be altered in disease states. However, there was no significant change in the level of expression of the *hCsx* transcript in human idiopathic dilated and ischemic cardiomyopathy as well as the two

congenital heart disease samples examined. It should be noted that these samples are from the end stage of heart failure and we can not rule out the possibility that levels of expression of *hCsx* may be altered in the earlier stages of the disease processes. Furthermore, our results neither rule out the possibility of an altered expression of the *hCsx* protein nor possible existence of point mutations in the *hCsx* gene in other forms of congenital cardiopathies (38).

By FISH, *hCsx* maps to the 5q34–q35 region. Synteny homology of mouse chromosome 17, where the *Csx/Nkx2.5* gene maps, suggested that the human homolog would likely reside on chromosome 6 (Fig. 7). The localization of *hCsx* to the distal portion of chromosome 5q is unexpected since adjacent human genes surrounding the mouse *Csx/Nkx2.5* locus have been assigned to human chromosome 6. However, no synteny has been studied in the specific region where the mouse *Csx/Nkx2.5* gene is located, the th20 region (31). The unambiguous assignment of *hCsx* to human chromosome 5q (Figs. 5 and 6 A–C), suggests the introduction of a novel synteny group between human chromosome 5q and mouse chromosome 17. There is another gene, *Fer*, located on human chromosome 5, which is also present on mouse chromosome 17, outside of the *t*-locus (Figure 7) (32). Alternatively, this disrupted region of homology may be due to a chromosomal rearrangement.

In the genetic cascade for heart morphogenesis, a number of genes are activated sequentially to produce proteins which are necessary for normal development. When any of these genes are altered, a phenotypic abnormality of the heart may occur, depending on where the gene is located in this epigenetic cascade (39–41). A strictly anatomical classification of congenital heart disease is evolving into a newer classification schema according to the affected developmental stage of cardiac morphogenesis (42). In general, duplications of any given chromosome portion produce a less severe phenotype than deletions (43). In the case of chromosome 5q, there are at least 25 cases reported in the literature with distal 5q trisomies. The predominant phenotype includes growth and mental retardation as well as facial abnormalities and a high incidence (at least 67%) of congenital heart disease. Among the cardiac malformations, ventricular septal defect is the most common in these patients (44). To date, no homozygous deletions have been reported in the distal portion of the chromosome 5q (43). This raises the possibility of

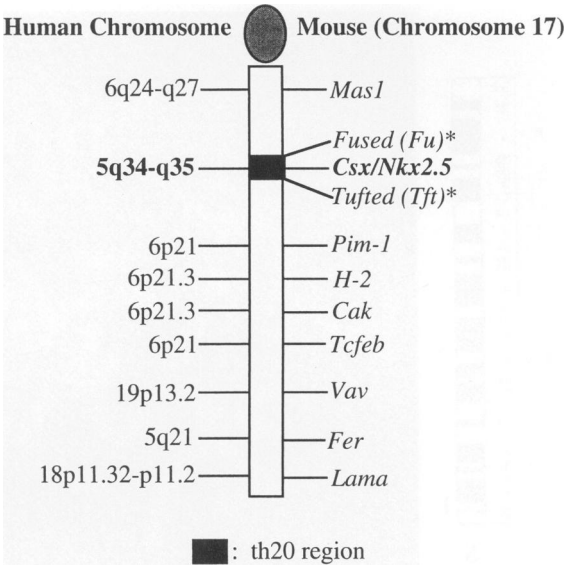

FIG. 7. Approximate composite map of mouse genes surrounding the *Csx/Nkx2.5* locus

On the left is the human chromosomal localization of the corresponding mouse genes showing the regions of mouse-human synteny homology (see text for details). *Mutated regions in which individual genes have not been cloned to date (31,32). The dark box indicates the th20 locus (see text for details).

an extremely critical role of this segment during human embryogenesis, with its loss resulting in fetal demise during the early embryonic period (45,46). In fact, a lethal spontaneous deletion of the th20 locus has been reported in the mouse; however, whether the lethality of this deletion is directly related to the loss of *Csx/Nkx2.5* or to other closely linked genes is not known (47).

In summary, we have isolated the first human member of the NK family of homeobox genes. This gene, *hCsx*, appears to be the homolog of murine *Csx/Nkx2.5* and *Drosophila tinman*. Since null mutations of *Csx/Nkx2.5* and *tinman* genes cause a severe early cardiac defect, it is possible that mutations in *hCsx* may be discovered in some forms of congenital heart disease. Future studies will test this hypothesis.

ACKNOWLEDGMENTS

We thank Ann Burgess, Madhu Prasad, and Mary Richmond for excellent technical assistance; T. Takahashi and P. D. Allen for RNA samples; and N. G. Copeland, T. W. Glover, S. A. Camper, and D. Law for helpful comments on

the manuscript. This work was supported in part by grants from the National Institutes of Health and the American Heart Association to SI.

REFERENCES

1. Weintraub H, Davis R, Tapscott S, et al. (1991) The myoD gene family: Nodal point during specification of the muscle cell lineage. *Science* **251:** 761–766.
2. Olson E, Klein WH. (1994) bHLH factors in muscle development: Deadlines and commitment, what to leave in and what to leave out. *Gene Dev.* **8:** 1–8.
3. Rudnicki MA, Schnegelsberg PNJ, Stead RH, Braun T, Arnold HH, Jaenisch R. (1993) MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75:** 1351–1359.
4. Gehring WJ. (1994) A history of the homeobox. In: Duboule D (ed). *Guidebook to the Homeobox Genes.* Oxford University Press, Oxford, pp. 3–10.
5. Gehring WJ, Qian YQ, Billeter M, et al. (1994) Homeodomain-DNA recognition. *Cell* **78:** 211–223.
6. Burglin TR. (1994) A Comprehensive classification of the homeobox genes. In: Duboule D (ed). Guidebook to the homeobox genes. Oxford University Press, Oxford, pp. 27–71.
7. Beeman RW. (1987) A homeotic gene cluster in the red flour beetle. *Science* **327:** 247–249.
8. Gehring WJ, Affolter M, Burglin T. (1994) Homeodomain proteins. *Annu. Rev. Biochem.* **63:** 487–526.
9. Kim Y, Nirenberg M. (1989) Drosophila NK-homeobox genes. *Proc. Natl. Acad. Sci. U.S.A.* **86:** 7716–7720.
10. Dohrmann C, Azpiazu N, Frasch M. (1990) A new *Drosophila* homeobox gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. *Gene Dev.* **4:** 2098–2111.
11. Duboule D. (1994) Description of the homeobox genes. In: Duboule D (ed). *Guidebook to the Homeobox Genes.* Oxford University Press, Oxford, pp. 75–246.
12. Price M, Lazzaro D, Pohl T, et al. (1992) Regional expression of the homeobox gene Nkx2.2 in the developing mammalian forebrain. *Neuron* **8:** 241–255.
13. Azpiazu N, Frasch M. (1993) Tinman and bagpipe: Two homeobox genes that determine cell fates in the dorsal mesoderm of *Drosophila. Gene Dev.* **7:** 1325–1340.
14. Bodmer R, Jan LY, Jan YN. (1990) A new homeobox-containing gene, *msh-2,* is transiently expressed early during mesoderm formation of *Drosophila. Development* **110:** 661–669.
15. Bodmer R. (1993) The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila. Development* **118:** 719–729.
16. Bodmer R. (1995) Heart development in *Drosophila* and its relation to vertebrates. *Trends Cardiovasc. Med.* **5:** 21–28.
17. Komuro I, Izumo S. (1993) *Csx*: A murine homeobox-containing gene specifically expressed in the developing heart. *Proc. Natl. Acad. Sci. U.S.A.* **90:** 8145–8149.
18. Lints TJ, Parsons LM, Hartley L, Lyons I, Harvey RP. (1993) *Nkx-2.5*: A novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119:** 419–431.
19. Lions I, Parsons LM, Hartley L, Harvey RP. (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeobox gene *Nkx2.5. Gene Dev.* **9:** 1654–1666.
20. Heikinheimo M, Scandrett JM, Wilson DB. (1994) Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Devel. Biol.* **164:** 361–373.
21. Edmondson DG, Lyons GE, Martin JF, Olson EN. (1994) *Mef2* gene expression marks the cardiac and skeletal muscle lineages during mouse embryogemesis. *Development* **120:** 1251–1263.
22. Tamkun MM, Knoth KM, Walbridge JA, Kroemer H, Roden DM, Glover DM. (1991) Molecular cloning and characterization of two voltage-gated K channel cDNA from human ventricle. *FASEB J* **5:** 331–337.
23. Takahashi T, Allen PD, Izumo S. (1992) Expression of A-, B-, and C-type natriuretic peptide genes in failing and developing human ventricles. *Circ. Res.* **71:** 9–17.
24. Sambrook J, Fritsch EF, Maniatis T. (1989) *Molecular Cloning: A Laboratory Manual.* 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Church GM, Gilbert W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **81:** 1991–1995.
26. Shepherd NS, Pfrogner BD, Coulby JN, et al. (1994) Preparation and screening of an arrayed human genomic library generated

with the P1 cloning system. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 2629–2633.
27. Pierce JC, Sauer B, Stenberg N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. *Proc. Natl. Acad. Sci. U.S.A.* **89:** 2056–2060.
28. Wechsler DS, Hawkins AL, Li X, Wang Jabs E, Griffin CA, Dang CV. (1994) Localization of the human Mxi1 transcription factor gene (*MXI1*) to chromosome 10q24–q25. *Genomics* **21:** 669–672.
29. Kozak M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acid. Res.* **15:** 8125–8132.
30. Tonissen KF, Drysdale TA, Lints TJ, Harvey RP, Krieg PA. (1994) *XNkx-2.5*, a *Xenopus* gene related to *Nkx-2.5* and *tinman*: Evidence for a conserved role in cardiac development. *Devel. Biol.* **162:** 325–328.
31. Himmelbauer H, Harvey RP, Copeland NG, Jenkins NA, Silver LM. (1994) High-resolution genetic analysis of a deletion on mouse chromosome 17 extending over the fused, tufted, and homeobox *Nkx2.5* loci. *Mamm. Genome* **5:** 814–816.
32. Forejt J, Artzt K, Barlow D, et al. (1994) Mouse chromosome 17. *Mamm. Genome* **5(Suppl):** S238–S258.
33. Moorman AFM, Lamers WH. (1994) Molecular anatomy of the developing heart. *Trends Cardiovasc. Med.* **4:** 257–264.
34. Litvin J, Montgomery M, Gonzalez-Sanchez A, Bisaha JG, Bader D. (1992) Commitment and differentiation of cardiac myocytes. *Trends Cardiovasc. Med.* **2:** 27–32.
35. John Sutton MG, Raichlen JS, Reichek N, Huff DS. (1984) Quantitative assesment of right and left ventricular growth in the human fetal heart: A pathoanatomic study. *Circulation* **70:** 935–941.
36. John Sutton MG, Gewitz MH, Shah B, et al. (1984) Quantitative assessment of growth and function of the cardiac chambers in the normal fetus: A prospective longitudinal echocardiographic study. *Circulation* **69:** 645–654.
37. Kenny JF, Plappert T, Doubilet P, et al. (1986) Changes in intracardiac blood flow velocities and right and left ventricular stroke volumes with gestational age in the normal human fetus: A prospective Doppler echocardiographic study. *Circulation* **74:** 1208–1216.
38. Dietz HCI, Pyeritz RE. (1994) Molecular genetic approaches to the study of human cardiovascular disease. *Annu. Rev. Physiol.* **56:** 763–796.
39. Nora JJ. (1993) Causes of congenital heart diseases: Old and new modes, mechanisms, and models. *Am. Heart J.* **125:** 1409–1419.
40. Bouvagnet P, Sauer U, Debrus S, et al. (1994) Deciphering the molecular genetics of congenital heart disease. *Herz* **19:** 119–125.
41. Zavala C, Jimenez D, Rubio R, Castillo-Sosa ML, Diaz-Arauzo A, Salamanca F. (1992) Isolated congenital heart defects in first degree relatives of 185 affected children. Prospective study in mexico city. *Arch. Med. Res.* **23:** 177–182.
42. Perry LW, Neill CA, Ferencz C, Rubin JD, Loffredo CA. (1994) Infants with congenital heart disease: the cases. In: Ferencz C, Rubin JD, Loffredo CA, Magee CA (eds). *Perspectives in Pediatric Cardiology, Epidemiology of Congenital Heart disease.* Futura Publishing Company, Mount Kisco, NY. Vol. 4, pp. 33–62.
43. Borgaonkar DS. (1991) *Chromosomal Variation in Man: A Catalog of Chromosomal Variants and Anomalies.* 6th Ed. Wiley-Liss, New York.
44. McDonald M, Maynard S, Sheldon S, Innis J. (1994) Unbalanced 5;16 translocation in a boy with papillary Thyroid carcinoma. *Am. J. Med. Genet.* **49:** 288–293.
45. Fryns JP, Kleczkowska A, Kubien E, Petit P, Van den Berghe H. (1984) Cytogenetic survey in couples with recurrent fetal wastage. *Hum. Genet.* **65:** 336–354.
46. Byrne J, Warburton D, Kline J, Blanc W, Stein Z. (1985) Morphology of early fetal deaths and their chromosomal characteristics. *Teratology* **32:** 297–315.
47. Lyon MF, Bechtol K. (1977) Derivation of mutant t-haplotypes of the mouse by presumed duplication or deletion. *Genet. Res.* **30:** 63–76.

Contributed by T. Yamada on October 11, 1995.