Genomics **59,** 203–212 (1999)
Article ID geno.1999.5871, available online at http://www.idealibrary.com on IDEAL®

Comparative Analysis of a Novel Gene from the Wolf–Hirschhorn/ Pitt–Rogers–Danks Syndrome Critical Region

Tracy J. Wright,* Jessica L. Costa,* Cleo Naranjo,* Phillipa Francis-West,† and Michael R. Altherr*,[1]

*Genomics Group, Life Sciences Division, MS M888, Los Alamos National Laboratory, Los Alamos, New Mexico 87545; and †Department of Craniofacial Development, UMDS, Guy's Tower, Floor 28, Guy's Hospital, London Bridge, London SE1 9RT, United Kingdom

Received January 27, 1999; accepted May 4, 1999

Wolf–Hirschhorn syndrome (WHS) is a multiple malformation syndrome characterized by mental and developmental defects resulting from the absence of a segment of one chromosome 4 short arm (4p16.3). Recently, Pitt–Rogers–Danks syndrome (PRDS), which is also due to a deletion of chromosome 4p16.3, has been shown to be allelic to WHS. Due to the complex and variable expression of these disorders, it is thought that WHS/PRDS results from a segmental aneusomy of 4p resulting in haploinsufficieny of an undefined number of genes that contribute to the phenotype. In an effort to identify genes that contribute to human development and whose absence may contribute to the phenotype associated with these syndromes, we have generated a transcript map of the 165-kb critical region and have identified a number of potential genes. One of these genes, WHSC2, which was identified with the IMAGE cDNA clone 53283, has been characterized. Sequence analysis defined an open reading frame of 1584 bp (528 amino acids), and transcript analysis detected a 2.4-kb transcript in all fetal and adult tissues tested. In parallel, the mouse homologue was isolated and characterized. Mouse sequence analysis and the pattern of expression are consistent with the clone being the murine equivalent of the human *WHSC2* gene (designated *Whsc2h*). The data from sequence and transcript analysis of this new human gene in combination with the lack of significant similarity to proteins of known function imply that it represents a novel gene. Most importantly, its location within the WHSCR suggests that this gene may play a role in the phenotype of the Wolf–Hirschhorn/Pitt–Rogers–Danks syndrome.

INTRODUCTION

The Wolf–Hirschhorn syndrome (WHS) is a multiple malformation syndrome characterized by mental and developmental defects resulting from the partial deletion of the short arm of one chromosome 4 (4p16.3). WHS patients exhibit a constellation of symptoms including severe growth deficiency, severe to profound mental retardation with onset of convulsions by the second year of life, microcephaly, sacral dimples, and characteristic facial features that include prominent glabella, hypertelorism ("Greek helmet appearance") micrognathia, highly arched eyebrows, down-turned "carp" mouth, and simple, lobeless ears. Other anomalies may include a variety of midline closure defects (cleft lip or palate, hypospadias, cryptorchidism), flexion/contracture deformities of hands and feet, skeletal defects (scoliosis, kyphosis), heart defects, hemangiomas, hypoplastic nipples, and eye defects (microphthalmia, iris defects, cataracts, strabismus) (Lurie *et al.*, 1980; Wilson *et al.*, 1981).

Pitt–Roger–Danks syndrome (PRDS) is also associated with a deletion of chromosome 4p16.3 (Pitt *et al.*, 1984; Clemens *et al.*, 1995). There is phenotypic overlap between PRDS and WHS, and it appears that the differences between these syndromes actually represent the spectrum of abnormality associated with this lost segment of chromosome 4 (Kant *et al.*, 1997; Wright *et al.*, 1998). This is supported by the identification of two patients with the same deletion (detectable at the molecular level) (Wright *et al.*, 1998). While one of these individuals had been characterized as having WHS, the other individual has been diagnosed as having PRDS.

The WHS phenotype was originally associated with cytogenetically observable alterations in the karyotype resulting from the terminal deletion of one 4p (Hirschhorn *et al.*, 1965; Wolf *et al.*, 1965). Subsequently, molecular techniques have been utilized to confirm the diagnosis of WHS where standard karyotype analysis failed to show a cryptic translocation (Altherr *et al.*,

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF101434 and AF101435.

[1] To whom correspondence should be addressed. Telephone: (505) 665-6144. Fax: (505) 665-6894. E-mail: altherr@lanl.gov.

0888-7543/99 $30.00

1991) and small terminal or interstitial deletions (Johnson *et al.*, 1994; Roulston *et al.*, 1991). These molecular analyses have been facilitated by the extensive physical maps and cloned resources available for 4p16.3 (Zuo *et al.*, 1993; Bates *et al.*, 1991; Baxendale *et al.*, 1993). A series of landmark cosmids was utilized to analyze patients using fluorescence *in situ* hybridization and to delineate the smallest region of overlap that defines the WHS critical region (WHSCR). At present the WHSCR is delimited to 165 kb (Wright *et al.*, 1997). Sequence similarity analysis of this 165-kb genomic segment using the computer program BLAST (Altschul *et al.*, 1990) identified nine cDNA clones in dbEST, representing potential transcription units (Wright *et al.*, 1997). In addition, six clusters of exons were identified using the exon prediction program XGRAIL (Uberbacher and Mural, 1991). Two of the apparently independent dbEST cDNA sequences and one or more of the GRAIL clusters have been merged and are contained within the gene *WHSC1*. *WHSC1* was the first gene to undergo extensive characterization in the WHSCR and has been described previously (Stec *et al.*, 1998).

The partial sequence of an IMAGE (integrated molecular analysis of the human genome and its expression (Auffray *et al.*, 1995)) cDNA clone, 53283, was identified in the WHS critical region transcript map and is demonstrated in this paper to represent a highly conserved gene based on its sequence similarity and expression profile in human and mouse. This gene has been named *WHSC2* (Wolf–Hirschhorn syndrome candidate 2). It was characterized by complete sequencing and transcript analysis. The mouse homologue was similarly identified by database searching and characterized using the same methods used for *WHSC2*. Comparative sequence analysis (cDNA vs genomic) of this gene in human identified 11 exons spanning a 26.2-kb region that falls entirely within the WHSCR (Wright *et al.*, 1997). Analysis of Northern blots using probes derived from *WHSC2* and the mouse homologue identified a ubiquitous 2.4-kb transcript in all tissues and developmental stages examined. Although no function has been associated with this gene, it represents a potential contributor to the WH/PRDS phenotype due to its location within the critical region.

MATERIALS AND METHODS

Sequence analysis of cDNA clones. cDNA clones were sequenced using an FS dye terminator cycle sequencing kit (Perkin–Elmer Cetus) on an ABI 373 DNA sequencing machine. Clones were sequenced initially with the vector primers M13 forward and reverse and finished using primer walking. DNA sequence was edited using the computer program ABI Sequence Analysis 2.0, and contigs were constructed from overlapping sequences using the computer program Seqman (from Lasergene DNAStar). Clones were sequenced at least once on each strand.

Sequence and polypeptide comparisons. Mouse and human sequence and polypeptide comparisons were carried out using the computer programs MegAlign and Seqman (from Lasergene DNAStar).

5′ and 3′ RACE. RACE was carried out using Marathon cDNA libraries (Clontech). Primers were chosen close to either the 5′ or the 3′ end of the available sequence. These were in combination with a primer to the cDNA adapter according to the manufacturer's instructions. Products were electophoresed on a 1–2% agarose gel according to the expected product size. Excess primers and dNTPS were removed from PCR products using exonuclease I and shrimp alkaline phosphatase. PCR products were then ligated into pCR vector (Invitrogen) and sequenced using the vector primers M13F and M13R or using PCR primers to confirm their identity.

Transcript analysis. Human and mouse Northern blots were purchased from Clontech and hybridized according to the manufacturer's instructions. Human blots contained approximately 2 μg poly(A)$^+$ RNA from a series of adult and fetal tissues. Mouse blots contained approximately 2 μg poly(A)$^+$ RNA from a series of adult tissues or embryonic stages. A *β-actin* control probe was used to ensure equal loading.

Mouse in situ hybridization. The 2.1-kb mouse cDNA clone 437082 was utilized to generate riboprobes. This cDNA was cloned between the *Not*I and the *Eco*RI sites of the vector pT3T7. Antisense probes were generated using T3 RNA polymerase using *Eco*RI linearized plasmid, and control sense riboprobes were synthesized using T7 RNA polymerase and *Not*I linearized plasmid.

Expression of *Whsc2* was analyzed by *in situ* hybridization to whole embryos or on tissue sections as described previously (Francis *et al.*, 1994). Briefly, embryos were fixed overnight in 4% paraformaldehyde and processed through a methanol series for whole-mount *in situ* hybridization or though an ethanol series into wax for *in situ* hybridization to tissue sections. Whole-mount *in situ* hybridization using digoxigenin-labeled RNA probes and *in situ* hybridization to tissue sections using ^{35}S-UTP-labeled RNA probes were performed as described previously (FrancisWest *et al.*, 1995). For radioactive *in situ* hybridizations, sections were cut at 8 μm and probed for the expression of *Whsc2* or *shh* using antisense riboprobes generated. Background levels of hybridization were determined by probing with the sense *Whsc2* riboprobe as a control.

RESULTS

A transcript map of the 165-kb WHSCR identified a number of putative genes in dbEST by sequence similarity (Wright *et al.*, 1997). One of these was defined by the cDNA clone, 53283, from the IMAGE consortium collection. Comparing the genomic sequence to the cDNA revealed three exons at the 5′ end and a single exon at the 3′ end (Wright *et al.*, 1997). Consequently, it was decided to analyze this clone further to determine whether it is a protein encoding gene and to evaluate the potential role of that gene in the Wolf–Hirschhorn syndrome.

Sequence Analysis of WHSC2

The cDNA clone 53283 was obtained from Research Genetics and verified by DNA sequencing using the vector primers M13F and M13R. The remainder of the clone was sequenced using a primer walking strategy. Both strands were sequenced at least once. Comparison of the sequence and predicted polypeptides against GenBank detected no significant similarities to any known gene or protein at the date of submission. Analysis of the sequence from cDNA clone 53283 using MapDraw (Lasergene DNAStar) detected an open reading frame of 1492 bp. No long open reading frames were detected in any of the other frames.

```
+1
ATGGCGTCCATGCGGGAGAGCGACACGGGCCTGTGGCTGCACAACAAGCTGGGGGCCACGGACGAGCTGTGGGCGCCGCCCAGCATCGCGTCCCTGCTCA
 M  A  S  M  R  E  S  D  T  G  L  W  L  H  N  K  L  G  A  T  D  E  L  W  A  P  P  S  I  A  S  L  L

CGGCCGCGGTCATCGACAACATCCGTCTCTGCTTCCATGGCCTCTCGTCGGCAGTGAAGCTCAAGTTGCTACTCGGGACGCTGCACCTCCCGCGCCGCAC
T  A  A  V  I  D  N  I  R  L  C  F  H  G  L  S  S  A  V  K  L  K  L  L  L  G  T  L  H  L  P  R  R  T

GGTGGACGAGATGAAGGGCGCCCTAATGGAGATCATCCAGCTCGCCAGCCTCGACTCGGACCCCTGGGTGCTCATGGTCGCCGACATCTTGAAGTCCTTT
  V  D  E  M  K  G  A  L  M  E  I  I  Q  L  A  S  L  D  S  D  P  W  V  L  M  V  A  D  I  L  K  S  F

CCGGACACAGGCTCGCTTAACCTGGAGCTGGAGGAGCAGAATCCCAACGTTCAGGATATTTTGGGAGAACTTAGAGAAAAGGTGGGTGAGTGTGAAGCGT
 P  D  T  G  S  L  N  L  E  L  E  E  Q  N  P  N  V  Q  D  I  L  G  E  L  R  E  K  V  G  E  C  E  A

CTGCCATGCTGCCACTGGAGTGCCAGTACTTGAACAAAAACGCCCTGACGACCCTCGCGGGACCCCTCACTCCCCCGGTGAAGCATTTTCAGTTAAAGCG
S  A  M  L  P  L  E  C  Q  Y  L  N  K  N  A  L  T  T  L  A  G  P  L  T  P  P  V  K  H  F  Q  L  K  R

GAAACCCAAGAGCGCCACGCTGCGGGCGGAGCTGCTGCAGAAGTCCACGGAGACCGCCCAGCAGTTGAAGCGGAGCGCCGGGGTGCCCTTCCACGCCAAG
  K  P  K  S  A  T  L  R  A  E  L  L  Q  K  S  T  E  T  A  Q  Q  L  K  R  S  A  G  V  P  F  H  A  K

GGCCGGGGGCTGCTGCGGAAGATGGACACCACCACCCCACTCAAAGGCATCCCGAAGCAGGCGCCCTTCAGAAGCCCCACGGCGCCCAGCGTCTTCAGCC
 G  R  G  L  L  R  K  M  D  T  T  T  P  L  K  G  I  P  K  Q  A  P  F  R  S  P  T  A  P  S  V  F  S

CCACAGGGAACCGGACCCCCATCCCGCCTTCCAGGACGCTGCTGCGGAAGGAACGAGGTGTGAAGCTGCTGGACATCTCTGAGCTGGATATGGTTGGCGC
P  T  G  N  R  T  P  I  P  P  S  R  T  L  L  R  K  E  R  G  V  K  L  L  D  I  S  E  L  D  M  V  G  A

TGGCCGAGAGGCGAAGCGGAGAAGGAAGACTCTCGATGCGGAGGTGGTGGAGAAGCCGGCCAAGGAGGAAACGGTGGTGGAGAACGCCACCCCGGACTAC
  G  R  E  A  K  R  R  R  K  T  L  D  A  E  V  V  E  K  P  A  K  E  E  T  V  V  E  N  A  T  P  D  Y

GCAGCCGGCCTGGTGTCCACGCAGAAACTTGGGTCCCTGAACAATGAGCCTGCGCTGCCCTCCACGAGCTACCTTCCCTCCACGCCCAGCGTGGTTCCCG
 A  A  G  L  V  S  T  Q  K  L  G  S  L  N  N  E  P  A  L  P  S  T  S  Y  L  P  S  T  P  S  V  V  P

CCTCCTCCTACATCCCCAGCTCCGAGACGCCCCCAGCCCCATCTTCCCGGGAAGCCAGCCGCCCACCAGAGGAGCCCAGCGCCCCGAGCCCCACGTTGCC
A  S  S  Y  I  P  S  S  E  T  P  P  A  P  S  S  R  E  A  S  R  P  P  E  E  P  S  A  P  S  P  T  L  P

AGCGCAGTTCAAGCAGCGGGCGCCCATGTACAACAGCGGCCTGAGCCCTGCCACACCCACGCCTGCGGCGCCCACCTCGCCTCTGACACCCACCACACCT
  A  Q  F  K  Q  R  A  P  M  Y  N  S  G  L  S  P  A  T  P  T  P  A  A  P  T  S  P  L  T  P  T  T  P

CCGGCTGTCGCCCCTACCACTCAGACACCCCCGGTTGCCATGGTGGCCCCGCAGACCCAGGCCCCTGCTCAGCAGCAGCCTAAGAAGAATCTGTCCCTCA
 P  A  V  A  P  T  T  Q  T  P  P  V  A  M  V  A  P  Q  T  Q  A  P  A  Q  Q  Q  P  K  K  N  L  S  L

CGAGAGAGCAGATGTTCGCTGCCCAGGAGATGTTCAAGACGGCCAACAAAGTCACGCGGCCCGAGAAGGCCCTCATCCTGGGCTTCATGGCCGGCTCCCG
T  R  E  Q  M  F  A  A  Q  E  M  F  K  T  A  N  K  V  T  R  P  E  K  A  L  I  L  G  F  M  A  G  S  R

AGAGAACCCGTGCCAGGAGCAGGGGGACGTGATCCAGATCAAGCTGAGCGAGCACACGGAGGACCTGCCCAAGGCGGACGGCCAGGGTAGCACAACCATG
  E  N  P  C  Q  E  Q  G  D  V  I  Q  I  K  L  S  E  H  T  E  D  L  P  K  A  D  G  Q  G  S  T  T  M

CTGGTGGACACAGTGTTTGAGATGAACTATGCCACGGGCCAGTGGACGCGCTTCAAGAAGTACAAGCCCATGACCAATGTGTCCTAGaaccacctgcctc
 L  V  D  T  V  F  E  M  N  Y  A  T  G  Q  W  T  R  F  K  K  Y  K  P  M  T  N  V  S  .

acagctggccgtcacttgtgggggtccacgggacgatggctttgccagcttaaagtaaccggatggcggacacctggcccccgaggtcccccggccgccg
ccctgctgctgacccagcctgttttaagttctggatgcatttctctggggtatttggggcttattttaaaattttaatatgggttctttttttgtgtgat
ttaaagacactttttggactcaacgttacatttttgaatgtagtaagtaaattaaccaaaaaagttacaacttcctaattttagtgacagctctgcctgt
ttgttagactcttacttttaaaatcttttctatttccctcgctggggcagtgccctcctacccccagggttgaggggaccaaggtggcacggtggtac
tgggggtgcggcagggacacccgaccacaccagagcgtgggagacggtgggccttgtcccctgcctgtgcctgcctgggagttttgtattcatcttttgt
atagttgtggacatttaagacagtctttgggtacctattttcattgtaaaactatctgaaccattaaagtcgagcttttctaaagaaaaaaaaaaaaaaa
```

FIG. 1. Sequence, and predicted protein, of *WHSC2* is shown. Sequence numbering begins at A in ATG and is given the designation bp +1. Coding sequence is shown in uppercase, and 3′ UTR is shown in lowercase. Location of the variant polyadenylation signal ATTAAA is shown by a box.

The 5′ End of WHSC2

Evidence obtained from sequencing of the cDNA clone suggested that the cDNA present in clone 53283 is incomplete. The large open reading frame identified in this clone runs directly to the cloning site. Therefore, 5′ RACE was carried out to extend the clone toward the translation and transcription start sites. 5′ RACE was carried out using a fetal brain cDNA library with a primer located at bp 102–123 of cDNA clone 53283 (Fig. 1). The products identified using 5′ RACE were cloned and sequenced using vector derived primers. These clones yielded an additional 100 bp of sequence information including a putative tranalational start codon in the appropriate context (Kozak, 1996) (Fig. 1). The A in the putative start codon was designated +1; this coordinate will be used throughout the remainder of the results. Only 5 bp of additional sequence was identified 5′ to this putative start codon. The total available sequence following this experiment was 2.2 kb. To determine whether the sequence immediately 5′ to the suspected ATG start codon could contain a transcription start site, the 500 bp of genomic sequence obtained from the cosmid sequence and located 5′ to the ATG was analyzed using the program NNPP/Eukaryotic (Reese and Eeckman, 1995). A putative transcriptional start site was identified at −11 with a probability of 0.98.

The 3′ End of WHSC2

A TAG termination codon was found at bp 1585–1587, yielding an open reading frame of 1584 bp (528 amino acids). The 3′ untranslated region has 596 bp of nucleotide sequence that is 52% AT-rich. A polyadenylation signal was found at bp 2161–2155; this is followed 17 bp downstream by the poly(A) tail. The polyadenylation signal, ATTAAA, is a variant of the canonical AATAAA sequence. However, systematic mutagenesis of the canonical AATAAA has shown that ATTAAA is a tolerated alternative and the most frequent variant of the perfect hexanucleotide sequence (Wahle and Keller, 1992). This variant polyadenylation signal is found in approximately 10% of transcripts (Colgan and Manley, 1997).

Genomic Structure of WHSC2

The intron/exon structure of *WHSC2* was determined by comparing the cDNA sequence with the genomic sequence using the computer program BLAST (Altschul *et al.*, 1990). The genomic sequence of this region was generated by the Sanger Centre, United Kingdom, from a 2.2-Mb cosmid contig of the region (Baxendale *et al.*, 1993). Comparison of the sequence of *WHSC2* with the available genomic sequence showed that it was contained entirely within cosmid 96a2 (Fig. 2). In addition, a base substitution was identified within the coding region, and a single basepair variation resulting from an insertion/deletion (indel) was detected in the 3′ UTR. The base substitution is found in exon 9 and is a transition from T (cDNA) to C (genomic). This change does not alter the amino acid codon. The indel identified in the 3′ UTR is an A insertion in cDNA 53283 sequence when compared to the genomic sequence. Comparison with the genomic sequence showed that *WHSC2* was transcribed in the centromeric to telomeric direction. The 2.2-kb of available sequence from *WHSC2* consists of 11 exons and 10 introns that are distributed over a 26.2-kb genomic region. The organization of the gene is depicted in Fig. 2. A total of 2205 bp of sequence has been deposited with the EMBL/GenBank Data Libraries under Accession No. AF101434. The size of each exon and intron and the sequence of the exon/intron boundary are shown in Table 1. All of the 5′ splice donor and 3′ splice acceptor sites conform to the GT/AG consensus. The intron sizes varied between 88 bp and 17 kb and are located between exons 5 and 6 and between exons 1 and 2, respectively (Table 1). The exon sizes range from 69 bp (exon 6) to 781 bp (exon 11). The suspected start codon was found within exon 1, and the termination codon was found in exon 11 (Fig. 2).

Transcript Analysis of WHSC2

To analyze the gene expression pattern of *WHSC2*, Northern blots containing poly(A)$^+$ RNA from human tissues were hybridized with cDNA clone 53283. Results from the hybridization of cDNA clone 53283 to poly(A)$^+$ RNA from a series of adult and fetal tissues show that WHSC2 is ubiquitously expressed (Fig. 3). The transcript is approximately 2.4 kb in size. It appears that there is a reduced level of expression in the adult lung relative to the other tissues (Fig. 3, left). The reduction in signal was not due to unequal loading but likely represents a real difference in the level of expression in the adult lung. This was confirmed following hybridization with a control probe (*β-actin*) and was repeated on three separate Northern blots. All of the results were consistent with the original findings. In contrast to the result in the adult lung RNA, the apparent underexpression of *WHSC2* in fetal liver is due to an underloading of RNA when compared to the level of a control probe (Figure 3, right).

Isolation of a Mouse Homologue to WHSC2

Sequence from *WHSC2* was used to search dbEST for related mouse clones. Two IMAGE cDNA clones, 437082 and 354189, were identified. The inserts from these clones were sequenced on both strands using a primer walking approach and shown to be overlapping. A total of 2133 bp of sequence has been deposited with the EMBL/GenBank Data Libraries under Accession No. AF101435. cDNA clone 354189 contained a poly(A) tail that had been primed from an internal poly(A) sequence. No other sequence differences were detected between cDNA clones 354189 and 437082. Comparison of the sequence of these clones to *WHSC2* showed an 84.2% identity at the nucleotide sequence level and a 93.3% identity at the amino acid level across the predicted coding region (Fig. 4). An open reading frame was present from the beginning of the cDNA clones to a termination codon at bp 1566–1568. The putative start codon was not present in either of these cDNA clones; the 5′ end of the available sequence corresponded to bp +22 in the human sequence. In an attempt to clone the 5′ end of this gene, 5′ RACE was utilized. No novel sequence was generated following five attempts at amplification from a mouse brain cDNA library (Clontech). The translation stop codon in the mouse clone was a TAA compared to a TAG in *WHSC2* but it was found in the same location. The variant polyadenylation signal ATTAAA was also conserved. To confirm that this was the 3′ end of the gene, 3′ RACE was carried out using a primer located near the 3′ end of the gene and a pool of mouse brain poly(A)$^+$ RNA. A single extension product was amplified and sequenced; this confirmed the location of the 3′ end of this gene. These data are highly suggestive that this is the murine homologue of *WHSC2* and therefore it was named *Whsc2h*.

Transcript Analysis of Whsc2h

To analyze the gene expression pattern of *Whsc2h*, Northern blots containing poly(A)$^+$ RNA from mouse tissues were hybridized with cDNA clone 437082. Results from the hybridization of cDNA clone 437082 to poly(A)$^+$ RNA from a series of adult tissues and embryonic stages show that it is ubiquitously expressed (Fig. 5). Use of a control probe (*β-actin*) confirms that

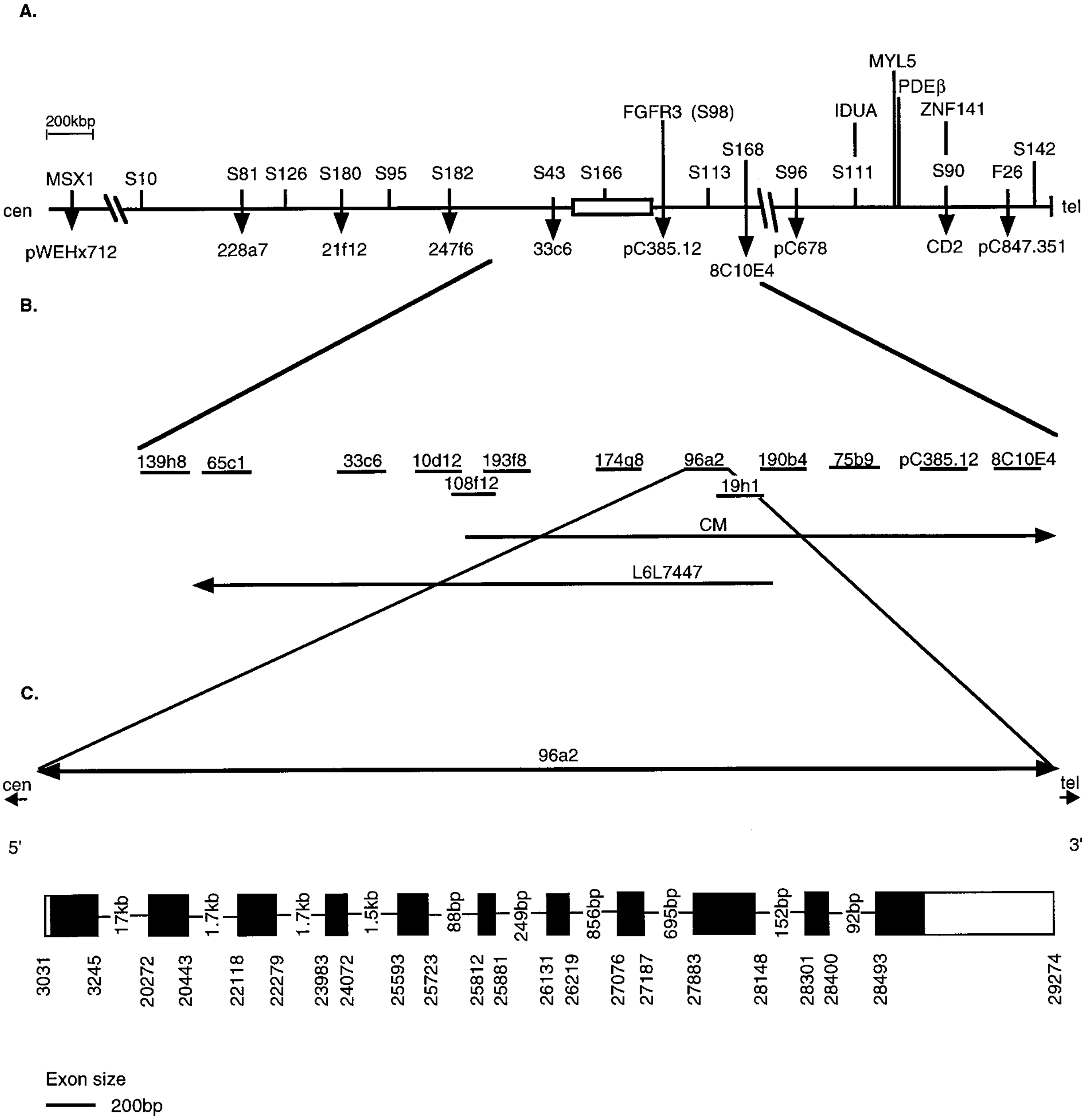

FIG. 2. The intron/exon and genomic structure of *WHSC2.* (**A**) Physical map of 4p16. The loci, shown above the line, corresponding to a series of "landmark" cosmids used as the probes, shown below the line, are indicated. The distance from D4S81 to the telomere is approximately 4.5 Mb. The genetic distance between *MSX1* and D4S10 is 3 cM. The critical region defined by the WHS patients in previous studies is depicted by an open box. (**B**) Second-tier cosmids used in FISH analysis determine the breakpoints in the patients defining the critical region. Cosmids are denoted by plate and row by column coordinates in the Los Alamos chromosome 4 library array. The continuous horizontal lines indicate the extent of the deletion in the designated patient; arrowheads indicate that the deletion continues either telomeric (CM) or centromeric (LGL7447). (**C**) The intron/exon and genomic structure of *WHSC2* is shown relative to the cosmid 96a2. Direction of transcription and cen/tel orientation are shown. Exons are shown by boxes, and intron sizes are shown between the exons. The locations of the suspected initiation and termination codons are denoted by the beginning and end of the filled-in boxes, respectively. The sequence coordinates of the exons within cosmid 96a2 (as defined by the Sanger Centre cosmid sequence) are shown.

the spike in expression seen at 11 days postcoitum (dpc) is due to unequal loading (Fig. 5, right). Use of *β-actin* confirmed a reproducible underrepresentation of the *Whsc2h* transcript in adult mouse skeletal muscle tissue (Fig. 5, right). In general, these results are similar to the data generated from the human expression analysis and confirm the likelihood that we have cloned the murine homologue of the *WHSC2* gene.

Mouse in Situ Analysis of Whsc2h

To evaluate the spatial and temporal distribution of *Whsc2h* expression in the developing embryo, *in situ* hybridization to whole embryos or to tissue sections between 10.5 and 13.5 dpc of embryonic development was performed using cDNA clone 437082 as a probe. *In situ* hybridization confirms the ubiquitous pattern of

TABLE 1

Intron/Exon Sequences of *WHSC2*

Exon	Intron 3′	Exon sequence 5′	Exon 3′	Intron 5′	Exon size (bp)	Intron size (bp)	Location in cosmid 96a2
1			GACGAG	gtaagg	215	17026	3,031–3,245
2	ttgcag	ATGAAG	AAAAGG	gtatgt	171	1674	20,272–20,443
3	cggcag	TGGGTG	AGAAGT	gtgagt	161	1703	22,118–22,279
4	ttccag	CCACGG	CCACCA	gtaagg	89	1520	23,983–24,072
5	ctccag	CCCCAC	GTGAAG	gtaggc	130	88	25,593–25,723
6	tggcag	CTGCTG	CTCTCG	gtgggg	69	249	25,812–25,881
7	ccacag	ATGCGG	ACGCAG	gtaggg	88	856	26,131–26,219
8	ttccag	AAACTT	CCCCAG	gtgagt	111	695	27,076–27,187
9	ttgcag	CCCCAT	CTCACG	gtaggt	265	152	27,883–28,148
10	ctgtag	AGAGAG	CCCGAG	gtatgt	99	92	28,301–28,400
11	ctgcag	AGAACC			781		28,493–29,274

expression. While the antisense probes demonstrate greater signal than that seen with the sense control probe, it did not appear to hybridize specifically to any region of the embryo (Figs. 6C and 6D). As a positive control for the *in situ* hybridization, adjacent sections were probed for the expression of *shh*. These hybridizations showed expression of *shh* in the developing CNS, gut, and facial primordia as previously reported (Fig. 6B) (FrancisWest *et al.*, 1995).

Computational Analysis of the Putative Protein, WHSC2

WHSC2 encodes a putative protein of 528 amino acids with a predicted molecular mass of 57 kDa (Fig. 1). The amino acid identity between the mouse and the human putative proteins is 93.3% as illustrated in Fig. 4. The amino-terminal 7 amino acids have not been identified in the mouse homologue.

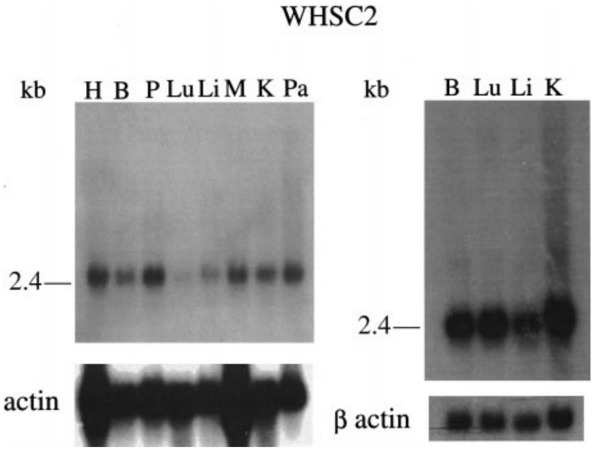

FIG. 3. Analysis of Northern blots containing adult and fetal RNA using cDNA clone 53283. Northern blots containing 2 μg poly(A)$^+$ RNA from various adult or fetal tissues were purchased from Clontech. The blots were hybridized with cDNA clone 53283 and a control probe *β-actin*. Blots were washed at 0.1x SSC, 0.1% SDS at 65°C and exposed overnight at −70°C. (**Left**) Northern blot containing poly(A)$^+$ RNA from a range of adult tissues. **H,** Heart; **B,** brain; **P,** placenta; **Lu,** lung; **Li,** liver; **M,** muscle; **K,** kidney; **Pa,** pancreas (**Right**) Northern blot containing poly(A)$^+$ RNA from a range of fetal tissues. **B,** Brain; **Lu,** lung; **Li,** liver; **K,** kidney.

The human and murine protein sequences were analyzed using different computational tools available via the World Wide Web. A transmembrane region with an amino-terminus outside/carboxyl-terminus inside orientation was predicted for both the human and the mouse proteins using TMPred (Hofmann and Stoffel, 1993). The location of this domain is shown in Fig. 4.

A hydropathic profile of the human and mouse proteins was generated using the Kyte–Doolittle method (Kyte and Doolittle, 1982). Both plots identified a predominately hydrophilic protein that extended to the carboxyl-terminus.

To determine the possible cellular localization of WHSC2, the program PSORTII was utilized (Nakai and Kanehisa, 1992). Analysis of both the human and the mouse protein sequences predicted nuclear localization with a reliability of 94.1%. Identical nuclear localization signals were found in both the human and the mouse proteins (Fig. 4). These are located in the human protein sequence starting at amino acids 166 and 272 (Fig. 4) and are composed of four or five amino acid combinations of lysine and arginine.

Analysis of both the human and the mouse protein sequences using the program COILS (Lupas *et al.*, 1991) predicted a potential coiled-coil region at amino acids 108–128 in the human sequence. This domain is shown in Fig. 4. Finally, no similarity to functionally characterized proteins was found.

DISCUSSION

Wolf–Hirschhorn and Pitt–Rogers–Danks syndromes are allelic variants resulting from the segmental aneusomy of chromosome 4p16.3. Although the deletions may extend millions of basepairs, encompassing much of chromosome 4p, analysis of WHS patients using a series of overlapping cosmids for FISH reduced the critical region to 165 kb. A number of potential transcriptional units were identified in this region (Wright *et al.*, 1997). At present it is impossible to determine whether the abnormalities seen are due

```
MASMRESDTGLWLHNKLGATDELWAPPSIASLLTAAVIDNIRLCFHGLSSAVKLKLLLGTLHLPRRTVDEMKGALMEIIQLASLDSDPWV
       --------------------------------------- ------------------------  --- ----- -------
       DTGLWLHNKLGATDELWAPPSIASLLTAAVIDNIRLCFHRLSSAVKLKLLLGTLHLPRRTVDEMNAALMDIIQLATLDSDPWV
         ^10       ^20       ^30       ^40       ^50       ^60       ^70       ^80       ^90

LMVADILKSFPDTGSLNLELEEQNPNVQDILGELREKVGECEASAMLPLECQYLNKNALTTLAGPLTPPVKHFQLKRKPKSATLRAELLQ
------------------ --------- --------- ---------------------------------------------------
LMVADILKSFPDTGSLNLDLEEQNPNVQEILGELREKVSECEASAMLPLECQYLNKNALTTLAGPLTPPVKHFQLKRKPKSATLRAELLQ
         ^100      ^110      ^120      ^130      ^140      ^150      ^160      ^170      ^180

KSTETAQQLKRSAGVPFHAKGRGLLRKMDTTTPLKGIPKQAPFRSPTAPSVFSPTGNRTPIPPSRTLLRKERGVKLLDISELDMVGAGRE
 ---------------------------------------------- ------ ----------- - -------------  ------
ISTETAQQLKRSAGVPFHAKGRGLLRKMDTTTPLKGIPKQAPFRSPTTPSVFSPSGNRTPIPPSRTPLQKERGVKLLDISELNTVGAGRE
         ^190      ^200      ^210      ^220      ^230      ^240      ^250      ^260      ^270

AKRRRKTLDAEVVEKPAKEETVVENATPDYAAGLVSTQKLGSLNNEPALPSTSYLPSTPSVVPASSYIPSSETPPAPSSREASRPPEEPS
--------- ------ --------------------------- -- ----------------------------  ------------
AKRRRKTLDTEVVEKPTKEETVVENATPDYAAGLVSTQKLGSLNSEPTLPSTSYLPSTPSVVPASSYIPSSETPPATPSREASRPPEEPS
         ^280      ^290      ^300      ^310      ^320      ^330      ^340      ^350      ^360

APSPTLPAQFKQRAPMYNSGLSPATPTPAAPTSPLTPTTPPAVAPTTQTPPVAMVAPQTQAPA--QQQPKKNLSLTREQMFAAQEMFKTA
------- ------------ ---  -------- - ------ -- ----------------  --------- ---------------
APSPTLPTQFKQRAPMYNSGVSPA--TPAAPTSPRTHTTPPAVTPTAQTPPVAMVAPQTQAPAPVQQQPKKNLSFTREQMFAAQEMFKTA
         ^370      ^380      ^390      ^400      ^410      ^420        ^430      ^440

NKVTRPEKALILGFMAGSRENPCQEQGDVIQIKLSEHTEDLPKADGQGSTTMLVDTVFEMNYATGQWTRFKKYKPMTNVS.
----------------------- --------------------------------------------------------
NKVTRPEKALILGFMAGSRENPCPEQGDVIQIKLSEHTEDLPKADGQGSTTMLVDTVFEMNYATGQWTRFKKYKPMTNVS.
 ^450      ^460      ^470      ^480      ^490      ^500      ^510      ^520
```

FIG. 4. Comparison of predicted amino acid sequence of *WHSC2* and *Whsc2h*. Polypeptides were compared using the MegAlign program (DNAStar). *WHSC2* amino acid sequence (**top**) vs *Whsc2h* amino acid sequence (**bottom**). Amino acid coordinates are given for the human protein (1–529). Mouse clones containing the amino-terminal 7 amino acids including the putative initiation codon were recalitrant to cloning (see text). The middle line shows the consensus sequence with a horizontal dash indicating conservation; identity between the two proteins is 93.3%. Dashed box represents the predicted transmembrane domain; solid boxes represent the nuclear localization signals; boldface box represents the predicted coiled coil domain.

to the loss of a single gene or the loss of several genes. The syndrome is characterized by mental retardation, suggesting defective neurological development, while the facial, skeletal, and heart defects strongly implicate abnormalities in production/survival or proliferation of neural crest, which arises from the neuro-epithelium. However, the abnormalities in the limb and urogenital tract suggest that WHS is not just the result of abnormal development of neuroepithelium and its derivatives.

To date, only one gene has been characterized in the WHSCR, *WHSC1* (Stec *et al.*, 1998). This gene merged two of the putative transcriptional units identified in the original transcript map, HFBEP10 and 194164 (Wright *et al.*, 1997). We have characterized another putative gene originally defined by the cDNA clone 53283 and named it *WHSC2*. Interestingly, the two genes are transcribed in opposite orientations with the 3′ ends being separated by only 523 bp.

Our results show that *WHSC2* is contained totally within the WHSCR and is composed of 11 exons spanning a 26-kb genomic segment (Fig. 2). Two nucleotide variants were identified by comparing the 53283 cDNA sequence to the colinear sequence derived from the genomic cosmid clone 96a2. These variants include one single base substitution within the coding region and a single base indel in the 3′ UTR. The base substitution is found in exon 9 and is a transition from T (cDNA) to C (genomic). This change does not alter the amino acid codon. The indel identified in the 3′ UTR is an A insertion in cDNA 53283 sequence when compared to the genomic sequence. The indel is found in the 3′ UTR and therefore will not affect the protein coding potential of this gene.

Transcript analysis of *WHSC2* using Northern blots containing both adult and fetal RNA tissues and cDNA clone 53283 as a probe identified a single, ubiquitously expressed transcript (Fig. 3). Interestingly, it appears that this transcript is underrepresented in the lung (Fig. 3, left). However, in the fetal lung the level of expression is similar to the other tissues analyzed (Fig. 3, right). Lung abnormalities that have been identified in WHS/PRDS patients include right and left isomerism of the lungs, incomplete lobation of the right lung in a 29-year-old, and both hypoplastic and multilobated lungs in a 31-year-old. Although difficult to reconcile with the fetal Northern blot findings, it is tempting to speculate that *WHSC2* expression in the lung may contribute to these abnormalities (Tachdijan *et al.*, 1992). Finally, there appears to be more WHSC2 transcript in fetal tissue than in adult when compared to the amount of control probe.

A second IMAGE cDNA clone, 267784, was identified in the same region as cDNA clone 53283 by sequence similarity searching (Wright *et al.*, 1997). The cDNA clone 267784 comprises 2 exons that span a 10.5-kb

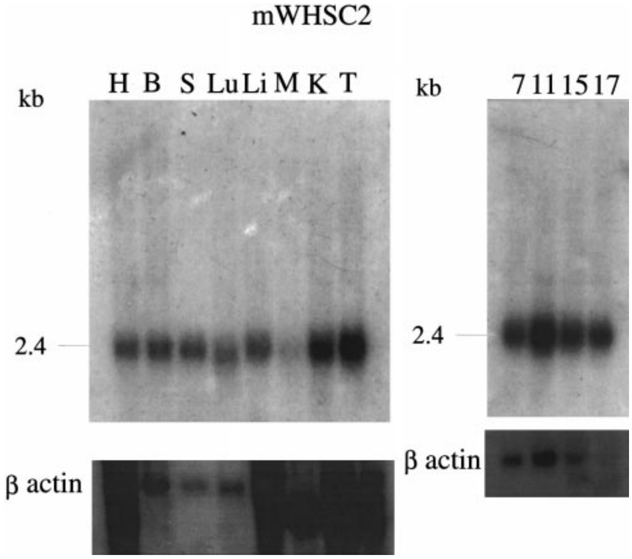

FIG. 5. Analysis of Northern blots containing adult and embryonic mouse RNA using cDNA clone 437082. Northern blots containing 2 μg poly(A)$^+$ RNA from various embryonic stages or adult mouse tissues were purchased from Clontech. The blots were hybridized with cDNA clone 437082 and a control probe *β-actin*. Blots were washed at 0.1x SSC, 0.1% SDS at 65°C and exposed overnight at −70°C. (**Left**) Northern blot containing poly(A)$^+$ RNA from a range of adult mouse tissues. **H,** Heart; **B,** brain; **S,** spleen; **Lu,** lung; **Li,** liver; **M,** muscle; **K,** kidney; **T,** testis. (**Right**) Northern blot containing poly(A)$^+$ RNA from mouse embryos at different stages. Lanes 1, 7 dpc; 2, 11 dpc; 3, 15 dpc; 4, 17 dpc.

genomic region. The cDNA clones 53283 and 267784 share the same polyadenylation site but diverge 174 bp 5′ to the poly(A) tail. The remainder of cDNA clone 267784 is found within intron 1 of *WHSC2*. Conversely, exons 2–10 of *WHSC2* are found within the intron of cDNA clone 267784. At this point, it is not known whether cDNA clone 267784 identifies a bona fide gene. However, the transcript data from *WHSC2* make it unlikely that cDNA clone 267784 is an alternative splicing product of *WHSC2*.

Once completed, the sequence of *WHSC2* was used to search dbEST, and two mouse IMAGE cDNA clones, 437082 and 354189, were identified. Analysis of these clones showed that cDNA clone 354189 was primed from an internal poly(A) tract but no other sequence differences were identified between the two clones. Sequence identity between the mouse and human sequences is 84.2% across the coding region and drops to 45% across the 3′ UTR. Amino acid identity is 93.3% between the two predicted polypeptides (Fig. 4). Therefore, we conclude that this gene is the murine homologue of *WHSC2*, and it was named *Whsc2h*.

Transcript analysis of *Whsc2h* using Northern blot analysis in a range of adult mouse tissues and a series of mouse developmental stages identified a single ubiquitous transcript in all samples (Fig. 5). *In situ* hybridization with the antisense *Whsc2h* probe on whole mount and tissue sections from 10.5 to 13.5 dpc did not reveal any specific areas of expression at any stage tested (Figs. 6C and 6D). At 10.5 dpc the majority of the organs have formed in the developing embryo, while at 13.5 dpc patterning of the embryo is almost complete and differentiation of the different cell types in various regions such as the cartilage and muscle in the limb has started (Fig. 6A). The *in situ* hybridization data together with the Northern data suggest that the *Whsc2h* transcript is expressed ubiquitously throughout the developmental stages studied (Figs. 5, 6C, and 6D). As was noted for the human transcript, there appears to be more Whsc2h RNA in fetal tissue than in adult when compared to the amount of control probe.

WHSC2 has a 1584-kb open reading frame encoding a 528-amino-acid protein with a predicted molecular mass of 57 kDa. In an effort to predict the cellular localization and possible function of this gene, several computational analyses were carried out with both the mouse and the human predicted polypeptides. The predicted protein WHSC2 has a hydrophilic amino-terminus followed by a transmembrane region and a hydrophilic carboxyl-terminus (Fig. 4). Interestingly, there are two predicted nuclear localization signals in both the human and the mouse proteins. These are found

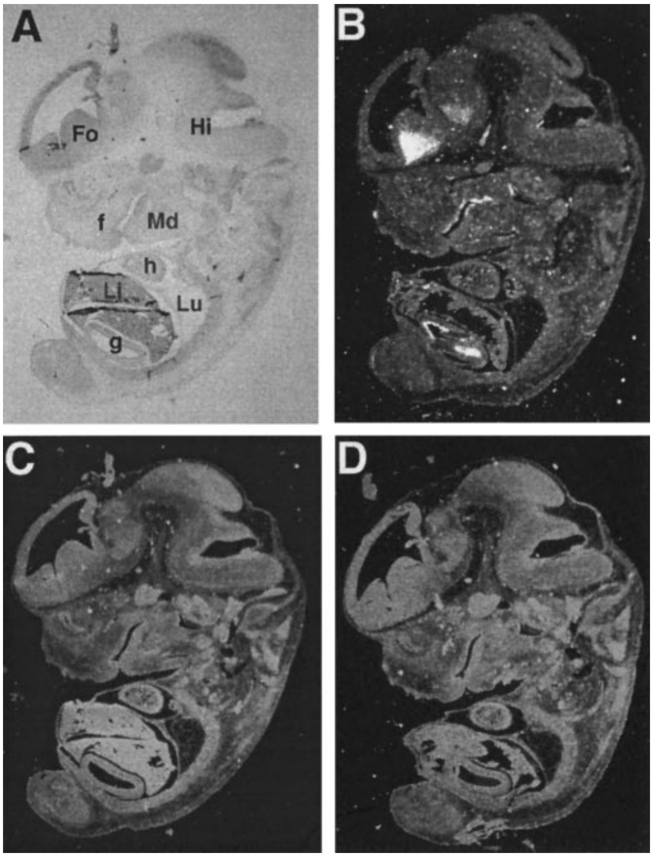

FIG. 6. *In situ* hybridization analysis of *Whsc2h*. (**A**) Bright-field and (**B–D**) adjacent dark-field figures of a sagittal sections through a 13.5 dpc mouse embryo that have been probed for the expression of (**B**) *shh* and (**C**) *Whsc2*. (**D**) Control section that has been probed with the sense *Whsc2* riboprobe and shows background levels of expression. In **B–D,** autoradiographic signal is seen as white grains under dark-field illumination. F, frontonasal mass; H, heart; G, gut, L, liver; Md, mandibular primordium; Fo, forebrain; Hi, hindbrain; Lu, lung.

starting at amino acids 166 and 272 in the human protein (Fig. 4). Each signal is composed of a 4- or 5-amino-acid combination of lysine and arginine. Therefore, it is possible that following the transmembrane region, the carboxyl-terminus is located within the nucleus as determined by the hydrophilicity/hydrophobicity and nuclear localization signals. This is in agreement with the program PSORTII, which predicted nuclear localization of the protein. Consequently we hypothesize that this peptide resides intranuclearly near the nuclear membrane. We are not able to speculate on the function at this time but the availability of the mouse clone should facilitate the development of mouse strains for genetic and biochemical analyses.

Although it is more difficult to associate function with a gene when there are no obvious indications provided by database similarities, it is possible that a gene with no known functional motifs and a ubiquitous pattern of expression can be involved in an interesting pathway. Recent studies have shown that the *Drosophila* gene Mothers against dpp (*Mad*) and its homologues play an important role in the intracellular signal transduction of the serine threonine kinase receptors (Takenoshita *et al.*, 1998). Initial identification of the Mad gene family and the predicted polypeptide from MAD revealed no similarity to known protein motifs (Sekelsky *et al.*, 1995). In mammals, one of the Mad homologues (MADH2/Smad2) was shown to be a mediator of TGF-β and activin signaling and is mutated in some colon and lung cancer cases (Takenoshita *et al.*, 1998). Analysis of this gene using Northern blot analysis has shown that MADH2 is ubiquitously expressed in all tissues examined (Takenoshita *et al.*, 1998). The pathway by which this ubiquitously expressed gene is involved in some cancer cases is not yet known.

Although the function of *WHSC2* remains to be determined, its location in the Wolf–Hirschhorn syndrome critical region and the identification of its mouse homologue, *Whsc2h*, are highly suggestive that it encodes a protein that may play a role in WHS/PRDS.

ACKNOWLEDGMENTS

Funding for this project was provided by the Los Alamos National Laboratory Directed Research and Development Fund (T.J.W., C.N., M.R.A.) and the Medical Research Council UK (P.F.W.). J.C. was funded by the Department of Energy through the Science and Energy Research Semester (SERS) Program. The authors thank Tom Peat and Norman Doggett for critically reading the manuscript.

Note added in proof. Recently, the authors were contacted by Dr. Yamaguchi, working in the laboratory of Hiroshi Handa at the Tokyo Institute of Technology, and were informed that *WHSC2* appears to be a member of the NELF protein complex (described in their paper in *Cell* **97:** 41–51, 1999) that participates in the regulation of RNA polymerase II transcription elongation.

REFERENCES

Altherr, M. R., Bengtsson, U., Elder, F. F. B., Ledbetter, D. H., Wasmuth, J. J., McDonald, M. E., Gusella, J. F., and Greenberg, F. (1991). Molecular confirmation of Wolf–Hirschhorn syndrome with a subtle translocation of chromosome 4. *Am. J. Hum. Genet.* **49:** 1235–1242.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Auffray, C., Behar, G., Bois, F., Bouchier, C., Dasilva, C., Devignes, M., Duprat, S., Houlgatte, R., Jumeau, M., Lamy, B., Lorenzo, F., Mitchell, H., Mariagesamson, R., Pietu, G., Pouliot, Y., Sebastiani-kabaktchis, C., and Tessier, A. (1995). IMAGE: Integrated molecular analysis of the human genome and its expression. *C.R. Acad. Sci. III* **318:** 263–272.

Bates, G. P., MacDonald, M. E., Baxendale, S., Youngman, S., Lin, C., Whaley, W. L., Wasmuth, J. J., Gusella, J. F., and Lehrach, H. (1991). Defined physical limits of the Huntington disease gene candidate region. *Am. J. Hum. Genet.* **49:** 7–16.

Baxendale, S., MacDonald, M. E., Mott, R., Francis, F., Lin, C., Kirby, S. F., James, M., Zehetner, G., Hummerich, H., Valdes, J., Collins, F. S., Deaven, L. J., Gusella, J. F., Lehrach, H., and Bates, G. P. (1993). A cosmid contig and high resolution restriction map of the 2 megabase region containing the Huntington's disease gene. *Nat. Genet.* **4:** 181–186.

Clemens, M., McPherson, E. W., and Surt, U. (1995). 4p microdeletion in a child with Pitt–Rogers–Danks syndrome. *Am. J. Hum. Genet.* **57:** A85.

Colgan, D. F., and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11:** 2755–2766.

Francis, P. H., Richardson, M. K., Brickell, P. M., and Tickle, C. (1994). Bone morphogenetic proteins and a signaling pathway that controls patterning in the the developing chick limb. *Development* **120:** 209–218.

FrancisWest, P. H., Robertson, K. E., Ede, D. A., Rodriguez, C., Izpisuabelmonte, J. C., Houston, B., Burt, D. W., Gribbin, C., Brickell, P. M., and Tickle, C. (1995). Expression of genes encoding bone morphogenetic proteins and sonic hedgehog in Talpid (TA(3)) limb buds: Their relationships in the signaling cascade involved in limb patterning. *Dev. Dyn.* **203:** 187–197.

Hirschhorn, K., Cooper, H. L., and Firschein, I. L. (1965). Deletion of short arms of chromosome 4–5 in a child with defects of midline fusion. *Humangenetik* **1:** 479–482.

Hofmann, K., and Stoffel, W. (1993). TMbase—A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374:** 166.

Johnson, W. P., Altherr, M. R., Blake, J. M., and Keppen, L. D. (1994). FISH detection of Wolf–Hirschhorn syndrome: Exclusion of D4F26 as critical site. *Am. J. Med. Genet.* **52:** 70–74.

Kant, S. G., van Haeringen, A., Bakker, E., Stec, I., Donnai, D., Mollevanger, P., Beverstock, G. C., Lindeman-Kusse, M. C., and van Ommen, G.-J. B. (1997). Pitt–Rogers–Danks syndrome and Wolf–Hirschhorn syndrome are caused by a deletion in the same region on chromosome 4p16.3. *J. Med. Genet.* **34:** 569–572.

Kozak, M. (1996). Interpreting cDNA sequences: Some insights from studies on translation. *Mamm. Genome* **7:** 563–574.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252:** 1162–1164.

Lurie, I. W., Lazjuk, G. I., Ussova, Y. I., Presman, E. B., and Gurevich, D. B. (1980). The Wolf–Hirschhorn syndrome. *Clin. Genet.* **17:** 375–384.

Nakai, K., and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14:** 897–911.

Pitt, D., Rogers, J. G., and Danks, D. M. (1984). Mental retardation, unusual face, and intrauterine growth retardation. *Am. J. Med. Genet.* **19:** 307–313.

Reese, M. G., and Eeckman, F. E. (1995). "Novel Neural Network Algorithms for improved eukaryotic Promoter Site Recognition," Seventh International Genome Sequencing and Analysis Conference, Hilton Head Island, SC.

Roulston, D., Altherr, M., Wasmuth, J. J., Christian, C., Graham, J., and Schreck, R. R. (1991). Confirmation of a suspected deletion of 4p16 by fluorescent in situ hybridization (FISH) with a cosmid probe. *Am. J. Hum. Genet.* **49:** 274.

Sekelsky, J., Newfeld, S. J., Raftery, L., Chartoiff, E., and Gelbart, W. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in *Drosophila melanogaster*. *Genetics* **139:** 1347–1358.

Stec, I., Wright, T. J., van Ommen, G.-J. B., de Boer, P. A. J., van Haeringen, A., Moorman, A. F. M., Altherr, M. R., and den Dunnen, J. T. (1998). WHSC1, a 90kb SET domain-containing gene, expessed in early development and homologous to a *Drosophila* dysmorphy gene maps in the Wolf–Hirschhorn syndrome critical region and is fused to *IgH* in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7:** 1071–1082.

Tachdijan, G., Fondacci, C., Tapia, S., Huten, Y., Blot, P., and Nessman, C. (1992). The Wolf–Hirschhorn syndrome in fetuses. *Clin. Genet.* **42:** 281–287.

Takenoshita, S., Mogi, A., Nagashima, M., Yang, K., Yagi, K., Hanyu, A., Nagamachi, Y., Miyazono, K., and Hagiwara, K. (1998). Characterization of the MADH2/Smad2 gene, a human *Mad* homolog responsible for the transforming growth factor-β and activin signal transduction pathway. *Genomics* **48:** 1–11.

Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88:** 11261–5.

Wahle, E., and Keller, W. (1992). The biochemistry of 3′-end cleavage and polyadenylation of messenger RNA precursors. *Annu. Rev. Biochem.* **61:** 419–440.

Wilson, M. G., Towner, J. W., Coffin, G. S., Ebbin, A. J., Siris, E., and Brager, P. (1981). Genetic and clinical studies in 13 patients with the Wolf–Hirschhorn syndrome [del (4p)]. *Hum. Genet.* **59:** 297–307.

Wolf, U., Reinwein, H., Porsch, R., Schroter, R., and Baitsch, H. (1965). Defizienz an den kurzen Armen eines Chromosoms nr. 4. *Humangenetik* **1:** 397–413.

Wright, T. J., Clemens, M., Quarrell, O., and Altherr, M. R. (1998). Wolf–Hirschhorn and Pitt–Rogers–Danks syndromes caused by overlapping 4p deletions. *Am. J. Med. Genet.* **75:** 345–350.

Wright, T. J., Ricke, D. O., Denison, K., Abmayr, S., Cotter, P. D., Hirschhorn, K., Keinanen, M., McDonald-McGinn, D., Somer, M., Spinner, N., Yang-Feng, T., Zackai, E., and Altherr, M. R. (1997). A transcript map of the newly defined 165kb Wolf–Hirschhorn syndrome critical region. *Hum. Mol. Genet.* **6:** 317–324.

Zuo, J., Robbins, C., Baharloo, S., Cox, D. R., and Myers, R. M. (1993). Construction of cosmid contigs and high-resolution restriction mapping of the Huntington disease region of chromosome 4. *Hum. Mol. Genet.* **2:** 889–899.