

PHOG, a candidate gene for involvement in the short stature of Turner syndrome

Jay W. Ellison^{1,+,*}, Zabihullah Wardak¹, Marian F. Young², Pamela Gehron Robey², Marion Laig-Webster¹ and Winston Chiong¹

¹Department of Pediatrics, University of California, San Francisco, CA 94143, USA and ²Craniofacial and Skeletal Diseases Branch, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892, USA

Received April 1, 1997; Revised and Accepted May 15, 1997

DDBJ/EMBL/GenBank accession no. U89331

The abnormalities seen in Turner syndrome (monosomy X) presumably result from haploinsufficiency of certain genes on the X chromosome. Gene dosage considerations lead to the prediction that the culpable genes escape X inactivation and have functional homologs on the Y chromosome. Among the genes with these characteristics are those residing in the pseudoautosomal regions (PAR) of the sex chromosomes. A pseudoautosomal location for a dosage-sensitive locus involved in stature has been suggested based on the analyses of patients with deletions of a specific segment of the short arm PAR; hemizygoty for this putative locus probably also contributes to the short stature in Turner individuals. We have isolated a gene from the critical deleted region that encodes a novel homeodomain-containing transcription factor and is expressed at highest levels in osteogenic cells. We have named the gene PHOG, for pseudoautosomal homeobox-containing osteogenic gene. Its deletion in patients with short stature, the predicted altered dosage in 45,X individuals, along with the nature of the encoded protein and its expression pattern, make PHOG an attractive candidate for involvement in the short stature of Turner syndrome. We have also found that the mouse homolog of PHOG is autosomal, which may help to explain the lack of a growth abnormality in mice with monosomy X.

INTRODUCTION

Turner syndrome is a common developmental disorder in females, characterized by four cardinal features: (i) short stature, (ii) ovarian failure, (iii) a variety of major and minor malformations and (iv) a high rate of embryonic lethality (1). It is classically caused by complete monosomy for the X chromosome (a 45,X karyotype; ref. 2), although certain X chromosome structural abnormalities that result in partial X monosomy can also give rise

to the Turner phenotype (1,3). As in other aneuploid disorders, the phenotypic abnormalities are most easily understood as resulting from altered dosage of certain genes. Genes expected to have an altered dose in Turner individuals are those which escape the process of X inactivation and which have functional homologs on the Y chromosome (4). Although such an idea for the pathogenesis of Turner syndrome was proposed over 30 years ago (3), no individual genes have emerged as strong candidates for involvement in the Turner phenotype. A gene pair, RPS4X/Y, has been proposed to play a role in the disorder (5), but evidence arguing against such a role has recently been reported (6).

A growing number of genes have the above-noted property of escape from X inactivation and expression from the Y chromosome (7). The prototypes of such genes are those in the pseudoautosomal regions (PAR) of the X and Y chromosomes, located at the distal ends of both the short and long arms (8). Besides the gene dosage considerations, another reason for implicating pseudoautosomal genes in Turner syndrome comes from observations of certain rare non-Turner patients who exhibit short stature. Ogata and colleagues have investigated patients having structural changes in the X and/or Y which result in deletions of portions of the short arm PAR. It was found that individuals who were hemizygous for the distal 750 kb of this region were short, while those with deletions elsewhere within the PAR were of normal stature (9,10). It was concluded that haploinsufficiency for a gene(s) in the distal deleted interval results in short stature, and these studies refined the localization of such a gene proposed a few years earlier based on a different set of X deletion patients (11).

Since virtually all Turner patients are monosomic for the short arm PAR, and since short stature is a cardinal feature of Turner syndrome, it seems reasonable to hypothesize that the dosage-sensitive locus described by Ogata *et al.* contributes to the short stature in Turner individuals. Currently no cloned genes have been assigned to the critical deleted interval, and so we have searched for expressed sequences in this region. In this paper we describe the isolation and characterization of a novel pseudoautosomal gene. Because of its location, its expression pattern, and the nature of the predicted protein, we believe it to be a very attractive candidate for involvement in the short stature phenotype of Turner syndrome. In addition we provide mapping data on the

*To whom correspondence should be addressed. Tel: +1 415 476 3386; Fax: +1 415 476 3339; Email: ellison@cgl.ucsf.edu

⁺Present address: Departments of Pediatrics and Human Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA23298, USA

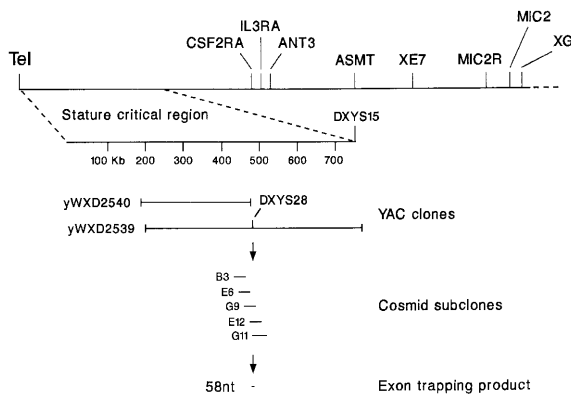


Figure 1. Strategy for cloning candidate stature genes. The top line shows a physical map of the 2.6 megabase pseudoautosomal region at the distal ends of the X and Y short arms. The positions of the known pseudoautosomal genes are indicated, according to Reid *et al.* (16). The distal 750 kb that is deleted in the set of patients with short stature is bounded proximally by the locus *DXYS15* as shown. Yeast artificial chromosome clones *yWXD2539* and *yWXD2540* are located in this region as indicated. A subset of the cosmid subclones derived from these YACs are shown; these clones form a small contig centered around the locus *DXYS28*, which has been localized to a site ~500 kb from the telomere. This cosmid contig yielded the 58 bp product from exon trapping experiments.

mouse homolog of this gene that could have important implications regarding the phenotypic differences of humans and mice with monosomy X.

RESULTS

Cloning of a new pseudoautosomal gene

A physical map of the short arm PAR is shown in Figure 1, and includes the locations of the genes currently assigned there. The region that is deleted in the patients of Ogata *et al.* (9,10) is indicated; it is bounded distally by the telomere and proximally by the locus *DXYS15*. None of the cloned pseudoautosomal genes lie within the deleted region. Yeast artificial chromosome (YAC)

clones were obtained and were determined by STS content and pulsed field electrophoresis to span most of the critical region, as indicated in the figure. The YACs were subcloned into a cosmid vector, and several small cosmid contigs were produced; one of these contigs, centered around the locus *DXYS28*, is shown in Figure 1. DNAs within each contig were pooled and used in exon trapping (12) experiments, and a single candidate exon was amplified: a 58 bp segment which encoded a portion of a homeodomain. This sequence was mapped back to the cosmid contig shown in Figure 1, and sequencing of cloned genomic DNA upstream of this trapped exon revealed another potential exon, based on an open reading frame and appropriate splice junction sequences. Primers were designed from both potential exons and were used in RT-PCR experiments with RNAs from a variety of human tissues, in order to ascertain an appropriate source of mRNA for further work.

RNA from cultured trabecular bone cells was used with the technique of 5'-RACE (13) to isolate transcribed sequences 5' of the trapped exon; the RACE subclone is labeled 5'-R in Figure 2A. Additional cDNA clones were obtained by sequential screening of a cDNA library constructed from primary cultures of bone marrow stromal fibroblasts. Over 30 independent clones were isolated from the library, and no evidence was found from the sequences of these clones to indicate alternative splicing. Several representative clones are shown in Figure 2A, which along with the 5'-RACE clone give a composite cDNA length of 4.6 kb. The complete cDNA and predicted protein sequences are presented in Figure 3. A homeobox is located in the middle of a coding region for a protein of 292 amino acids. The region downstream of the homeodomain contains a small proline-rich stretch with similarity to SH3 binding sites (14,15), believed to be important for protein-protein interactions. The coding region is flanked by large 5' and 3' untranslated regions, with the latter containing an *Alu* repeat element in the position shown in Figures 2A and 3. Characterization of cloned genomic DNA and comparison with the cDNA sequence revealed that the gene is comprised of six exons, whose organization is shown in Figure 2B. The first exon is non-coding, and the homeobox is divided

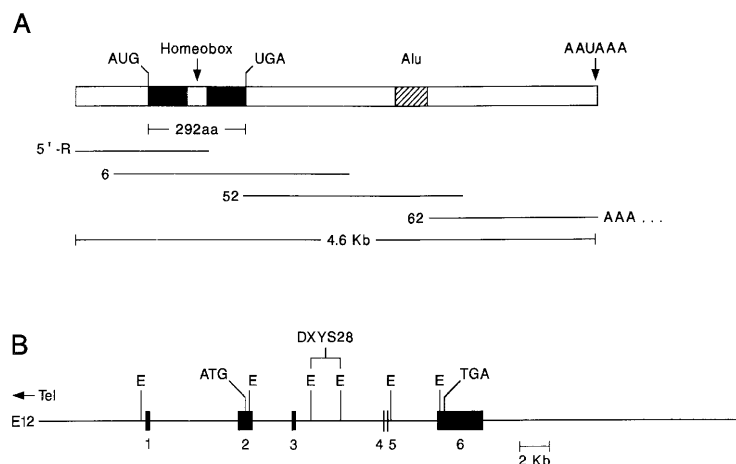


Figure 2. cDNA and genomic clones. (A) The bar figure represents the composite mRNA structure deduced from overlapping cDNA clones. The coding sequence is bounded by the AUG and UGA codons, and the position of the homeobox is indicated. An *Alu* repeat element is located in the 3' untranslated region as shown. (B) The insert of the cosmid E12 is shown, with its orientation with respect to the telomere indicated. The vertical lines marked 'E' denote *EcoRI* restriction sites. The STS *DXYS28* was found by clone blot hybridization to be located within the indicated *EcoRI* fragment. Exons 1-6 are represented by the labeled black boxes.

```

1   CGCCTGCTTTTGCCTGGTCTCGAGAACAGGGGCTCCCCACACTTTGTTTTTTTTTTGGTTTGTATTATTCGTTCCGCGCGTCTCTTTCTACTGC
101  AAACAGAAATGGGAGGGTGGACAGCGGGTAGGAGCGGATCAGACGCCAGGACGCAGCAGCCGAGTCCGCACAGGTTTGCGGGAGGTGGTGACCCGG
201  CTGGGGACGCCAGGACGCGAATGAACCTCCGGGGCAGCCTCGGGGCTCGCTCAGAGCTTGGAAAAC'GGAG'TTCTTTTCTCCGGCCACGGAGAGAA
301  CGCGGGTAACTGTGTGGGGGCTCGGGCCCTCGCCCCCTCTCGCGCGCGCTCTCCCTTCCAAAATGGGATCTTTCCCCCTTCGCACCAAGGTG
401  TACGGACGCCAAACAGTGTAGAAATGAGAAGAAAGCAATTGCCGGCTGGGGGGTGGGGGAGACACAGCGTCTCTCGTGCCTCCGCGCGGAGCCCGG
501  AGACCAGTAATTGCACCAGACAGGCAGCGCATGGGGGCTGGCGAGGTCCGCGGTATAAATAGTGAGATTTCCAATGGAAAGGCGTAAATAACAGCGC
M E E
601  TGGTGTATCCACCCGCGCACGGGCGTCTCTCCGCGGGGAGACGCGCGCATCCACCAGCCCCGGTGTCTGCCAGCCCCGGCCACGCGCATGGAAG
L T A F V S K S F D Q K S K D G N G G G G G G G K K D S I T Y R
701  AGCTCACGGCTTTTGTATCCAAGTCTTTTGACCAGAAAAGCAAGGACGGTAACGGCGGAGGCGGAGGGCGGGAGGTAAAGAGATTCCATTACGTACCG
E V L E S G L A R S R E L G T S D S S L Q D I T E G G G H C P V H
801  GGAAGTTTTGGAGAGCGGACTGGCGCGCTCCCGGAGCTGGGGAGCTCGGATTCAGCCTCCAGGACATCACGGAGGGCGGGCCACTGCCCGGTGCAT
L F K D H V D N D K E K L K E F G T A R V A E G I Y E C K E K R E D
901  TTGTTCAAGGACCAGTAGACAATGACAAGGAGAACTGAAAGAATTCGGCACCCGCGAGAGTGGCAGAAGGGATTTATGAATGCAAAGAGAAGCGCGAGG
V K S E D E D G Q T K L K Q R R S R T N F T L E Q L N E L E R L F
1001 ACGTGAAGTCGGAGGACGAGGACGGGACGCCAAGCTGAAACAGAGGGCGAGCCGACCAACTTCACGCTGGAGCAGCTGAACGAGCTCGAGCGCATCTT
D E T H Y P D A F M R E E L S Q R L G L S E A R V Q V W F Q N R R
1101 CGACGAGACCCATTACCCCGACGCTTCATGCGCGAGGAGCTCAGCCAGCGCTGGGGCTCTCCGAGGCCCGGTCAGGTTTGGTTCCAGAACCGGAGA
A K C R K Q E N Q M H K G V I L G T A N H L D A C R V A P Y V N M G
1201 GCCAAGTGGCCGCAACAAGAGAATCAGATGCATAAAGCGTCACTCTGGGCACAGCCAAACCCTAGACGCTGCCGAGTGGCACCCCTACGTCAACATGG
A L R M P F Q Q V Q A Q L Q L E G V A H A H P H L H P H L A A H A
1301 GAGCCTTACGGATGCTTTTCCAACAGGTCAGGCTCAGTGCAGCTGGAAGCGTGGCCACCGGCACCCGACCTGCACCCGACCTGGCGGCGCACCG
P Y L M F P P P P F G L P I A S L A E S A S A A A V V A A A A K S
1401 GCCCTACCTGATGTCCCCCGCGCCCTTCGGGCTGCCATCGCGTCTGGCCGAGTCCGCTCCGGCCCGCCCGTGGTTCGCGCCGCGCCGCAAAAGC
N S K N S S I A D L R L K A R K H A E A L G L END
1501 AACAGCAAGAATCCAGCATCGCCGACCTGCGGCTCAAGGCGGGAAGCACCGGAGCCCTGGGCTGTGACCCGCGCGCAGCCCCGCGCGCCGGGA
1601 CTCCCGGGCTCCGCGCACCCCGCTGCACCGCGCTCTGACCTCAACCCCGCTGGAGCTCTCCGCGGCCACCGTCTCCGGGACCCCGGAGCTC
1701 CTGCAAGAGGCTGAGGAGGGAGGCTCCCGGCACCGTCCACGCACGACGACCCAGACCTCGCGGAGATGGTGCAGAAGGCGGAGCGGGTGAAGCGCCG
1801 TGCGTCCAGCCCGGCTCTCCAAGGCTGCCGTCGCTCTGGGACCTGGAGAAGGTAACCCCGCTGGTGTGCTCTCTCTGCTATACCCATG
1901 CATCGGTTAACTACACACGTTTGAAGATCCTTAGAGTCTATTGAAACTGCAAAGATCCCGGAGCTGGTCTCCGATGAAAATGCCATTTCTTCTGTTGC
2001 CAACGATTTTCTTACTACCATGCTCCTTCTTCACTCCCGAGAGGCTGGGAACGGGTGTGGATTTGAATGTGGACTTCGGAATCCCAGGAGGCAGGGGC
2101 CGGGCTCTCTCCACCGCTCCCGGAGCCTCCAGGCAGCAATAAGGAAATAGTTCT'CTGGC'FGAGGCTGAGGACGTGAACCGCGGCTTTGGAAAGGG
2201 AGGGGAGGGAGACCCGAACCTCCACGTTGGGACTCCCACGTTCCGGGACCTGAATGAGGACCGACTTTATAACTTTTCCAGTGT'TTGAT'TCCCAAAT
2301 GGGTCTGGT'TTGT'TTGGAT'TGGT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT
2401 CGTGTGTCAAGGTGTCACTGATATGCAGCATTAACCTTACTGACATGGAGTGAAGTGAATATTTATAAATATTAAGATTAATAAATAAATAAATAAAGCCG
2501 TGCACCTTTGACCCGTCACGTCACGTCGAAAGGGT'TACCTCTTCTCCAGCGCTGGCCGCTGGCAGTGAAGGTCCTTTTGCAAAAATCACGGG
2601 TGTAGAGATGGCCCGGGCGCTGGGAGTGTGGTGTGTTTCTGAAGGGATAAAAAGAGGGCACGGTGGTGCAAGATATCAGTTTGGTACTGAGCTG
2701 TTTCTGGTTGGGAAGCGTAAAAGCCAGGAGAGATCCAGAGAGTTTCAAGTTTTCAGATGTAGGTGGTTCAGCTTTTCTTCTCCCTACTCCATC
2801 TTCTGCTTCCCCAGTCTTTTATTTCTTTGTTT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT'TT
2901 ACCGCCACCTCCGCCCCGGGTTCAAGCGATGCTCCTGCTCAGCCTCCGAGTAGCTGGGACTACAGGCACCTGCCACCACCACCCCGCCATATTTT
3001 TTGTATTTATAGTAGAGAGCGGCTTACCGTGTGGCCAGGCTCGTCTCGGAACTCCTGACCTCAGGTGATCTGCCCGCTCGGCTCCCAACGTGCC
3101 CCAGTTTTATAACAGCAGATAGCAACTTGTCTGTCACAGCTGGCATGGGCTGGACAGTGTCTTGAATGACCTAACCAAAAACATTCAGGGTTCTGCC
3201 CCAGATTTCCGGGAGATCCAGTTCATGTTCTGATTTGGTTTCTGGGAACACAGCAAGGGTTTGGTACCTCCGAGAAGATCCATCTGCATGATTTGGCA
3301 TTAGTTACCACAGCCAGCCAGAGAGAACTATCTTCTCCAACATTTACTAACATCCACTGGTCAACTCTTATTTCCATAACACATTTGCATCTTTTC
3401 TGGATCAAGCTTGGTGGT'TTCTTCTTAACTTCTGATTTAGATACTTCTCCCTCGAGGTGGGGATAAAAAGAAAAAAAACAACCTTTT'TT'TT'TT
3501 CCGCATAACACTTCTATCTTGTCACTGAGCTGAACGTAGATCCATTTCGACCCGCTCATTGTATCTTCTGATATCTTATACAAAACAAAAGTCC
3601 CTTTCAACATTTTTTATGTCAAAATGTTACAACCGCTGTAAAATGACGGAGAGAGAGAGAAAAGAAATCCAGACATTAACGGTATTAGAGAGTTGCCTCA
3701 TTCATCCATTTTCTTAAAAGCTGGAATTAATAAAGAGAGAGAGGCTTTCGGTGGACACGTTTACATTAATAAATAAATAAATAAATAAATAAATAA
3801 CTGGTGCCTAATTTATTAAGAGAATTAGCTTAGCGATGATATCTGATATTTCTCGACACACGTTGGCTAAGTTGATGCCATTTATAAATGTTTATTG
3901 AAATTTGATATTTAATGAGAAGCCGTTAAGGAATGTAGACAATATCCCGTTTCAAAGCTATGAAATGTCTATTTATTTGAAAGGGGATGTGGCTTCACG
4001 AGTTTACGCCATTTGACGTGCAGTCCCGTGGGAAGGAGGCAAAAGCCCTGCTTCTTACTTTGTGATGTATGTGGATTTGTTATTTATTTT'TT'TT
4101 CTTGCTCGGACGTTTATAAATATGTAATTTAATTTATGTCGAGTGTAAATTTGACATCGCGTTCGATTTATTTTATATTTCTGAAAACGTTTGCTTT
4201 TTCTTTTCCCTCCCCATTTGACGACATAGCGGCCCGGGTCCGGTTACAAATACATCTACAGATATTTTCAGGGATTGCTTCAGATGAAAACAAATCA
4301 CACACGTTTCCCAACCAACAGTCTTTCACATTTCTATCCCTCTGTTATTGTCCGCGAGGCGGTGAGGGGTAGAAAAAAAACAACAAACAAACAAAAA
4401 AAAACCAAAAAAACCCCTGAGTTTCTGTTGACGCTTCTCTCTTCTGTTTCAATAACTCAATGTTGAGTGCAGCAACAGACTGTATCTTTGT
4501 GACGCCCGTAGTATGAATGTACATCTTGTAAAAC'TGAGATA'PAAATAACTTATAAATTTGT 4565

```

Figure 3. PHOG cDNA sequence and deduced amino acid sequence. The mRNA synonymous strand of the composite PHOG cDNA sequence is listed, along with the predicted amino acid sequence. Nucleotide numbers are noted for the leftmost residues of each line. A number of in-frame stop codons are present upstream of the initiator ATG shown. Arrowheads mark the positions of exon-intron junctions. The homeobox and predicted amino acid residues are boxed. The proline-rich stretch in the C-terminal region of the predicted protein that has similarity to SH3 binding sites is underlined. The sequence in the 3' untranslated region from residues 2820–3098 that is underlined represents an *Alu* repeat element. The poly A addition signal is bold and underlined. This sequence has been deposited in GenBank (accession no. U89331).

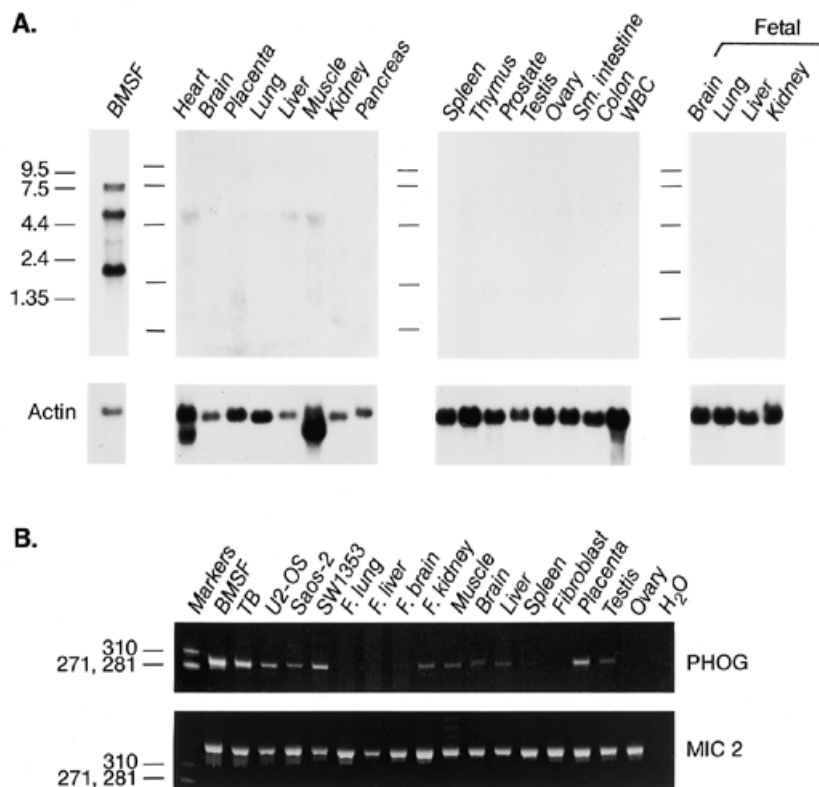


Figure 4. Expression of the PHOG gene in human tissues. (A) Autoradiogram showing the results of northern analysis of RNA from the indicated tissues, using a cDNA clone containing sequences from exons 1–4 as a hybridization probe. Sizes in kb of RNA standards are listed at the left and indicated by ticks for each filter. BMSF, bone marrow stromal fibroblasts. The bottom photo shows signals obtained by hybridizing the same filters to a β -actin probe. (B) Photograph of RT-PCR products run on a 5% polyacrylamide gel. Sizes in bp of DNA standards are listed to the left. The top photo shows results using PCR primers derived from exons 2 and 4 of the PHOG gene, while the bottom photo shows the products obtained using primers for the gene MIC2, a ubiquitously expressed pseudoautosomal gene. BMSF, bone marrow stromal fibroblasts; TB, trabecular bone cells; Saos-2 and U-2 OS are osteosarcoma cell lines; SW1353 is a chondrosarcoma cell line.

between exons 3 and 4; the latter is the exon that was trapped. The locus *DXYS28* lies within intron 3, thus placing the gene ~500 kb from the telomere (16), with the transcriptional orientation as indicated in Figure 2B.

Expression studies

The expression of the gene was examined in a variety of human fetal and adult tissues, with the results summarized in Figure 4. Northern analysis (Fig. 4A) revealed a faint band at 4.8 kb in heart, skeletal muscle and liver, but transcripts were undetectable in most tissues, even after the month-long exposure shown in the figure. In contrast, three prominent transcripts were observed in RNA from bone marrow stromal fibroblasts: besides the 4.8 kb transcript that (after accounting for a poly A tail) matches the length of the composite cDNA of Figure 2A, transcripts of 7.5 and 2.0 kb were also seen. We cannot immediately account for these transcripts based on our cDNA clones, although several possible explanations exist, including partially processed transcripts, alternative polyadenylation/splicing, and transcripts from related genes. The stromal fibroblasts are pluripotent stem cells with the capacity to differentiate into several different cell types, among them osteoblasts and chondrocytes (17). Separate cultures of these cells from different individuals all gave strong signals by northern analysis and/or RT-PCR, as did primary cultures of trabecular bone cells, which have an osteoblast-like phenotype

(18). Figure 4B shows results of RT-PCR experiments that demonstrate expression from both of these primary bone-derived cultures, as well as from two osteosarcoma cell lines, (Saos-2, U-2 OS) and a chondrosarcoma cell line (SW1353). Several other tissues showed expression by RT-PCR, though northern signals were weak or absent (compare with Fig. 4A). Thus the highest and most consistent levels of expression that we have observed is in osteogenic cells. For this reason we have named the gene PHOG, for pseudoautosomal homeobox-containing osteogenic gene.

Mapping of the mouse homolog

A search of sequence databases with the PHOG sequence revealed strong similarity to a mouse homeobox-containing gene that has very recently been cloned. An alignment of the protein sequences of PHOG and the mouse gene, *OG-12* (19) is shown in Figure 5. The homeodomains of the predicted proteins are identical, and there is substantial amino acid similarity (80%) in the C-terminal segments.

The corresponding nucleotide similarities are 81% (homeobox) and 76% (3' coding region). Upstream of the homeodomains the human and mouse proteins are quite different: other than three small blocks of near amino acid identity, the two proteins are almost completely dissimilar. In addition, both the human and mouse proteins have small polyglycine stretches, but these are in different locations in the two proteins. Nonetheless because of the

```

PHOG MEELTAFVSKSFDQKSKDGNNGGGGGGGKKSITRYEVLESGLARSRELGTSDDSLQDIT
OG-12 -----V-E --EA-----PLRGAKEPGCVPEGRDRR

E      GGG      HCPVHLFKDHVDNDKEKLE
SSPAVRAA---GGAGGGGGGGGGGGAGGGGAGGGAGGGRS-RELDMGAERSREPGRS

FGTARVAEGIYECKEKREDVKSEDEDGQTKLQRRSRTNFTLEQLNELERLFDETHYPDA
PRLTEVSPCLKDRKDDAKGM-----I-----

FMREELSQRLLGLSEARVQVWFQNRRAKCRKQENQMHHKGVILGTANHLDACRVAPYVNMGA
-----L-----LI-A-SQEE-----V--

LRMPFQQVQAQLQLEGVAHAHPLHPLAAHAPYLMFPPPPFGLPIASLAE SASAAAV
-----DSA-----P-----M---A-----L-T-AD-----S-

VAAAA KSNKNSIADRLRLKARKHAEALGL
-----AA-TT-----K---A-----
    
```

Figure 5. Comparison of the predicted protein sequence of PHOG with that of its mouse homolog, *OG-12*. The amino acid sequence of the protein encoded by PHOG is listed on the top line in one-letter code. Below that sequence is listed the sequence of the homologous mouse protein. Where the residues are identical, a dash is placed in the mouse sequence. Where differences occur, the relevant mouse residue is listed. Gaps were introduced into either sequence to maintain alignment.

overall sequence similarity, as well as an identical intron-exon organization for the two genes, there is little doubt that *OG-12* is the mouse homolog of PHOG. Indeed the differences in the N-terminal region are reminiscent of the trend of extensive divergence of human pseudoautosomal genes and their mouse homologs (20).

The mouse homologs of only two other human pseudoautosomal genes have been cloned, and these mouse genes map not to either of the sex chromosomes, but rather to chromosomes 14 (20,21) and 19 (22). We therefore set out to map the homolog of PHOG in the mouse. We did so by first designing PCR primers flanking an intronic dinucleotide repeat reported in the *OG-12* partial genomic sequence (19). These primers yielded polymorphic PCR products in *Mus domesticus* and *Mus spretus* DNAs, and we used the interspecific mapping panels from the Jackson Laboratory (23) to genetically map *OG-12* in the mouse. The gene maps to the proximal-middle portion of chromosome 3; the genetic map of the relevant interval in one of the Jackson backcross panels is shown in Figure 6. This chromosomal assignment has interesting implications not only with respect to the evolution of mammalian sex chromosomes, but also regarding differences in the phenotype of monosomy X in humans and mice, as discussed below.

DISCUSSION

The genes expected to show altered dosage in 45,X individuals is but a subset of all the genes on the sex chromosomes. These genes are the ones expressed in two doses in both sexes: from both the X and Y in males, and from both the active and inactive X in females (4). Among the growing list of such genes, most probably play no role in Turner syndrome, because one gene copy is sufficient for normal development to occur. This conclusion is based on the observation that hemizygous carriers of distal Xp deletions typically display no phenotypic abnormalities other than short stature (11,24). The various abnormalities of Turner syndrome thus likely result from haploinsufficiency for a relatively small number of genes. The identification of these dosage-sensitive genes, and the determination of which part of the phenotype they are involved in, has been difficult. This difficulty has been in part due to the phenotypic variability of the disorder

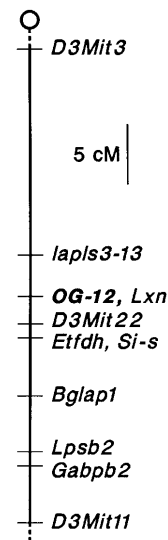


Figure 6. Genetic map from the Jackson BSS backcross showing part of chromosome 3. The map is depicted with the centromere toward the top. A 5 cM scale bar is shown to the right of the figure. Missing typings were inferred from surrounding data where assignment was ambiguous. Raw data from The Jackson Laboratory were obtained from the World Wide Web address <http://www.jax.org/resources/documents/cmdata>. Typing of *M.domesticus* and *M.spretus* *OG-12* alleles was done as described in Materials and Methods.

and the resulting lack of genotype-phenotype correlations in Turner patients with partial monosomy X. This is why we think the pseudoautosomal deletion patients studied by Ogata *et al.* (9,10) are so important; not only are they deleted for genes expected to have altered dosage in Turner syndrome, but their phenotype is a cardinal feature of the disorder. This seems more than coincidence, and it thus seems reasonable to hypothesize that the same locus that leads to short stature in these individuals when present in single dose also plays a role in the short stature of Turner individuals, though in the latter other genes may also contribute.

Since PHOG lies in the critical deleted region of the PAR, it satisfies the genotype-phenotype criteria for a candidate stature

locus. Because of its tissue-restricted expression, it has not been possible to examine the X inactivation status of PHOG in the usual manner. This is normally done by studying expression in rodent-human hybrid cell lines that retain active or inactive human X chromosomes, but since such hybrids are typically derived from fibroblasts, in which PHOG is not expressed (see Fig. 4B), such studies could not be performed. Since the gene is pseudoautosomal, one would predict that, like all other short arm pseudoautosomal genes studied to date (25–28), it would escape X inactivation, and thus have an altered dose in Turner patients. The nature of the protein encoded by PHOG (a homeodomain-containing transcription factor) and its predominant expression in osteogenic cells furthermore make PHOG a plausible candidate for involvement in linear skeletal growth. Lastly, there are precedents for dosage effects of genes encoding transcription factors (29,30), including homeobox-containing genes (31).

Taken together, the above considerations make a strong case for the hypothesis that haploinsufficiency of PHOG contributes to the short stature phenotype of Turner syndrome. The proof of such a hypothesis may be difficult, since the complete Turner phenotype (even its growth component) likely results from the interactions of a number of genes. Finding mutations of PHOG in patients with short stature would provide strong evidence for its proposed role, but given the many etiologies of short stature, such patients may be difficult to identify. Another way to support PHOG's candidacy would be to take advantage of an animal model. However, the monosomy X mouse does not have the somatic features of Turner syndrome (32). Although reduced weight gain of certain XO mice has been reported, this was attributed to a paternal X imprinting effect (33). The autosomal location that we have determined for *OG-12*, the mouse homolog of PHOG, could have an important implication: if PHOG indeed has a dose-dependent role in growth in humans and *OG-12* has a similar function in mice, then the lack of a growth abnormality in XO mice could at least in part be explained by the retention of two copies of *OG-12* in these animals. The obvious extension of this line of thinking is to examine a mouse with a single copy of the PHOG homolog; such studies have the potential not only for inferring the role of PHOG in humans, but also for generating a mouse model for part of the Turner phenotype.

MATERIALS AND METHODS

Cells and nucleic acids

Yeast artificial chromosome (YAC) clones were obtained from the Genome Center at Washington University. Multiple tissue northern blots and RNAs were purchased from Clontech. The cell lines Saos-2, U-2 OS and SW 1353 were obtained from the American Type Culture Collection. Poly A RNAs were prepared from these sources using the FastTrack kit from Invitrogen. Trabecular bone cells were prepared from pieces of normal bone under IRB approved procedures as described previously (18). Bone marrow stromal fibroblasts were cultured from normal bone marrow fragments as described (34). Total RNA was prepared from these two types of cultured cells, and poly A RNA was isolated by oligodT cellulose chromatography. DNAs from mouse interspecific backcross animals (23) were obtained from the Jackson Laboratory.

Characterization of genomic sequences

YAC clones from the distal PAR that were partially characterized by the Washington University Genome Center were further

analyzed for STS content and by pulsed field electrophoresis to rule out chimerism. Clones yWXD2539 and yWXD2540 were chosen for further experiments. Cosmid subclones with human inserts were produced from these YAC clones as described (35). Restriction mapping and hybridization to known pseudoautosomal STSs allowed the assembly of several small cosmid contigs, including the one containing PHOG that contains the locus *DXYS28*. The exon organization of PHOG was determined by restriction mapping, hybridization to different cDNA fragments, and sequencing of genomic DNA subclones.

Exon trapping

DNAs from pools of 4–6 cosmids were separately digested with *EcoRI*, *BamHI* and *BglIII*, and fragments were shotgun cloned into the exon trapping vector pSPL3 (12). Pooled plasmid DNAs were then used with the exon trapping kit from Gibco BRL, according to the manufacturer's instructions. PCR amplification products were subcloned into the vector pAMP10, and double-stranded templates were sequenced using the Sequenase kit from US Biochemicals. The sequences obtained were used to search sequence databases using BLAST.

cDNA clones

Poly A RNA (0.5 µg) from trabecular bone cells was used as a template for cDNA synthesis using the cDNA synthesis primer from Clontech's Marathon cDNA Amplification Kit. Following adaptor ligation, two rounds of PCR were performed using an antisense primer from the trapped exon, along with the nested primers provided in the kit. A 1.2 kb fragment was amplified and subcloned into the TA cloning vector pCR2.1 (Invitrogen), and was designated clone 5'-R. A 0.3 kb fragment from the 3' end of this clone was used to screen a cDNA library constructed as described (36) from bone marrow stromal fibroblast RNA. Sequential screens were performed using probes derived from 3' cDNA clones, and sequencing of these clones allowed the assembly of the cDNA contig whose sequence is presented in Figure 3.

Expression studies

Poly A RNA (2 µg) from bone marrow stromal fibroblasts was subjected to northern blot analysis as described (36). This filter strip was hybridized along with the commercial multiple tissue blots as described (36), using the 5'-R cDNA fragment. The autoradiogram shown in Figure 4A was exposed for 4 weeks with an intensifying screen. The same filters were rehybridized with a β-actin probe, and were exposed for 6 h. For RT-PCR experiments, cDNA was synthesized in 20 µl reactions using 0.5 µg poly A RNA and random hexamer primers; 1 µl of the cDNA reaction was then used in 20 µl PCR reactions using a sense primer from exon 2 (5'-GGACCACGTAGACAATGACAA-GG-3') and an antisense primer from exon 4 (5'-ACTTGGCTC-TCCGGTTCTGGAAC-3') at a concentration of 0.5 µM each. Thirty five cycles of amplification were carried out by denaturation at 94°C for 45 s, annealing at 65°C for 45 s and extension at 72°C for 1.5 min (the extension time for the first cycle was 15 min). PCR products were electrophoresed on a 5% polyacrylamide gel and photographed after ethidium bromide staining. Control PCR reactions were done using primers for the *MIC2* gene as described (37).

Genetic mapping of the mouse PHOG homolog

A dinucleotide repeat polymorphism was found within intron 2 of the *OG-12* gene which allowed us to distinguish *domesticus* and *spretus* alleles. PCR was done using primers flanking this polymorphism; the sense primer was 5'-CTGGTAATGAG-AYTGTCTTGC-3' and the antisense primer was 5'-CAGTTCA-GGGGACTGAAG-3'. PCR was done using 50 ng of DNA from each of 188 backcross animals [both the BSB and BSS backcrosses (23)], under the following conditions: initial denaturation at 94°C for 4 min, followed by 35 cycles of 94°C for 45 s, 58°C for 45 s and 72°C for 1.5 min. This resulted in the amplification of a 138 bp *M.domesticus* allele and a 171 bp *M.spretus* allele. The PCR products were run on a 10% polyacrylamide gel and visualized by ethidium bromide staining. Typing data were sent to the staff at the Jackson Laboratory for linkage analysis.

ACKNOWLEDGEMENTS

We thank the staff at the Jackson Laboratory for data analysis and assistance in figure production. We also thank C. Epstein, W. Miller, F. Chehab, E. Salido and L. Shapiro for comments on the manuscript. This work was supported by a Basil O'Connor Starter Scholar Research Award from the March of Dimes (#5-FY96-0562) to J.W.E., who was also supported in part by a Child Health Research Center grant (HD28825) to UCSF.

REFERENCES

- Lippe, B. (1991) Turner syndrome. *Endocrinol. Metab. Clin. North Am.* **20**, 121–152.
- Ford, C.E., Jones, K.W., Polani, P.E., deAlmeida, J.C., and Briggs, J.H. (1959) A sex- chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **i**, 711–713.
- Ferguson-Smith, M.A. (1965) Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. Med. Genet.* **2**, 142–155.
- Zinn, A.R., Page, D.C., and Fisher, E.M.C. (1993) Turner syndrome: the case of the missing sex chromosome. *Trends Genet.* **9**, 90–93.
- Fisher, E.M.C., Beer-Romero, P., Brown, L.G., Ridley, A., McNeil, J.A., Bentley Lawrence, J., Willard, H.F., Bieber, F.R., and Page, D. (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**, 1205–1218.
- Geerkens, C., Just, W., Held, K.R., and Vogel, W. (1996) Ullrich-Turner syndrome is not caused by haploinsufficiency of RPS4X. *Hum. Genet.* **97**, 39–44.
- Disteche, C.M. (1995) Escape from X inactivation in human and mouse. *Trends Genet.* **11**, 17–22.
- Rappold, G.A. (1993) The pseudoautosomal regions of the human sex chromosomes. *Hum. Genet.* **92**, 315–324.
- Ogata, T., Goodfellow, P., Petit, C., Aya, M., and Matsuo, N. (1992) Short stature in a girl with a terminal Xp deletion distal to DXYS15: localisation of a growth gene(s) in the pseudoautosomal region. *J. Med. Genet.* **29**, 455–459.
- Ogata, T., Yoshizawa, A., Muroya, K., Matsuo, N., Fukushima, Y., Rappold, G., and Yokoyama, S. (1995) Short stature in a girl with partial monosomy of the pseudoautosomal region distal to DXYS15: further evidence for the assignment of the critical region for a pseudoautosomal growth gene(s). *J. Med. Genet.* **32**, 831–834.
- Ballabio, A., Bardoni, B., Carrozzo, R., Andria, G., Bick, D., Campbell, L., Hamel, B., Ferguson-Smith, M.A., Ginelli, G., Fraccano, M., Maraschio, P., Zuffardi, O., Guioli, S., and Camerini, G. (1989). Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. *Proc. Natl. Acad. Sci. USA* **86**, 10001–10005.
- Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Troffater, J.A., and Buckler, A.J. (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**, 98–105.
- Chenckik, A., Moqadam, F., and Siebert, P. (1995) Marathon cDNA amplification: a new method for cloning full-length cDNAs. *CLONTECH-niques X*, 5–8.
- Ruibao, R., Mayer, B.J., Cicchetti, P., and Baltimore, D. (1993) Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**, 1157–1161.
- Gout, I., Dhand, R., Hiles, I.D., Fry, M.J., Panayotou, G., Das, P., Truong, O., Totty, N.F., Hsuan, J., Booker, G.W., Campbell, I.D. and Waterfield, M.D. (1993) The GTPase dynamin binds to and is activated by a subset of SH3 domains. *Cell* **75**, 25–36.
- Reid, K., Mertz, A., Nagajara, R., Trusgnich, M., Riley, J.H., Anand, R., Lehrach, H., Page, D.C., Ellison, J.W., and Rappold, G. (1995) Characterization of a YAC contig spanning the pseudoautosomal region. *Genomics* **29**, 787–792.
- Friedenstein, A.J., Chailakhyan, R.K., and Gerasimov, U.V. (1987) Bone marrow osteogenic stem cells: *in vitro* cultivation and transplantation in diffusion chambers. *Cell Tissue Kinet.* **20**, 263–272.
- Gehron Robey, P., and Termine, J.D. (1985) Human bone cells *in vitro*. *Calcif. Tissue Int.* **37**, 453–460.
- Rovescalli, A.C. Asoh, S., and Nirenberg, M. (1996) Cloning and characterization of four murine homeobox genes. *Proc. Natl. Acad. Sci. USA* **93**, 10691–10696.
- Ellison, J.W., Li, X., Francke, U., and Shapiro, L.J. (1996) Rapid evolution of human pseudoautosomal genes and their mouse homologs. *Mamm. Genome* **7**, 25–30.
- Miyajima, I., Levitt, L., Hara, T., Bedell, M.A., Copeland, N.G., Jenkins, N.A., and Miyajima, A. (1995) The murine interleukin-3 receptor α subunit gene: chromosomal localization, genomic structure, and promoter function. *Blood* **85**, 1246–1253.
- Disteche, C.M., Brannon, C.I., Larsen, A., Adler, D., Schorderet, D.F., Gearing, D., Copeland, N.G., Jenkins, N.A., and Park, L.S. (1992) The human pseudoautosomal GM-CSF receptor alpha subunit gene is autosomal in mouse. *Nature Genet.* **1**, 333–336.
- Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S.H., Thurston, S.J., and Birkenmeier, E.H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5**, 252–274.
- Curry, C.J., Magenis, R.E., Brown, M., Lanman, J.T. Jr., Tsai, J., O'Lague, P., Goodfellow, P., Mohandas, T., Bergner, E.A., and Shapiro, L.J. (1984) Inherited chondrodysplasia punctata due to a deletion of the terminal short arm of the X chromosome. *N. Engl. J. Med.* **311**, 1010–1015.
- Fialkow, P.J. (1970) X-chromosome inactivation and the Xg locus. *Am. J. Hum. Genet.* **22**, 460–463.
- Goodfellow, P., Pym, B., Mohandas, T., and Shapiro, L.J. (1984) The cell surface antigen locus, MIC2X, escapes X-inactivation. *Am. J. Hum. Genet.* **36**, 777–782.
- Ellison, J., Passage, M., Yu, L.-C., Yen, P., and Shapiro, L. (1992) Directed isolation of human genes that escape X inactivation. *Somat. Cell Mol. Genet.* **18**, 259–268.
- Schillel, K., Weiss, B., Wohlr, D., and Rappold, G. (1993) A human pseudoautosomal gene, ADP/ATP translocase, escapes X inactivation whereas a homologue on Xq is subject to X inactivation. *Nature Genet.* **3**, 82–87.
- Amati, B., and Land, H. (1994) Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr. Opin. Genet. Dev.* **4**, 102–108.
- Younger-Shepherd, S., Vaessin, H., Bier, E., Jan, L.Y., and Jan, Y.H. (1992) *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in Drosophila sex determination. *Cell* **70**, 911–922.
- Matsuo, I., Kuratani, S., Kimura, C., Takeda, N., and Aizawa, S. (1995) Mouse *Otx2* functions in the formation and patterning of rostral head. *Genes Dev.* **9**, 2646–2658.
- Cattanach, B.M. (1962) XO mice. *Genet. Res.* **3**, 487–490.
- Thornhill, A.R., and Burgoyne, P.S. (1993) A paternally imprinted X chromosome retards the development of the early mouse embryo. *Development* **118**, 171–174.
- Kuznetsov, S.A. and Gehron Robey, P. (1996) Growth requirements of bone marrow stromal fibroblast colony formation *in vitro*. *Calcif. Tissue Int.* **59**, 265–270.
- Carrozzo, R., Ellison, J., Yen, P., Taillon-Miller, P., Brownstein, B.H., Ballabio, A., and Shapiro, L.J. (1992). Isolation and characterization of a yeast artificial chromosome (YAC) contig around the human steroid sulfatase gene. *Genomics* **12**, 7–12.
- Ellison, J., Ramos, C., Yen, P., and Shapiro, L. (1992) Structure and expression of the human pseudoautosomal gene XE7. *Hum. Mol. Genet.* **1**, 691–696.
- Brown, C.J., Flenniken, A.M., Williams, B.R.G., and Willard, H.F. (1990) X chromosome inactivation of the human TIMP gene. *Nucleic Acids Res.* **18**, 4191–4195.