Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11579–11583, December 1992
Genetics

Identification and genetic mapping of a homeobox gene to the 4p16.1 region of human chromosome 4

(single-stranded conformational polymorphism/degenerate oligonucleotide/craniofacial)

H. S. STADLER*, B. J. PADANILAM*, K. BUETOW†, J. C. MURRAY‡, AND M. SOLURSH*

*Department of Biology, University of Iowa, and ‡Department of Pediatrics, University of Iowa Hospitals and Clinics, Iowa City, IA 52242; and †Fox Chase Cancer Center, Philadelphia, PA 19111.

Communicated by Stanley M. Gartler, August 31, 1992 (received for review June 3, 1992)

ABSTRACT **A human craniofacial cDNA library was screened with a degenerate oligonucleotide probe based on the conserved third helix of homeobox genes. From this screening, we identified a homeobox gene, *H6*, which shared only 57–65% amino acid identity to previously reported homeodomains. *H6* was physically mapped to the 4p16.1 region by using somatic cell hybrids containing specific deletions of human chromosome 4. Linkage data from a single-stranded conformational polymorphism derived from the 3′ untranslated region of the *H6* cDNA placed this homeobox gene more than 20 centimorgans proximal of the previously mapped *HOX7* gene on chromosome 4. Identity comparisons of the *H6* homeodomain with previously reported homeodomains reveal the highest identities to be with the Nk class of homeobox genes in *Drosophila melanogaster*.**

The homeobox was originally identified as a 180-base-pair (bp) region of DNA in the *Drosophila melanogaster* antennapedia (*Antp*) gene, where mutations in *Antp* resulted in transformation of antenna to leg-like body parts (1–4). DNA sequence analysis of other *Drosophila* developmental genes revealed that many of these genes, such as sex combs reduced (5), deformed (6), labial (7), and caudal (8), also contain a homeobox region and illustrates the importance of homeobox-containing genes in regulating *Drosophila* development. Similarly, homeobox genes are also present in other metazoans including *Caenorhabditis elegans* (9, 10), *Xenopus laevis* (11, 12), *Mus musculus* (13, 14), and *Homo sapiens* (15, 16).

A role for homeobox genes in mammalian development has been suggested by the association of several aberrant mouse and human phenotypes with mutations in *Pax* genes, which contain two DNA binding motifs, a paired box and a homeobox. These phenotypes include small eye (17) and splotch (18) for the mouse as well as Waardenburgs syndrome in humans (19). Additional evidence for the developmental importance of homeobox genes comes from their targeted disruption in mouse embryo-derived stem cells and subsequent use in creating transgenic mice homozygous for the mutant form of the homeobox gene (20). By this method, the developmental functions of *Hox-1.5* and *-1.6* have been determined (21, 22). In both cases, offspring homozygous for the mutant form of the gene demonstrated the loss of specific tissues and structures including thymus, parathyroid, and thyroid for *Hox1.5* and loss of outer, inner, and middle ear structures for *Hox1.6*.

Proteins encoded by homeobox genes have been shown to function as transcription factors (23) capable of binding DNA (24). Structural analysis of homeodomains indicate that these proteins form a helix–turn–helix motif (25–27), with the third helix serving as a DNA binding recognition helix (27).

Studies of the recognition helix revealed a highly conserved eight-amino acid region specific to many homeobox genes (28). This suggests that DNA probes based on the conserved region of the third helix would serve as powerful tools to screen for homeobox genes. Recently, researchers using a degenerate oligonucleotide probe based on the conserved region of the third helix identified several additional homeobox genes in *C. elegans* (9) and *M. musculus* (29).

Recognizing the developmental importance of homeobox-containing genes and the efficacy by which third-helix probes detect them, we used a similar approach to identify homeobox genes that are expressed in the developing human craniofacial region. Probing a cDNA library constructed from human embryonic craniofacial tissue, we identified a homeobox gene that maps to the 4p16.1 region of human chromosome 4.§

MATERIALS AND METHODS

Oligonucleotide Probe. An oligonucleotide probe derived from previously reported third-helix probes (9, 28) was used in all library screenings. The sequence of the probe was AARATXTGGTTYCARAAYMGXMGX, where X is A, C, T, or G; R is A or G; Y is C or T; and M is A or C.

Hybridization. The oligonucleotide was end-labeled with [γ-^{32}P]dATP (7000 Ci/mmol; 1 Ci = 37 GBq) (ICN) as described (30). Plaques were fixed to nitrocellulose filters as described (31) and hybridized overnight at 42°C. Filters were washed with tetramethylammonium chloride at 54°C to select for full-length hybridization of the oligonucleotide probe and were placed against film as described (32).

cDNA Library. A cDNA library was constructed from mRNA derived from the craniofacial region of human embryos ranging from 42 to 53 days of gestation. Approximately 1 μg of poly(A)$^+$ RNA was used as template to produce cDNA using a cDNA synthesis kit and protocols described by the manufacturer (Pharmacia). *Eco*RI/*Not* I linkers (Pharmacia) were ligated to the double-stranded cDNA product, which was in turn ligated to Lambda ZAP II vector arms (Stratagene). The average insert size of the library was 1.8 kilobases (kb). Approximately 1×10^6 plaques were detected before amplification. Nonrecombinant plaques were estimated at 5% as determined by blue/white color selection.

Clone Isolation and Sequencing. Plasmids from positive plaques were rescued from the Lambda ZAP II host by *in vivo* excision with R408 helper phage as described by the manufacturer (Stratagene). Bluescript plasmids containing the cDNA inserts were sequenced by the Sanger dideoxynucleotide method (33) using sequence-specific primers and the Sequenase version 2.0 kit (United States Biochemical). Regions of high secondary DNA structure were sequenced by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: SSCP, single-stranded conformational polymorphism.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M99587).

using either single-stranded templates or terminal transferase as described by the manufacturer (United States Biochemical). DNA sequence analysis was done using a Vax mainframe and the Genetics Computer Group (GCG) software package (34). Programs used in DNA sequence analysis included SEQED, MAP, and TRANSLATE, whereas identity comparisons between *H6* and related homeobox genes were done with the programs FASTA, PILEUP, LINEUP, and PRETTY. All nucleotide and protein data base searches used the most recent versions of the GenBank and Swiss-Prot data bases provided by the GCG software package.

PCR Conditions. A 250-bp fragment was amplified from human genomic DNA by PCR with primers derived from the 3′ untranslated region of the *H6* gene. The primer sequences were 5′-CAGCGCAACAAAGGAAAACT-3′ for the forward primer and the reverse complement of a primer from the 3′ untranslated region, 5′-GTATGGAATAAAAAGGGACA-3′. Amplifications were done with 5 ng of genomic DNA and 0.5 unit of *Taq* DNA polymerase as recommended by the manufacturer (Stratagene). PCR conditions were 40 cycles of 94°C for 55 sec, 54°C for 45 sec, and 72°C for 55 sec.

Single-Stranded Conformational Polymorphism (SSCP) Conditions. The 250-bp fragment amplified from genomic DNA demonstrated a SSCP when electrophoresed at 20 W for 3 hr in 0.5× TBE (1× TBE = 90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3). DNA samples were denatured at 85°C for 5 min in a loading buffer containing 95% formamide, 0.1% bromophenol blue, and 0.1% xylene cyanol and immediately placed on ice before loading. The gels consisted of 10% acrylamide (acrylamide/bisacrylamide, 49:1), 5% (vol/vol) glycerol, and 0.5× TBE. Gel temperatures were maintained during the 3-hr run by constantly cooling the glass plates with a small commercial house fan. Nucleotide sequencing for each allele was achieved by double-stranded sequencing of the cloned amplification product derived from homozygotes.

```
CCGATCAGCTGTCGGCGCGCACTCGCTCCCGGCCCGGCCCAGCCCAGCCCGGCGCGGAGGCCGCCGCCTGCCCTGCGGGCGCCGACGCCAGCGGTCCGGGTAGCAGCTCCAGGGCCGGCCC 121

GCGCGTGCGCCGGGAGCCGCGCCCACCATCCCCAGCGGACCGAGGAGCCCGGCCGAGCCCGAGAAGCCCGCGGCCGCGATGCCTGACGAGCTGACGGAGCCCGGGCGCGCCACGCCGGCC 241
1                                                                                                                   M  P  D  E  L  T  E  P  G  R  A  T  P  A

CGCGCCTCCTCCTTCCTCATCGAGAACCTGCTGGCGGCCGAGGCCAAGGGCGCAGGGCGCGACCCAGGGCGACGGCAGCCGGGAGGACGAGGAGGAGGACGACGACGACCCCGAAGACGA 361
15  R  A  S  S  F  L  I  E  N  L  L  A  A  E  A  K  G  A  G  R  D  P  G  R  R  Q  P  G  G  R  G  G  G  R  R  R  P  R  R  R

GGACGCCGAGCAGGCCGGCGGCGACGGCTACAGCGGCGGCGACAGTTGCTCGCGGGCACCGGGCCCGGCGGGGAGGCGCGGGCCGTGCCTGCTCGGGCGGGCGCGCTGGGCCTCGGTCCT 481
55  G  R  R  A  G  R  R  R  R  L  Q  R  R  R  Q  L  L  A  G  T  G  P  G  G  E  A  R  A  V  P  A  R  A  G  A  L  G  L  G  P

CGGCCGCCCCCCCGGTCCCGGGCCGCCCTTCGCTCTGGGCTGCGGAGGCGCAGGCCGCTGGTACCCACGGGCGCACGGTGGCTATGGAGGCGGCCTCAGTCCTGACACCAGCGACCGGGAC 601
95  R  P  P  P  G  P  G  P  P  F  A  L  G  C  G  G  A  G  R  W  Y  P  R  A  H  G  G  Y  G  G  G  L  S  P  D  T  S  D  R  D

TCACCGGAGACGGGCGAGGAGATGGGCCGTGCGGAGGGCGCCTGGCCGCGAGGCCCGGGCCGGAGCGGTGCAGCGGGAGGCAGCGGAGCTGGCGGCGCGTGGCCCGGCGGCCGGCACGGA 721
135  S  P  E  T  G  E  E  M  G  R  A  E  G  A  W  P  R  G  P  G  R  S  G  A  A  G  G  S  G  A  G  G  A  W  P  G  G  R  H  G

GGAGGCTCGGAGCTGGCCGAGGTCCCTGCGGCGGCTGGGGAGACACGCGGCGGCGTTGGCGTGGGCGGCGGCCGAAAGAAGAAGAGCGCGCACAGTCTTCTCCCGCAGCCAGGTCTTCCAG 841
175  G  G  S  E  L  A  E  V  P  A  A  A  G  E  T  R  G  G  V  G  V  G  G  G  R  K  K  K  T  R  T  V  F  S  R  S  Q  V  F  Q

                                                                                                                    AARATXTGGTTYCARAAYMGXMGX
CTGGAATCGACCTTCGACCTGAAGCGCTACCTGAGCACGGCCGAGCGCGCCGGCCTGGCCGCCTCCCTGCAGCTCACCGAGACGCAGGTTAAGATCTGGTTCCAGAACCGCCGCAACAAG 961
215  L  E  S  T  F  D  L  K  R  Y  L  S  T  A  E  R  A  G  L  A  A  S  L  Q  L  T  E  T  Q  V  K  I  W  F  Q  N  R  R  N  K

TGGAAGCGGCACGTGGCAGCCGAGCTGGAGGCGGCCAGCCTGTCCCCGCCGGGAGCGCACGCCTGGTCCGCGTGCCGGTGCTCTACCACGAAAGCCCCCCGGCCGCGGCCGCCGCTGGGC 1081
255  W  K  R  H  V  A  A  E  L  E  A  A  S  L  S  P  P  G  A  H  A  W  S  A  C  R  C  S  T  T  K  A  P  R  P  R  P  P  L  G

CCCCGGCCACCCTGCCCTTCCCGCTGGCCCGCCGGCCCGCGCCGCCCCCACCGCTGCTCGGCTTCTCCGGGGCCCTCGCCTACCCGCTGGCCGCCTTCCCGGCCGCCGCCTCCGTGCCCT 1201
295  P  R  P  P  C  P  S  R  W  P  A  G  P  R  R  P  H  R  C  S  A  S  P  G  P  S  P  T  R  W  P  P  S  R  P  P  P  P  C  P

TTCTGCGGGCGCAGATGCCTGGCCTGGTGTGAGCCCCGCCTGCCGGGCCCTCTCCCCACGACCCTGTGGACCTGTGTGGACGCGCGTTCAGCGGCAGGCGCAGGGCTCAGGGGGCGTTAG 1321
335  F  C  G  R  R  C  L  A  W  C  E  P  R  L  P  G  P  L  P  T  T  L  W  T  C  V  D  A  R  S  A  A  G  A  G  L  R  G  R  *

GGAAGGGATGGTCGCTCCTGCGGCCTCCTAGATACCTCGGGAGCGCAGGCCGCGGCCGGCGGGCCTCAGCTCCTGTGGGGAGCGCCTCTAGAATGTAATGGGACGCCCCACCCATTTGCC 1441

AGGCTGGATCCCCACTCGAACAGGGGGCCATGCAGAGACTCTGGGCTGCGCAGCCCCCGGCGCCACGGCCACCCCCCGGCCTCAGCGAGGAGCGGTCGGCAATGGCCACCCGGGGCAGCT 1561

            [CAGCGCAACAAAGGAAAACT]
GCCCTCAGCCAAGCCCAGCGCAACAAAGGAAAACTACGAACCGGCTGTCCAAGGCTGAGCGGTGACTGTCCCCACAGACTGCCCCCAACACTAAACGTCCCTTTCCTGGGACCCAGACAG 1681

                                                                                                                    C*
CAGGCCGGCCCGGACGGTGTGCTACCCCTCTCGCGGGTGCTGGTGGAGAAAGGACCCTGGACCTGTGGGTCCATCGTCCGTTCCAGGAGCAGGCAGGCTGGGGCTCTCTGCAGACGTTCG 1801

                          G*          [GTATGGAATAAAAAGGGACAT]
CAGCCTCCGGGTTGTTGTTTTTTTAAATGAATCTACTTATTTGCGTATGGAATAAAAAGGGACATTTTCT 3′                                                  1870
```

FIG. 1. Complete nucleotide and translated amino acid sequence of *H6*. Nucleotide positions 932–955 represent the exact match of the 4096-fold degenerate oligonucleotide probe. Bracketed regions represent sites for forward and reverse primers used in the SSCP and mapping analysis. Nucleotides superscripted by an asterisk represent the two transitions causing the SSCP. Homeobox region and polyadenylylation signal are underlined.

```
                5                                                    55    Identity

H6              RTVFSRSQVF QLESTFDLKR YLSTAERAGL AASLQLTETQ VKIWFQNRRN      100%
HOX7            RTPFTTAQLL ALERKFRQKQ YLSIAERAEF SSSLSLTETQ VKIWFQNRRN       60%
Mus Hox7.1      RTPFTTAQLL ALERKFRQKQ YLSIAERAEF SSSLSLTETQ VKIWFQNRRA       60%
HTR-A2          RTAYSRSQLF ELEKEFHFDK YISRPRRVEL ASSLNLTERH IKIWFQNRRM       57%
NK3             RAAFSHAQVF ELERRFAQQR YLSGPERSEM AKSLRLTETQ VKIWFQNRRY       65%
NK1             RTAFTYEQLV SLENKFKTTR YLSVCERLNL ALSLSLTETQ VKIFQNRRTK       62%
Hm40            RTAFTYEQLV ALENKFKTTR YLSVCERLNL ALSLSLTETQ VKIWFQNRRT       62%
Gallus Hox7     RTAFTSEQLL ELEKEFHCKK YLSLTERSQI AHALKLSEVQ VKIWFQNRRA       57%
Consensus       RT-F---QL- -LE--F---- YLS--ER--- --SL-LTE-Q VKIWFQNRR-       --
                   Helix 1             Helix 2              Helix 3
```

FIG. 2. Comparison of translated amino acid sequences for H6 and most closely related homeodomains. Identity scoring is based only on the homeodomain region and is relative to the H6 homeodomain. Dashes within consensus sequence represent >1 amino acid substitution at the respective position. Data are from the following sources: H6, this work; HOX7, ref. 16; Mus Hox7.1, ref. 14; HTR-A2, ref. 45; NK3 and NK1, ref. 46; Hm40, ref. 47; *Gallus* Hox7, ref. 48.

Human Chromosomal Assignment. Initial assignment of *H6* to human chromosome 4 was achieved by hybridization of radiolabeled *H6* DNA to Southern blots containing DNA derived from a mouse–human somatic cell hybrid (HA4) containing only human chromosome 4 (35).

Mapping. *H6* was assigned to the region 4p16.1 using the PCR conditions and primers described earlier to amplify the 250-bp fragment from a DNA panel of hamster–human somatic cell hybrids containing differential deletions of human chromosome 4 (36). Confirmation of this initial assignment was achieved by detecting linkage between a collection of markers on chromosome 4 and the 250-bp amplification product from *H6*, which demonstrates a SSCP when electrophoresed on a nondenaturing acrylamide gel as described (37, 38). The SSCP was typed against 50 three-generation pedigrees with large sibships (average size, 7.6) made available by the Centre d'Etude du Polymorphisme Humain as described (39). Pairwise linkage analysis was performed using the logarithm of odds (lod) score method of Morton (40) and the linkage program MLINK (41). Allele frequencies were determined by gene counting from the parents. Significance of the observed results was determined using the $Z(\theta) \geq 3$ to accept linkage and $Z(\theta) \leq -2$ to reject linkage (40). Linkage was evaluated by using a group of loci on chromosome 4 known to be on 4p (42).

Restriction fragment length polymorphisms for *H6* were also screened for using DNA from seven unrelated individuals. No restriction fragment length polymorphisms were detected with the following restriction enzymes: *Alu* I, *Bam*HI, *Bgl* I, *Bgl* II, *Bst*EI, *Bst*EII, *Dru* I, *Eco*0109, *Eco*RI, *Hae* III, *Hin*cII, *Hin*dIII, *Mbo* II, *Msp* I, *Pst* I, *Pvu* I, *Pvu* II, *Rsa* I, *Sac* I, *Sac* II, and *Taq* I.

RESULTS

Isolation and Characterization of *H6*. The degenerate oligonucleotide was used to screen $\approx 2 \times 10^5$ cDNA plaques. Twenty clones were isolated and insert DNA was fixed to Zetabind filters (AMF–Cuno) and reprobed with the degenerate oligonucleotide. Seventeen clones did not hybridize with the oligonucleotide, whereas the remaining three clones hybridized strongly. Restriction mapping and DNA sequence analysis of the positive clones indicated that the clones contained the same 1.8-kb cDNA insert and were assigned the name *H6*. The sequence of *H6* is shown in Fig. 1.

Sequence analysis of the 5′ end of *H6* cDNA revealed a high G+C content (82% for bases 1–240) and several rare-cutting restriction enzyme sites including *Not* I, *Bss*HII, and *Sma* I. Similarly, the 5′ region also contained numerous *Hpa* II restriction sites common to CpG islands, which have been found to reside near the transcription sites of vertebrate genes (43).

A single initiation codon was found in the 5′ region at position 200. This codon agrees favorably with Kozak's rules (44) and represents the only in-frame ATG in the 5′ region. Furthermore, the presence of CpG islands upstream of this ATG is consistent with its role as an initiation codon. Further analysis of the 3′ region revealed a polyadenylylation signal at positions 1851–1856; however, no poly(A) tail was found. Therefore, it is likely that some of the 3′ region of the *H6* cDNA was lost after oligo(dT) selection.

Analysis of the Translated *H6* cDNA. Translation of *H6* (Fig. 2) indicates that the cDNA encodes a complete homeodomain with conservation of the 9 invariant homeodomain-specific amino acids: Arg-5, Gln-12, Leu-16, Tyr-25, Leu-40, Trp-48, Phe-49, Asn-51, and Arg-53 (4). Amino acid comparisons of the *H6* homeodomain with the Swiss-Prot data base indicate that the highest identities are shared with several invertebrate homeobox genes including *D. melanogaster Nk-3* and *Nk-1* (46), *Helobdella triserialis HTR-A2* (45), and *Apis mellifera*

Table 1. Physical mapping of *H6* to 4p16.1

Somatic cell line	Portion of chromosome 4 contained	*H6*-specific amplification
HA4	Entire chromosome	+
HHW 416	Entire chromosome	+
HHW 986	4q35.1–4q35.2	–
HHW 582	4q25–4q35.2	–
HHW 848	4q21.1–q35.2	–
HHW 886	4p15.33–4q35.2	–
HHW 892	4p16.2–q35.2	+
HHW 842	4p16.3, 4p14–q35.2	–
HHW 693	4p16.3–p15.1	+

Physical mapping of *H6* using primers specific for the human form of *H6* and genomic DNA derived from somatic cell hybrids containing differential deletions of human chromosome 4.

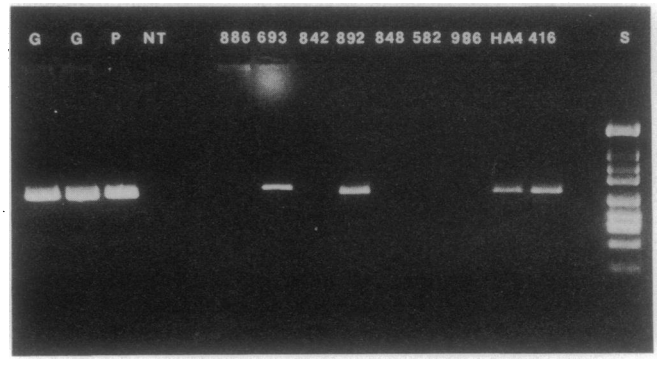

FIG. 3. The 250-bp PCR product electrophoresed on a 3% agarose gel. Lanes: G, total human genomic DNA; P, plasmid containing *H6* cDNA; NT, no template; 886, 693, 842, 892, 848, 582, 986, HA4, and 416, somatic cell hybrids containing specific portions of human chromosome 4 described in Table 1. S, pIBI31 digested with *Msp* I as a size marker.

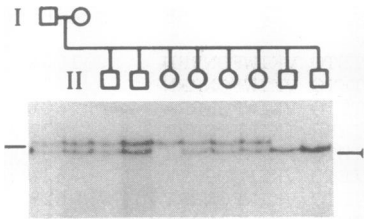

FIG. 4. *H6* SSCP demonstrated in a 10-person family. Horizontal bars indicate upper and lower bands in this two-allele system.

H40 (47) (Fig. 2). *HOX7* (16) represents the highest human and vertebrate homeodomain identity with *H6* at 60%. Interestingly, the *H6* homeodomain also demonstrates the rare substitution of threonine for arginine at position 43 as seen in *HTR-A2* (45), *D. melanogaster* labial (7), *M. musculus Hox7.1* (14), and *HOX7* (16).

Mapping and Chromosomal Localization. Somatic cell hybrid analysis and results of PCR on DNA samples from hamster–human somatic cell hybrids containing differential deletions of human chromosome 4 are shown in Table 1 and Fig. 3. Amplification products in the HA4 and HHW 416 cell lines confirm the assignment of *H6* to human chromosome 4. Products in the hybrid lines HHW 693 and HHW 892 localize the gene to the region 4p15.1–4p16.2, whereas the lack of an amplification product in either HHW 886 or HHW 842 narrows localization of *H6* to 4p16.1. A separate PCR analysis of hamster genomic DNA showed no amplification products with the primer pair used in this analysis (data not shown). DNA sequence analysis of the amplified products confirmed the amplification of the correct 250-bp region within the *H6* gene (data not shown).

Linkage to Markers on 4p. A SSCP was detected for the 250-bp PCR product derived from the 3′ untranslated region of *H6* (Fig. 4). This polymorphism consisted of two alleles, which followed Mendelian inheritance and Hardy–Weinberg distributions. Heterozygosity for the SSCP was 0.29. Linkage was detected between *H6* and both *RAF1P1* and *HOX7* (Table 2). These results confirm the initial physical localization of *H6* as both *RAF1P1* and *HOX7* have been mapped to the 4p16.1 region (16, 49). DNA sequence analysis of the PCR product from individuals homozygous for either allele identified two differences in nucleotide sequence: a T → C transition at nucleotide 1798 and an A → G transition at nucleotide 1831. Both transitions were present in one homozygous individual, whereas all other individuals homozygous for the same allele contained only the A → G transition at base 1831. This suggests that the A → G transition at base 1831 is the predominant cause of the SSCP, whereas the transition at base 1798 had no further effect on band migration under the conditions described. No linkage disequilibrium was detected between the *H6* SSCP and either a *RAF1P1 Bgl* I restriction fragment length polymorphism (49) or a *HOX7* dinucleotide CA repeat polymorphism (50).

Multipoint analysis of *H6* within the sex-averaged human chromosome 4 index map reveals its most likely localization to be between *HOX7* and *RAF1P1* (odds of 257:1); however, localization should also be expanded to encompass the region between *HOX7* and *D4S145* as exclusion from this broader domain could not be done at odds of 1000:1 (Fig. 5).

FIG. 5. Localization of *H6* within distal p arm of human chromosome 4. Crossbar indicates the most likely placement of *H6* within the sex-averaged map. Longitudinal line indicates the range of localization that could not be excluded with odds >1000:1. Distances shown are centimorgans (cMs) using Kosambi interference.

DISCUSSION

The use of oligonucleotide probes based on the conserved third helix of the homeobox represents a powerful method by which diverse classes of homeobox genes may be identified. Previous strategies for identifying vertebrate homeobox-containing genes used reduced stringency hybridizations with a particular *Drosophila* homolog. While this strategy is effective for identifying gene homologs and family members, its capacity to identify additional classes of homeobox genes is limited by its own nucleotide sequence and would not prove as efficacious in the identification of newly discovered homeobox genes.

Estimates of the number of homeobox genes in a particular species have been as high as 1% of the total number of genes (9). This suggests that many homeobox genes and their developmental roles remain uncharacterized. Consequently, the screening of tissue-specific cDNA libraries with a homeobox-specific probe can identify homeobox genes involved in development or regulation of that tissue. In this report, we used a homeobox-specific probe to screen a human craniofacial cDNA library and identified an additional homeobox gene.

Genetic mapping and DNA sequence analysis of the *H6* cDNA identified similarities to members of the Msh and Nk classes of homeobox genes. For the Msh class, *H6* was most closely related to *HOX7*, including 60% amino acid identity within the homeodomain, a G+C-rich 5′ region (51), and the rare substitution of threonine at position 43 of the homeodomain. However, the lack of significant amino acid identity between *H6* and *HOX7* genes outside the homeodomain, as well as previous reports of 5′ G+C-rich regions (52), suggests that the similarities between these two genes may be circumstantial.

Furthermore, while both *H6* and *HOX7* map to the 4p16.1 region, it is unlikely that these genes are members of a homeobox cluster. Support for this point is provided by the linkage data that places *H6* >10 centimorgans (±1 lod) proximal to the *HOX7* locus. This would correspond to ≈10

Table 2. Pairwise linkage analysis for *H6* with *RAF1P1* and *HOX7*

Comparison	0	0.001	0.05	0.10	0.20	0.30	0.40	Z_{max}	$\hat{\theta}$
H6–RAF1P1	−∞	7.81	10.66	10.0	7.8	5.07	2.10	10.55	0.05
H6–HOX7	−∞	−29.1	1.35	5.11	6.59	5.25	2.60	6.62	0.19

megabases of DNA based on an average estimate for physical and genetic distances in this region and correlates well with previous reports (51) that detected no homeobox genes within 20 kb proximal or distal to the *HOX7* locus. Similarity to members of the *Drosophila* Nk class include a 65% amino acid identity with the Nk-3 homeodomain and a 62% identity with the Nk-1 homeodomain. However, as in the case of HOX7, little amino acid identity was detected outside the homeodomain. This suggests that *H6* is an unusual homeobox gene that exhibits characteristics common to both the Msh and Nk homeobox gene families. Recently, the homeobox gene *Gsh-3* (26) was localized to a corresponding region on mouse chromosome 5. A comparison of the homeodomains for *Gsh-3*, *H6*, and *HOX7* reveals them to be distinct classes of homeobox genes; however, their chromosomal location suggests the possibility that a number of divergent homeobox genes may reside in the 4p region. Further identification of homeobox genes in this region or in the syntenic region on mouse chromosome 5 could give insight into which class of homeobox genes *H6* belongs.

The localization of *H6* to the 4p16.1 region also places it in the area associated with Wolf–Hirschorn syndrome (WHCR) (16, 53). However, the current map position of *H6* does not place it in the region most strongly associated with WHCR. Therefore, while it is unlikely that *H6* may determine the WHCR phenotype, it is still possible that mutations in *H6* may influence its etiology. Future genetic analysis including linkage studies with the mouse homolog of *H6* with previously mapped aberrant mouse phenotypes, as well as studies to determine the spatial and temporal expression patterns of this gene, should elucidate its potential role in development.

The authors wish to thank J. J. Wasmuth and M. R. Altherr for the HHW somatic cell hybrids used in the physical mapping, A. Killary for the HA4 somatic cell hybrid, T. Bürglein for his helpful suggestions on screening libraries with degenerate oligonucleotides, and D. Y. Nishimura for his assistance in primer design and mapping. This work has been supported by National Institute of Health Grants DE09170 and HG00355.

1. McGinnis, W., Levine, M. S., Haefen, E., Kuroiwa, A. & Gehring, W. J. (1984) *Nature (London)* **308**, 428–433.
2. Bender, W., Akam, M., Karch, F., Beachy, P. A., Pfeifer, M., Spierer, P., Lewis, E. B. & Hogness, D. S. (1983) *Science* **221**, 23–29.
3. Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirotta, V., Scalenghe, F. & Kaufman, T. C. (1983) *Cell* **35**, 763–776.
4. Gehring, W. J. (1987) *Science* **236**, 1245–1252.
5. Graham, A., Papalopulu, N. & Krumlauf, R. (1989) *Cell* **57**, 1424–1438.
6. Regulski, M., McGinnis, N., Chadwick, R. & McGinnis, W. (1987) *EMBO J.* **6**, 766–777.
7. Diederich, R. J., Merrill, V. K. L., Pultz, M. A. & Kaufman, T. (1989) *Genes Dev.* **3**, 399–414.
8. Duprey, P., Choudhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J. L. & Gruss, P. (1988) *Genes Dev.* **2**, 1647–1654.
9. Bürglin, T. R., Finney, M., Coulson, A. & Ruvkun, G. (1989) *Nature (London)* **341**, 239–243.
10. Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Muller, F. & Waterston, R. H. (1991) *Nature (London)* **351**, 703.
11. Ruiz i Altaba, A. & Melton, D. A. (1989) *Cell* **57**, 317–326.
12. Su, M.-W., Suzuki, H., Solursh, M. & Ramirez, F. (1991) *Development* **111**, 1179–1187.
13. Monaghan, A. P., Davidson, D. R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S. S. & Hill, R. E. (1991) *Development* **112**, 1053–1061.
14. Hill, R. E., Jones, P. F., Rees, A. R., Sime, C., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E. & Davidson, D. R. (1989) *Genes Dev.* **3**, 26–37.
15. Boncinelli, E., Acampora, S. R., Pannese, M., D'Esposito, M. & Simeone, A. (1988) *Hum. Reprod.* **3**, 880–886.
16. Ivens, A., Flavin, N., Willamson, R., Dixon, M., Bates, G., Buckingham, M. & Robert, B. (1990) *Hum. Genet.* **84**, 473–476.
17. Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. & van Heyningen, V. (1991) *Nature (London)* **354**, 522–525.
18. Epstein, D. J., Vekemans, M. & Gros, P. (1991) *Cell* **67**, 767–774.
19. Tassabehji, M., Read, A. P., Newton, V. E., Balling, R., Gruss, P. & Strachan, T. (1992) *Nature (London)* **355**, 635–638.
20. Capecchi, M. R. (1989) *Science* **244**, 1288–1292.
21. Chisaka, O. & Capecchi, M. R. (1991) *Nature (London)* **350**, 473–479.
22. Chisaka, O., Musci, T. S. & Capecchi, M. R. (1992) *Nature (London)* **355**, 516–520.
23. Ko, H.-S., Fast, P., McBride, W. & Staudt, L. M. (1988) *Cell* **55**, 135–144.
24. Mihara, H. & Kaiser, E. T. (1988) *Science* **242**, 925–927.
25. Shepherd, J. C., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. (1984) *Nature (London)* **310**, 70–71.
26. Laughon, A. & Scott, M. P. (1984) *Nature (London)* **310**, 25–31.
27. Wharton, R. P. & Ptashne, M. (1985) *Nature (London)* **316**, 601–605.
28. Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta.* **989**, 25–48.
29. Singh, G., Kaur, S., Stock, J. L., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. & Potter, S. S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10706–10710.
30. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
31. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–183.
32. Wood, W. I., Gitscher, J., Lasky, L. A. & Lawn, R. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.
33. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
34. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
35. Ning, Y., Weber, J. L., Killary, A. M., Ledbetter, D. H., Smith, J. R. & Pereira-Smith, O. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5635–5639.
36. Wasmuth, J. J., Carlock, L. R., Smith, B. & Immken, L. L. (1986) *Am. J. Hum. Genet.* **39**, 397.
37. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2766–2770.
38. Orita, M., Suzuki, Y., Sekiya, T. & Hayshi, K. (1989) *Genomics* **5**, 874–879.
39. Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J. & White, R. (1990) *Genomics* **6**, 575–577.
40. Morton, N. E. (1955) *Am. J. Hum. Genet.* **41**, 277–318.
41. Lathrop, G. M., Lalouel, J.-M., Julier, C. & Ott, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3443–3446.
42. Mills, K. A., Buetow, K. H., Xu, Y., Weber, J. L., Altherr, M. R., Wasmuth, J. J. & Murray, J. C. (1992) *Genomics* **14**, 209–219.
43. Bird, A. P. (1987) *Trends Genet.* **3**, 342–347.
44. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
45. Wedeen, C. J., Kostricken, R. G., Matsumura, I. & Weisblat, D. A. (1990) *Nucleic Acids Res.* **18**, 1908.
46. Kim, Y. & Nirenberg, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7716–7720.
47. Walldorf, U., Fleig, R. & Gehring, W. J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9971–9975.
48. Fainsod, A. & Gruenbaum, Y. (1989) *FEBS Lett.* **250**, 381–385.
49. Buetow, K. H., Shiang, R., Yang, P., Berdahl, L., Leysens, N., Nakamura, Y., Lathrop, M., Green, P., Ritty, T., Wise, M. & Murray, J. C. (1991) *Am. J. Hum. Genet.* **48**, 911–925.
50. Padanilam, B. J., Stadler, H. S., Lee, B., Ramirez, F., Murray, J. C. & Solursh, M. (1992) *Hum. Mol. Genet.* **1**, 407–410.
51. Hewitt, J. E., Clark, L. N., Ivens, A. & Williamson, R. (1991) *Genomics* **11**, 670–678.
52. Galliot, B., Dollé, P., Vigneron, M., Featherstone, M. S., Baron, A. & Duboule, D. (1989) *Development* **107**, 343–359.
53. Wilson, M. G., Towner, J. W., Coffin, G. S., Ebbin, A. J., Siris, E. & Brager, P. (1977) *Hum. Genet.* **59**, 297–307.