Mammalian Genome 11, 890–898
DOI: 10.1007/s003350010166

Mammalian Genome
Incorporating Mouse Genome

Comparative genomic sequence analysis of the Williams syndrome region (LIMK1-RFC2) of human Chromosome 7q11.23

Duane W. Martindale,[1] Michael D. Wilson,[1] Diana Wang,[1] Robert D. Burke,[1] Xianming Chen,[2] Vincent Duronio,[2] Ben F. Koop[1]

[1]Department of Biology, Centre for Environmental Health, P.O. Box 3020, University of Victoria, Victoria, British Columbia, V8W 3N5 Canada
[2]Department of Medicine, Jack Bell Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

Received: 27 December 1999 / Accepted: 18 May 2000

Abstract. Williams syndrome (WS) is a complex neurodevelopmental disorder arising from a microdeletion at Chr band 7q11.23, which results in a hemizygous condition for a number of genes. Within this region we have completely characterized 200 kb containing the genes LIMK1, WBSCR1, and RFC2. Evidence was also found for WBSCR5 in this region, but not the previously proposed genes WSCR2 and WSCR6. The syntenic region in mouse was also sequenced (115 kb) and characterized, and a comparative sequence analysis with a percent identity plot (PIP) easily allowed us to identify coding exons. This genomic region is GC rich (50.1% human, 49.9% mouse) and contains an unusually high abundance of repetitive elements consisting primarily of Alu (45.4%, one of the highest levels identified to date) in human, and the B family of SINES (30.6% of the total sequence) in mouse. WBSCR1 corresponds to eukaryotic initiation factor 4H, identified in rabbit, and is herein found to be constitutively expressed in both human and mouse, with two RNA and protein products formed (exon 5 is alternatively spliced). The transcription pattern of WBSCR5 was also examined and discussed along with its putative amino acid sequence.

Introduction

Williams syndrome (WS; also known as Williams-Beuren syndrome) is a genetic disorder affecting multiple physiological systems [Online Mendelian Inheritance in Man (OMIM) 194050; http://www.ncbi.nlm.nih.gov/omim]. WS is characterized by a characteristic craniofacial dysmorphology, cardiovascular disease, hypertension, infantile hypercalcemia, dental abnormalities, hernias, and diverticuli, as well as a distinct cognitive and personality profile (Morris et al. 1988; Bellugi et al. 1990; Keating 1997). Although most WS individuals have a short attention span, reduced spatial skills, and mild mental retardation, many also tend to have a good memory, expressive language, and musical talent. The typical WS personality includes a combination of friendliness and anxiety (Bellugi et al. 1999a, 1999b).

The genetic basis of WS is just beginning to be understood. The starting point for much of the recent work stems from the finding that WS results from hemizygous microdeletions at Chr 7q11.23 that include the elastin gene (ELN) (Ewart et al. 1993). The commonly deleted interval includes the genes sequenced and analyzed in this paper: LIMK1 (encoding LIM kinase-1) (Frangiskakis et al. 1996; Osborne et al. 1996; Tassabehji et al. 1996); WBSCR1 (William-Beuren syndrome critical region 1; originally called WSCR1 and encoding eIF4H), (Osborne et al. 1996; Richter-Cook et al. 1998); WBSCR5 (originally called WSCR5 and tentatively identified by an EST) (Osborne et al. 1996); and RFC2 (encoding replication factor C subunit 2) (Osborne et al. 1996; Peoples et al. 1996). See Fig. 1A for a schematic physical map incorporating these genes and others within this region.

The size of the commonly deleted region associated with WS has been found to span approximately 1.5–2 Mb of DNA (Osborne et al. 1996; Perez Jurado et al. 1996; Robinson et al. 1996; Urban et al. 1996; Wang et al. 1997; Meng et al. 1998; Wu et al. 1998; Hockenhull et al. 1999). Recently, Botta et al. (1999) described two patients with the full Williams syndrome phenotype who carried deletions from just before the ELN gene to just past marker D7S1870 within the GTF2I gene (Fig. 1A); this 'minimal' region defines a region estimated to be approximately 850 kb. The sequences analyzed in this paper are within this smaller defined region.

The WS commonly deleted region at human Chr 7q11.23 has been found to have conserved synteny to a region on the distal portion of mouse Chr 5 (DeBry and Seldin 1996; DeSilva et al. 1999). Comparative physical mapping of the human and mouse WS regions indicates that while the duplicated regions bounding the human common deleted region are absent in mouse (DeSilva et al. 1999), the expected genes residing between *Eln* and *p47-phox* are present.

In this paper we report the sequence of both the human and syntenic mouse regions containing the known genes LIMK1, WBSCR1, and RFC2 and a novel gene, WBSCR5. The transcription patterns of WBSCR1 and WBSCR5 are examined and a brief analysis of the product of the mouse *Wbscr1* gene is included. Alternative splicing is demonstrated for both of these genes. Unusual characteristics of this genomic region are noted and discussed.

Materials and methods

DNA sequencing and analysis. A sublibrary was constructed (Rowen and Koop 1994) for each of the cosmids and the BAC clone (Fig. 1B,C). Briefly, DNA from each cosmid or BAC was isolated and randomly sheared by nebulization. The sheared DNA was then fractionated by agarose gel electrophoresis, and fragments (2–4 kb) were collected, blunt-ended, and cloned into M13mp19 by standard techniques. Random clones from each sublibrary were sequenced with the aid of ABI 373 or 377 sequencing machines and fluorescently labeled primers (ABI, Amersham). DNAStar software was used for gel trace analysis and contig assembly, as well as DNA and protein alignments. DNA and protein sequences were examined against available public databases by using the various Blast programs available through the network server at the National Center for Biotechnology. Repeat elements were characterized by using RepeatMasker2 (A.F.A. Smit and P. Green; http://ftp.genome.washington.edu/cgi-bin/RepeatMasker). Comparative sequence alignment was done with PipMaker (http://globin.cse.psu.edu/cgi-bin/pipmaker), which produces the percent

Database Accession Numbers: U63721, AF045555, AF139987.

Correspondence to: Ben Koop; E-mail: bkoop@uvic.ca

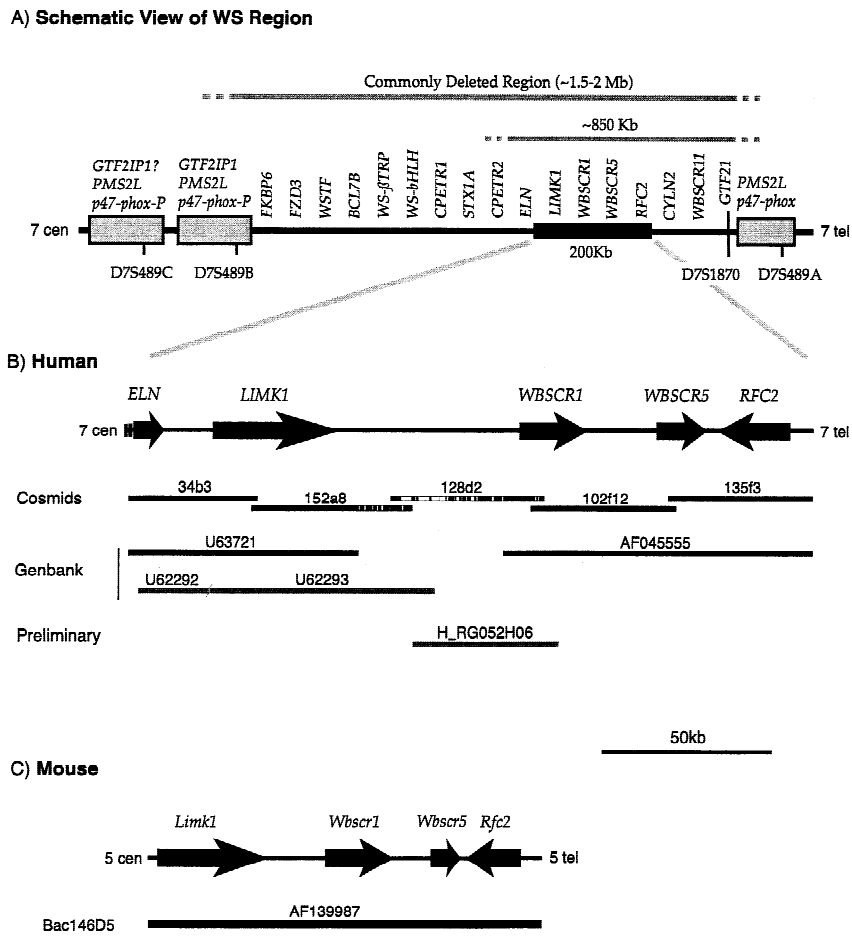

Fig. 1. Physical and transcription map of the 200-kb region downstream of ELN commonly deleted in Williams syndrome. **A)** Schematic view of the human WS region (not to scale). Grey lines delineate the extent of the commonly deleted region (extending from D7S489B to D7S1870; adapted from DeSilva et al. 1999a) and the 850 region deleted in Botta et al. (1999). **B)** For the human sequences, an overlapping set of cosmid clones (Osborne et al. 1996) were sequenced to completion (34b3, 102f12, 135f3) or near completion (152a8, 128d2). The completed sequences (Genbank U63721 and AF045555) contain the full genomic sequences of LIMK1, WBSCR1, RFC2, and the novel gene WBSCR5. By combining our data with those of Frangiskakis et al. (1996) (U62292 and U62293) and unreleased data from Washington University at St. Louis (H_RG052H06), a single continuous sequence of 199192 bp was obtained. The unsequenced regions of cosmids 152a8 and 128d2 are shown in white. Sequences from other laboratories (e.g., U62292) are grey. **C)** Bac146D5, containing mouse ES-129/Sv1 genomic DNA, was obtained from Genome Systems Inc. (St. Louis, Miss.) using PCR primers specific for *Limk1* and sequenced to completion (114898 bp). It was found to contain the full genomic sequence of *Limk1, Wbscr1, Wbscr5,* and *Rfc2.*

identity plot (PIP). The dot-plot output was also obtained with the PipMaker program, which generates alignments by using the alignment engine BlastZ (Schwartz et al. 2000). Putative exons were identified by a variety of programs (Xgrail_1.3, FGENES, FGENESH, FGENES-M; http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html).

Isolation and characterization of clones. The human genomic DNA cosmid clones used in this study were obtained from the Chr 7-specific cosmid library of the Lawrence Livermore National Laboratory (LL07NCC01). They were identified, mapped, and initially analyzed as described previously (Osborne et al. 1996). The mouse genomic DNA BAC clone (BAC146D5) was obtained from the mouse ES-129/SvJ I BAC library of Genome Systems, Inc. (St. Louis, Mo., USA) by using PCR probes for the mouse *Limk1* (LIMK1F: 5′-CGC TTC AGT TGA CCA TCT AGG-3′ and LIMK1R: 5′-CTG AGT TTG GGT TCA TTC CTG-3′).

Cloning and expression of Wbscr1 coding fragment of mouse eIF4H. The complete coding region of *Wbscr1* (681 or 741 bp) was amplified with the PCR primers WBSCR-C1F (5′-CGG GAT CCC GCG GAC TTC GAT ACC TAC GAC-3′) and MWS681R (5′-GGG GTA CCC TTC TTG CTC CTT CTG AAC CA). The PCR products from the amplified mouse liver cDNA were T-A cloned into pGEM-T vector (Promega, Madison WI USA) and sequenced. Two clones, L2 (spliced exons form) and L5, were subcloned into the expression vector pQE30 (Qiagen, Inc, Valencia, CA, USA) and transformed into *E. coli* strain M15 [pREP4]. The L2/pQE30 construct codes for a 25-kDa 6xHis-tag recombinant protein, and the L5/pQE30 construct codes for a 27-kDa 6xHis-tag recombinant protein. The expressed proteins were confirmed with immunoblots with anti-MRGS.6xHis antibody (Qiagen) and affinity-purified on a Ni-NTA agarose column (Qiagen).

Production of monoclonal antibodies. Female Balb-C mice were immunized and boosted intraperitoneally at 3-week intervals with purified L2 fusion protein emulsified in complete and incomplete Freund's adjuvant. Five days after the fourth injection, the spleen was removed and fused with the mouse myeloma x63Ag8.653 (ATCC) following standard fusion procedures. Hybridomas were plated in HAT containing methylcellulose, and individual colonies were picked and grown in RPMI 1640 (GIBCO). Supernatants from the immunoglobulin-secreting clones were screened and characterized for their immunostaining of NIH 3T3 cells and immunoblotting of L2 and L5 fusion proteins. Two Hybridomas 9B4 and 2A3 were selected.

Characterization of monoclonal antibody 9B4 by Western blot and immunocytochemistry. Multiple tissues from mouse were dissected out and sonicated in ice-cold PBS containing 1 mM 4-(2-aminoethyl)-benzenesulfonyl fluoride (AEBSF). 100 μg of crude protein from each sample and 5 μg of purified L2 fusion protein were run on 15% SDS-polyacrylamide gel electrophoresis and electrophoretically transferred onto PVDF membrane. The membrane was blocked in 5% skim milk powder, 0.1% Tween-20 in TBS, and then incubated with supernatant from 9B4 hybridoma cells and subsequently with AP-conjugated goat anti-mouse immunoglobulin (Sigma) and detected by using NBT/BCIP (Roche Diagnostics, Laval, Quebec, Canada). Relative molecular mass was determined with prestained low MW standards (GIBCO).

The newly confluent NIH 3T3 cells were fixed in ice-cold 100% methanol for 8 min and incubated with hybridoma 9B4 supernatant overnight. The cells were rinsed in PBS containing 0.1% Tween-20 and incubated with FITC-conjugated goat anti-mouse immunoglobulin (Sigma). The labeled cells were viewed with confocal laser scanning microscopy (Zeiss LSM 410).

Results

In a previous study (Osborne et al. 1996), we described the partial sequencing of eight overlapping cosmids containing human DNA

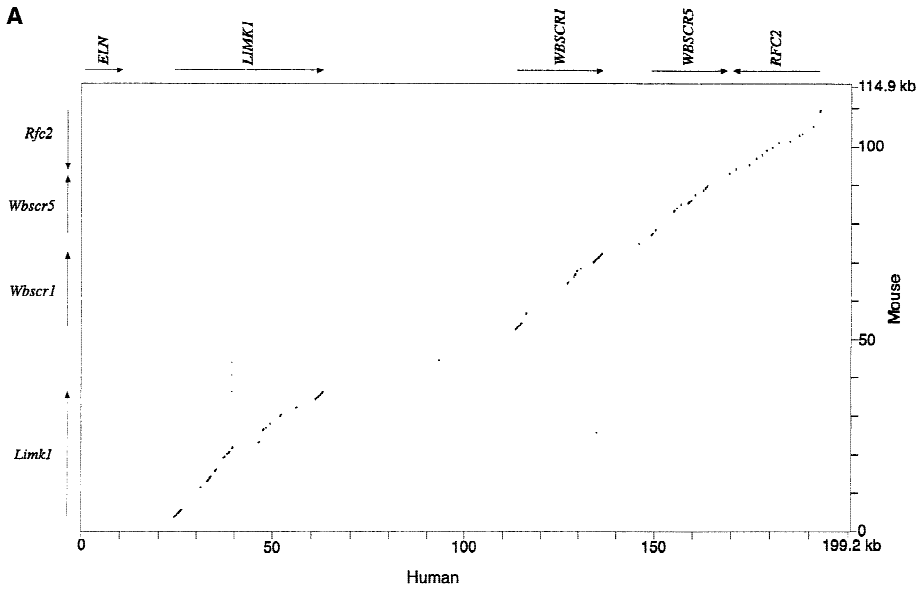

Fig. 2. A) Dot-plot display of human (x axis) and mouse (y axis) shows statistically significant local alignments between mouse and human as a diagonal line. **B)** Percent Identity Plot (PIP). The nucleotide positions for the mouse sequence is shown on the X axis, and the percent sequence identity with the corresponding human sequences (50%–100%) is shown on the Y axis. Genes are labeled based upon sequences that correspond to known or determined cDNA sequences. The locations of mouse repeats were determined by using RepeatMasker2.

telomeric of, and containing the 3′ end of the elastin (ELN) gene. These cosmids were 'scanned' in an effort to discover genes contained in a region that is commonly deleted in WS individuals. In the current study, we completed the genomic sequence containing and surrounding the LIMK1 gene (contained in cosmids 34b3 and 152a8; Genbank accession number U63721; 67 kb), as well as the genomic sequence surrounding and containing the WBSCR1 and RFC2 genes (contained on cosmids 128d2, 102f12, and 135f3; Genbank accession number AF045555; 90 kb). The region between these two completed sequences has been filled with information from our partial sequences, Genbank accession U62293 (Frangiskakis et al. 1996) and preliminary data from the Washington University (WU) sequencing project (from H_RG052H06, a BAC clone from the M.I. Simon and H. Shizuya library, Research Genetics). The preliminary WU sequence fills a 19.9-kb gap, consisting almost entirely of SINE repeats, between the known Genbank sequences. We have also sequenced 64% of this gap so we have confidence that this sequence is correct. The total length of the sequence represented in Fig. 1B is 199192 bp.

The region on mouse Chr 5 syntenic to human Chr 7q11.23 (LIMK1-RFC2) was also sequenced (Fig. 1C; total length, 114898 bp). The entire sequence was contained in the BAC clone 146d5 obtained from Genome Systems Inc. (see Materials and methods). The number, order, and orientation of the genes are conserved between human and mouse. Note that the intergenic regions are substantially less in the mouse.

From available cDNA sequences, the boundaries of the coding exons were determined for LIMK1, WBSCR5, and RFC2 (note 1–189 of the RFC2 cDNA sequence MM002914 corresponds to cyclin B2). The original published cDNA matching WBSCR1 (HUMORF) was found to be missing an alternatively spliced exon (exon 5). We had noticed two sizes of PCR products obtained from mouse cDNA when using primers that amplified the complete coding region (see Fig. 3Ca). Sequencing of these two PCR products revealed the sequence of exon 5. The exact exon-intron organization of the mouse genes was determined from published cDNAs for *Limk1* [MMRMALMK1 (1–316 did not correspond to our sequence); MMU14166 (*Kiz-1*, has full first exon); MMU15159 (first 5 bp did not correspond to our sequence)], homology to human cDNA sequences, ESTs and direct sequencing of the cDNA (*Wbscr1* and *Wbscr5*)

Three of the genes (LIMK1, WBSCR1, and RFC2) were found to be highly conserved between human and mouse. The nucleotide identity ranged from 86.7% for RFC2 to 91.4% for WBSCR1 (large, i.e., with exon 5), whereas the percent amino acid identity ranged from 95.2% for LIMK1 to 99.1% for WBSCR1 (small, i.e., without exon 5) (Table 1). RFC2 has a five-amino acid difference in the length of the non-conserved N-terminal end between the two species (30 aa for RFC2 and 25 aa for *Rfc2*). A non-conserved N-terminal end of variable size is seen in other RFC2 proteins whose complete sequence has been determined [chicken (Oberholtzer et al. 1994), fruit fly (Harrison et al. 1995), yeast (Cullmann et al. 1995), and archaebacteria (gene AF2060, accession number AE000961)]. In contrast, the coding region of WBSCR5 is significantly different in size between human and mouse; the human coding sequence is 120 nt longer than that of the mouse. Most of this extra sequence (108 nt) is found in exon 12 of the human gene (Fig. 4). In addition, WBSCR5 is less highly conserved than the other genes in this study, with the human and mouse genes having only 73.4% nucleotide identity (60.6% amino acid identity) over their matching coding regions (Table 1).

A comparison between the human and mouse sequences is shown in Fig. 2. After identifying and masking repeat elements with RepeatMasker2, the resulting sequences and exon locations were submitted to the PipMaker program website where Fig. 2A and B were produced. The data are drawn as a traditional dot-plot (Fig. 2A) and a percent identity plot (PIP; Fig. 2B), with the human DNA segments (50%–100% identical) mapped relative to the mouse sequence. The coding exons of the highly conserved genes (*Limk1, Wbscr1,* and *Rfc2*) are easily identified by visual inspection of the PIP. Even though *Wbscr5* is poorly conserved and contains small exons, they also are visible upon inspection of the PIP. Other regions of conservation such as the promoter regions of the genes indicate regions for further research.

The region between *Limk1* and *Wbscr1* is particularly barren of any likely exons based upon sequence similarity (Fig. 2A,B), gene prediction programs (XGrail, FGENES etc.) and ESTs. The rare ESTs found in this area do not correspond to predicted exons and contain repeat element sequences suggesting background transcription.

This genomic region is GC rich (50.1% human; 49.9% mouse) and contains an unusually high abundance of repetitive sequences (59.3% and 41.5% of the human and mouse sequences, respectively) (Table 2, Fig. 2B). The majority of the repeats are SINES, represented by Alu in the human sequence (45.4% of the overall sequence) and the B family of SINES in the mouse (B1, B2 and

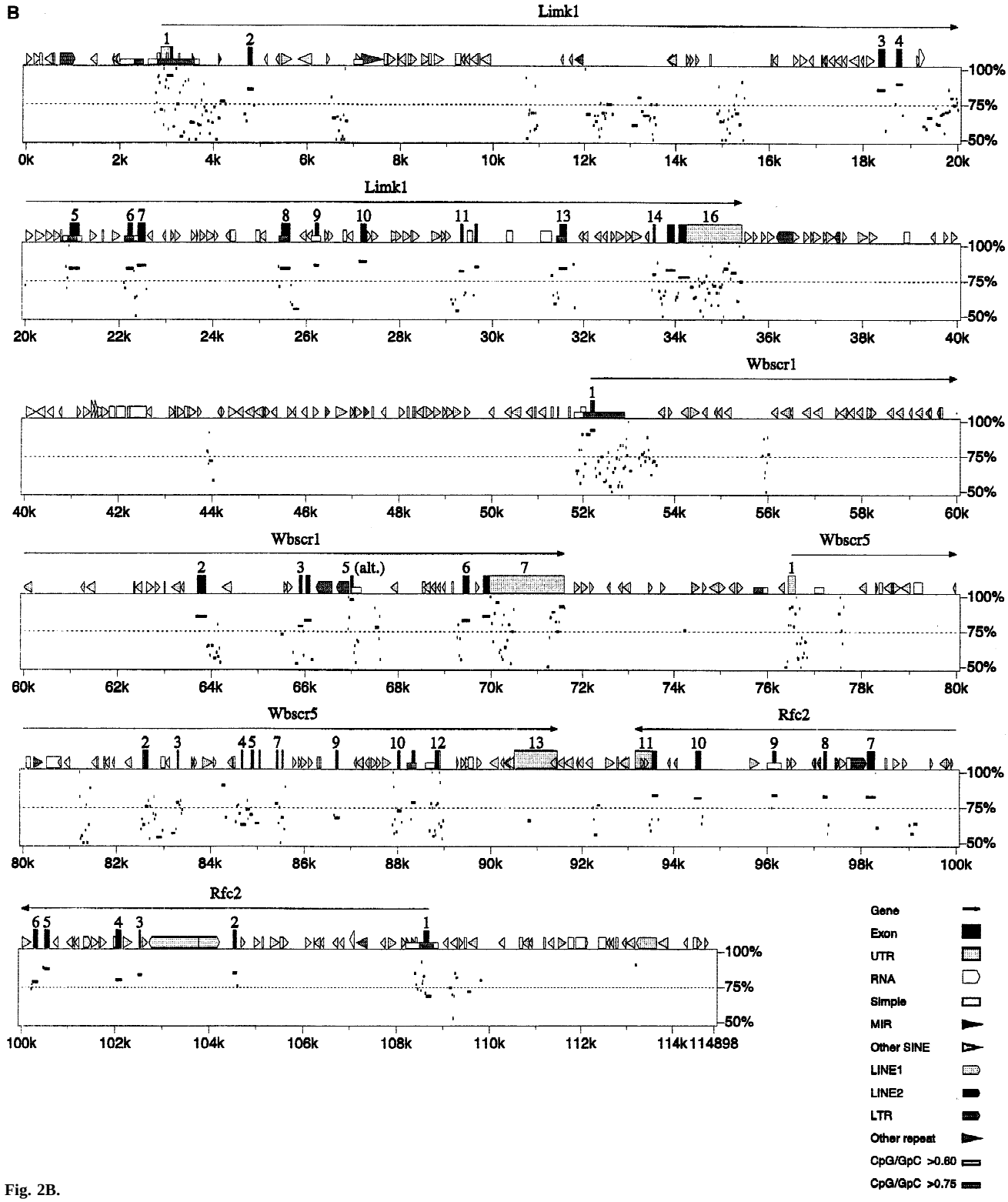

Fig. 2B.

B4; 30.6% of the overall sequence). Even though the sizes of the coding regions in these regions are similar between these two species, the human region is significantly larger (~179 kb) than that of the mouse (~115 kb). This difference can be largely accounted for by differences in the abundance of repetitive elements, since the size of the syntenic regions with the repetitive elements removed are 72.9 kb (human) and 67.2 kb (mouse).

An analysis of WBSCR1 transcription (Fig. 3A) and translation (Fig. 3B) was undertaken to confirm its functionality. Northern blot analysis (Osborne et al. 1996) had indicated a transcript of 2.5 kb which corresponds to the predicted transcript size of 2486 nt (minus exon 5) or 2546 nt (with exon 5) estimated from genomic, EST, and cDNA sequencing. The initial poor hybridization of a WBSCR1 cDNA probe to mouse genomic DNA and mRNA seen in the previous study was attributed to poor conservation of the gene, but our current analysis shows that it is highly conserved

Table 1. Gene and protein information.

Genes		Protein			%GC		Conservation	
		aa	MW	pI	Coding region	3rd position	% aa identity	% nt identity
LIMK1	human	647	72600	6.80	61.3	81.3	95.2	87.9
	mouse	647	72792	6.73	57.3	70.9	(616/647)	
WBSCR1 (large)	human	248	27388	7.11	53.5	49.6	98.8	91.4
	mouse	248	27344	7.11	55.1	51.6	(245/248)	
WBSCR1 (small)	human	228	25202	8.09	53.7	50.9	99.1	90.9
	mouse	228	25185	8.08	55.2	55.3	(226/228)	
WBSCR5	human	243	26549	4.59	58.6	66.7	60.6	73.4[a]
	mouse	203	22875	4.63	54.2	65.5	(123/203)	
RFC2	human	354	39161	6.26	52.2	63.6	97.2[b]	86.7[c]
	mouse	349	38724	6.26	54.5	70.5	(315/324)	

[a] Over the 609 nt of mouse coding sequence (the human sequence includes an extra 120 nt).
[b] Not including poorly conserved N-terminal amino acids (30 aa of RFC2 and 25 aa of *Rfc2*).
[c] Not including poorly conserved 5′ nucleotides (90 nt of RFC2 and 75 nt of *Rfc2*).

(91.4% identical over 744 bp of coding sequence; Table 1). We found the gene to be active in all tissues examined in both human and mouse (Fig. 3A) and further identified alternate splicing in both species (Fig. 3Ca). There is a marked difference in relative abundance of the two splice variant mRNAs between mouse and human tissues (Fig. 3Ca). The two mRNA variants are approximately the same levels in mouse, whereas in humans the smaller variant is by far the most abundant. This relative difference was the same in all tissues examined. In an analysis of protein expression, a monoclonal antibody produced against *Wbscr1* (9B4 antibody) binds to two bands (25 and 27 kDa) in protein samples from heart, spleen, lung, and kidney. In liver and skeletal muscle the antibody binds to a single band of about 25 kDa even though both mRNA splice variants are present in these tissues. In mouse 3T3 cells prepared for immunofluorescence with 9B4, there is immunoreactive material in the cytoplasm of all cells. The fluorescence is granular and frequently appears as a strong perinuclear ring. In some cells there are scattered foci of immunofluorescence in the nucleus (data not shown).

We isolated and sequenced the mouse cDNA representing *Wbscr5* in a separate study (see Discussion). A human cDNA for WBSCR5 was recently deposited in GenBank (AK002099) that matches the transcript we had predicted from ESTs, gene prediction programs, and similarity to the mouse gene. An analysis of WBSCR5 transcription (Fig. 3A) indicates that it is present in all tissues examined although at different levels. PCR amplification and subsequent sequencing of the products demonstrated that the human transcript was alternately spliced (Fig. 3Cb), with most transcripts missing exon 2. Exon 2 is non-coding, present in two forms (119 or 189 bp), and contains an Alu repeat. An exon equivalent to human exon 2 was not found in mouse (thus, exon 2 of mouse *Wbscr5* corresponds to exon 3 of human WBSCR5).

Discussion

Genes in WS. Williams syndrome (WS) is caused by haploinsufficiency for genes in a microdeletion at Chr band 7q11.23. The genetic distance between the most centromeric deleted locus (D7S489B) and the most telomeric (D7S1870) (see Fig. 1A) is 2 cM (1.5–2.5 Mb), with the breakpoints clustering approximately 1 cM on either side of ELN (Perez Jurado et al. 1996; Robinson et al. 1996; Urban et al. 1996; Osborne et al. 1997a, 1997b). Recently, two patients with the full Williams phenotype have been found to have deletions of less than 1 Mb extending from just before ELN to just past D7S1870 (Botta et al. 1999). In addition to ELN, genes known to exist in this smaller region are the ones we have sequenced in this paper (LIMK1, WBSCR1, WBSCR5, and RFC2) as well as CYLN2, WBSCR11, GTF2I, and possibly CPETR2.

Only two of the genes found within the WS critical region (ELN and LIMK1) have been convincingly shown to account for some of the multiple phenotypic features of WS. Point mutations in ELN as well as small deletions of ELN have been associated with heart and vascular problems without additional WS features (Ewart et al. 1993; Olson et al. 1995; Tassabehji et al. 1997). The other gene shown to account for a specific WS feature is LIMK1, which codes for LIM-kinase 1 (Mizuno et al. 1994; Okano et al. 1995). Frangiskakis et al. (1996) demonstrated that members of a rare family with a hemizygous deletion of 84 kb containing only the ELN and LIMK1 genes had some of the WS features associated with ELN (SVAS and some mild WS facial features), as well as the visuospatial constructive cognitive deficit observed in WS. The association of LIMK1 with a specific cognitive defect has been the subject of several recent reviews (e.g., Rosenblatt and Mitchison 1998). However, the genetic and psychometric testing of four patients with small deletions within the WS critical region has more recently suggested that LIMK1 does not contribute to the WS phenotype (Tassabehji et al. 1999). Clearly further studies must be done these conflicting results.

A separate study (Frangiskakis et al. 1996) presented the sequence of the human genomic LIMK1 region (Genbank U62292 and U62293) as we were completing the sequence of the same region (U63721). The gap between U62292 and U62293 of 232 bp is filled in our sequence. This gap contains the beginning of LIMK1 exon 1, is 88% GC, and contains two GC boxes in the opposite orientation. The ELN and LIMK1 exons were identical between our sequences. Sequence of clones 5′ of LIMK1 exon 6 differed by 0.2% and therefore likely represents two alleles. Sequences from 3′ of LIMK1 exon 6 differ by 0.03%. Significant differences include a small deletion (5′-CGCCGGG-3′) that was seen in the first LIMK1 intron of their sequence and a deletion of 445 bp found 1470 bp upstream of the LIMK1 start codon in our sequence, which we suspect is a cloning artifact that arose during the growth of the cosmid through a recombination event between two 23 bp identical repeats found in adjacent Alu sequences.

To aid in the detection of hemizygosity associated with WS, we identified several dinucleotide repeats in this genomic region. We constructed PCR primers to amplify a CA-repeat found in the 13th intron of LIMK1 (data not shown) and found PCR fragments ranging from 172 bp to 192 bp in the initial screen of ten individuals. In addition, Mari et al. (1998) used the genomic sequence reported here to detect a highly polymorphic CA-repeat 10 kb downstream of LIMK1 that they subsequently found useful in detecting hemizygosity in WS patients. It is anticipated that microsatellite markers throughout the commonly deleted region will allow the detection and extent of chromosomal deletions in WS patients to be determined (Nickerson et al. 1995). This information will in turn enable a more complete diagnosis and prognosis for WS patients.

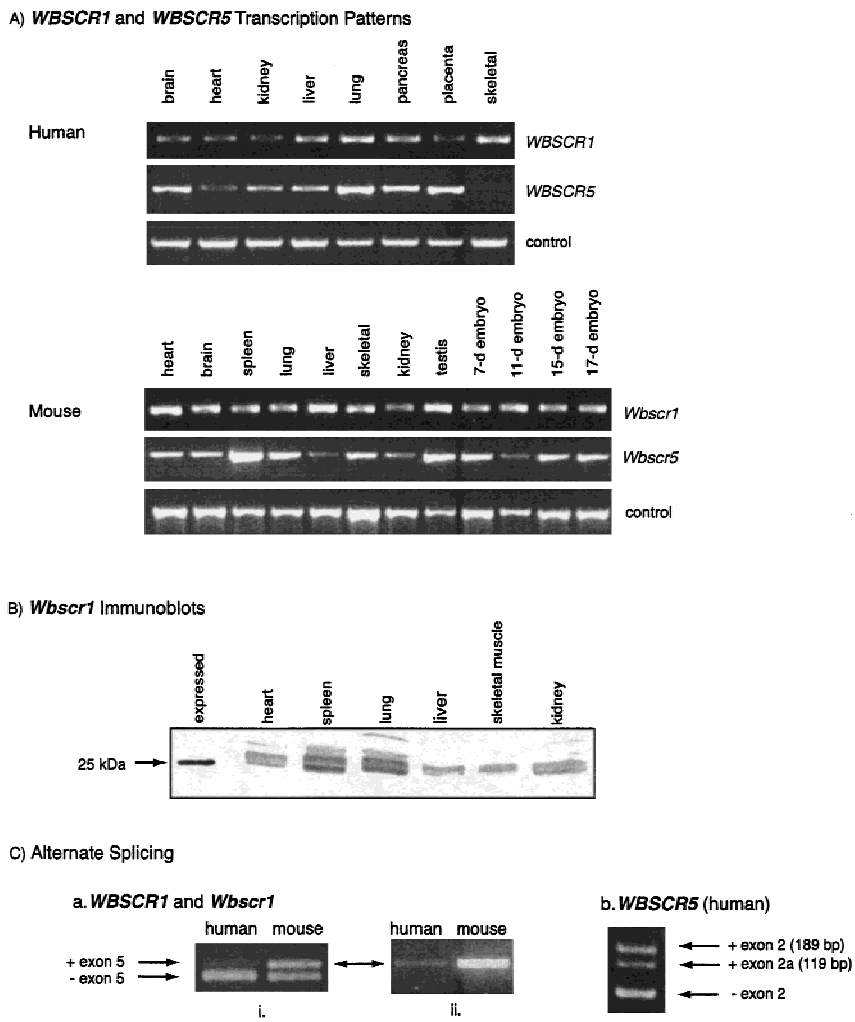

Fig. 3. A) The expression of WBSCR1 and WBSCR5 in different tissues of human and mouse. PCR products using Multiple Tissue cDNA panels from CLONTECH Laboratories, Inc. For WBSCR1: the primers used against the human cDNAs (MHWBSCR-F: 5′-GGA GAC GGC AAA TGG CGG ACT T-3′ within exon 1; MHWBSCR-exR: 5′-ATC CCA TCC ACC TCT AGA TTC TCG-3′ within exon 5) amplify a 457-bp product containing the alternately spliced exon. The primers used against the mouse cDNAs are the same as those used for amplifying the complete coding region (681 or 741 bp) of *Wbscr1* for antibody production (WBSCR-C1F and MWS681R). Results shown after 36 cycles for human and 30 cycles for mouse. For WBSCR5: the primers used against human cDNAs (hW5-259f: 5′-ACA GTT CTT GGA AAC CCA CTC G-3′; hW5-1488r: 5′-AAA ACC CCA GCA ACC AAC AAT A-3′) amplify a 1229-bp product. The primers used against mouse *Wbscr5* cDNAs (5C1-722f: 5′-GAT GTT TCC AGA AGC CCT CA-3′; 5C1-1631r: 5′-TAT TTC CCT ATC ACC GAC GC-3′) amplify a 909 bp product. Results shown after 40 cycles for human and 38 cycles for mouse. The controls used for the human and mouse MTC panels are the G3PDH Control Amplimer Set (CLONTECH). **B)** *Wbscr1* immunoblots prepared using 9B4 as a the primary antibody. Lane 1 is the expressed protein (25-kDa). **C)** Alternate splicing of WBSCR1 and WBSCR5: **a)** WBSCR1; **i.** Amplification of both forms of WBSCR1 cDNA. MHWBSCR-F and MHWBSCR-R (5′-TTG GGG/A TTG GCT ACT TGA TTG AGG-3′) were used against human and mouse pools of cDNA (30 cycles). The upper band (696 bp) contains the alternately spliced exon 5, while the lower band (636 bp) does not. **ii.)** Specific amplification of WBSCR1 cDNA containing exon 5. Primers MHWBSCR-F and MHWBSCR-exR (specific for exon 5) were used to amplify only cDNAs that contain the alternately spliced exon (30 cycles). **b)** WBSCR5; a primer from the first exon (hW5-98f: 5′-GTG GTG AGG AAC CCT GGA CTC T-3′) was combined with hW5-1488r (exon 14) to amplify three distinct bands of 1390, 1320, and 1201 bp (shown after 40 cycles). These distinct bands were gel purified and sequenced by using ABI dye terminators.

Other genes in the WS critical region sequenced in this study include RFC2, which codes for the 40-kDa ATP-binding subunit of replication factor C (RF-C) (also known as activator 1). RF-C is a multimeric protein complex containing five different subunits (RFC1–5) and plays a vital role in the elongation of DNA catalyzed by DNA polymerase δ and ϵ (Pan et al. 1993). The human RFC2 product has 97.2% amino acid identity to the mouse product (Table 1) and 60.1% amino acid identity to the product of the homologous yeast gene, *RFC4* (Cullmann et al. 1995). Both Osborne et al. (1996) and Peoples et al. (1996) speculate that a reduction in the quantity of RFC subunit 2 could lead to reduced DNA replication efficiency, accounting for WS symptoms such as growth deficiency, but there are currently no data demonstrating a specific role for this gene in the WS phenotype.

Our analysis of the human gene WBSCR1 (Osborne et al. 1996 and this study) and the analysis of the rabbit eukaryotic initiation factor 4H (eIF4H; Richter-Cook et al. 1998) suggest that these are orthologous genes. Richter-Cook et al. (1998) identified the eIFH4 protein from rabbit reticulocyte lysate on the basis of its ability to stimulate translation in an in vitro globin synthesis assay deficient in eIF4B and eIF4F. The authors demonstrated that the 25-kDa rabbit eIF4H protein stimulates the in vitro activities of EIF4B and EIF4F in globin synthesis, as well as the in vitro RNA-dependent ATPase activities of EIF4A, EIF4B, and EIF4F. Recently it has also been shown that eIF4H stimulates the helicase activity of eIF4A (Rogers et al. 1999). Our analysis reveals that the amino acid sequences of three rabbit EIF4H tryptic fragments are identical to the human WBSCR1 protein. The gene WBSCR1 encodes a protein product of 228 or 248 amino acids through alternate splicing of exon 5 (Fig. 3B, C). In both human and mouse, we found the gene to be active and to produce both mRNA forms in all tissues examined (Fig. 3A). In all cases, the smaller form was much more abundant in human cells, whereas in mouse, the two mRNAs were present in approximately equivalent amounts. Although the two protein products (25 kDa and 27 kDa) were seen in most mouse tissues, liver and skeletal muscle contained only the 25-kDa form. The specific function of the two forms of WBSCR1 will have to await further analysis. Richter-Cook et al. (1998) had noted that rabbit eIF4H preparations had two distinct pIs (7.8 and 8.5), which may correspond to the presence of the two forms of eIF4H (predicted pIs of 7.1 and 8.1). The WBSCR1 protein and eIF4H both contain RNA recognition motifs (RNP1 and RNP2) found in RNA-binding proteins (Osborne et al. 1996; Richter-Cook et al. 1998). No functional motif was identified within the

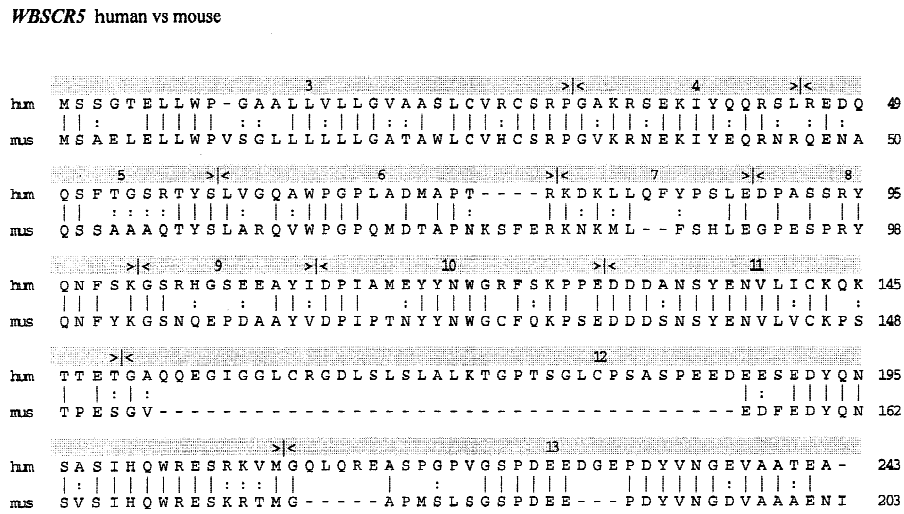

Fig. 4. Alignment of the predicted WBSCR5 protein products of human (hum) and mouse (mus). Identical amino acids (|) and chemically similar (:) amino acids are noted. The human exons coding for the each section of the protein are indicated in the shaded region above the alignment. Note that the corresponding mouse exon would be one less because of the absence of exon 2 in mouse.

Table 2. Distribution of repetitive elements.

	Human (179,196 bp)		Mouse (114,898 bp)	
	Number of elements	Percentage of sequence	Number of elements	Percentage of sequence
SINES:	335	46.48	299	31.70
Alu	314	45.38	—	—
MIR	21	1.1	—	—
B1	—	—	145	13.59
B2	—	—	63	8.97
B4	—	—	71	8.07
ID	—	—	18	0.96
LINES:	22	2.91	4	1.85
LINE1	14	2.45	4	1.85
LINE2	7	0.38	0	0
LTR elements:	32	6.30	9	1.65
MaLR	8	1.39	8	1.34
Retroviruses	7	1.39	1	0.31
MER4_group	14	2.94	0	0
DNA elements:	10	1.50	4	0.63
MER1 type	4	0.76	3	0.55
MER2 type	3	0.43	1	0.09
Mariners	1	0.03	0	0
Small RNA	2	0.06	3	0.21
Simple repeats	43	1.36	52	2.79
Low complexity	19	0.52	12	0.44
Total repeats		59.33		41.54

The distribution of human and mouse repetitive elements over the mouse region contained in Bac146D5 and the syntenic region in human (the region shown in Fig. 1B minus the first 20 kb containing the ELN gene).

alternatively spliced exon 5 of WBSCR1. We determined the location of the protein product of WBSCR1 to be cytoplasmic, frequently perinuclear, and sometimes intranuclear, which is consistent with its role in mRNA binding and utilization in the initiation of RNA synthesis. It is thought that eIF4H may influence the competition between different mRNAs for translation (Richter-Cook et al. 1998). If so, lower amounts of WBSCR1 in WS patients may cause a change in the proteins that get preferentially translated.

We originally predicted the existence of WBSCR5 from the presence of an EST sequence (H23560) adjacent to RFC2 (Osborne et al. 1996; originally called WSCR5). A larger sequence of the WBSCR5 transcript (AK002099) recently deposited into the databanks was found to contain the H23560 and H23535 EST sequences. Thus, the two EST sequences that originally suggested two genes (WBSCR5 and WSCR6; Table 3 in Osborne et al. 1996) has been shown to be one gene. We (X. Chen and V. Duronio) obtained and sequenced the mouse cDNA representing *Wbscr5* while studying genes that are induced in hemopoietic cells and up-regulated in response to cytokines. *Wbscr5* was found to be particularly highly expressed in mast cells (manuscript in preparation). An analysis of WBSCR5 transcription in human and mouse (Fig. 3A) indicates that it is present in all tissues examined although at differing levels. The human transcript was found to be alternately spliced (Fig. 3Cb), with most transcripts missing a non-coding, Alu-containing exon 2 (189 bp) or exon 2a (a shortened 119-bp version of exon 2). Evidence of a corresponding sequence to this exon 2 has not been found in mouse *Wbscr5* transcripts. Since translation is predicted to begin in exon 3 of human WBSCR5 (exon 2 of mouse *Wbscr5*), this alternate splicing is not expected to affect the sequence of the protein product (Fig. 4). No similarity was found between the predicted WBSCR5 product and other proteins in the databases. Although the predicted proteins of mouse and human differ by 40 amino acids, their pI values are virtually identical (4.63 and 4.59 respectively; Table 1) and the program PSORT (http://psort.nibb.ac.jp/) predicted both mouse and human proteins to be nuclear with an equal probability. The function of WBSCR5 and its possible associations with WS await further investigation.

Chromosomal environment. Based on the 3rd codon position nucleotide content from the elastin gene, 7q11.23 was suggested to be part of a gene-rich, GC-rich, H3 isochore (Duret et al. 1995). The region of 7q11.23 that we have examined has an overall GC content of 50.1% and could, therefore, be considered to be part of an H2 isochore. Isochores are extended genomic regions of 100–300 kb with similar GC content (reviewed by Bernardi 1995). The two richest isochores are H3 (%GC >53%) and H2 (%GC from >47% to 52%) and are associated with a higher density of genes. The prediction of one gene per 25–50 kb in an H2 isochore (Gardiner 1996) is met in this region. High Alu content is also typical in regions with a very high GC content (Bernardi 1995; Chen et al. 1996; Flint et al. 1997). There is a general trend in which Alu elements, which are themselves rich in GC, become more abundant as the GC richness of the genomic region increases (Mazzarella and Schlessinger 1997). The number of repeats found in this region of 7q11.23 (45.4% Alu; 59.3% total repetitive elements) is, however, substantially higher than the average seen for H3 isochores (approximately 20% Alu, 35% total repetitive elements) and the average (14% Alu) seen in the genome overall (Duret et al. 1995). For example, our sequence has a GC content similar to that found in a 225-kb gene-rich cluster on Chr 12p13 and its syntenic region on mouse Chr 6 (Ansari-Lari et al. 1998; 17 genes; human 51.35%; mouse 49.68% GC). However, the number of repetitive elements is significantly lower in the 12p13 region (33.36% human, 26.39% mouse) than in our region. This abun-

dance of similar elements creates a high potential for recombination and may partially explain the volatility of this region. Recombination between Alu repeats has been implicated in deletions of LIMK1 (Frangiskakis et al. 1996) and ELN (Olson et al. 1995).

Traditional dot-plot display of human (x axis) and mouse (y axis) reveals that there are no duplicated or inverted segments in the syntenic regions (Fig. 2A). The dot-plot shows significant local alignments between mouse and human as a diagonal line and further shows extra sequence in the human genome between LIMK1 and WBSCR1 (51 kb vs 17 kb for the mouse). This region in the human was found to have an extremely high level of repeat elements (83%, including 60% SINEs and 17% LTR elements). A survey of a percent identity plot (PIP) (Schwartz et al. 2000) allows one to see the likely locations of coding exons and important regions of control (Fig. 2B). All of the *Limk1*, *Wbscr1*, *Wbscr5*, and *Rfc2* coding exons were identified. In addition, other regions of sequence conservation, particularly in the promoter region and in the 3′ untranslated region of some of the genes in this study, are identified and can be targeted for further study.

There are still only a few human-mouse sequence comparisons that cover more than 100 kb. These comparative studies can provide unique evolutionary insights, particularly for regions susceptible to breakage and deletion such as the WS region. Human-mouse genomic sequence comparisons also allow the identification of conserved regions and aid in the identification of genes and their regulatory elements. The combined use of EST databases and conserved regions have been particularly helpful in identifying new genes and alternately spliced exons in the WS region. Coding exons have been shown to be clearly visible when the syntenic sequence from mouse and human are compared by PIP analysis, and genes not detected when the sequence from only one organism is analyzed can become obvious when two are compared (Fig. 2; Ansari-Lari et al. 1998). No evidence was found in this study for genes other than the three previously identified (LIMK1, WBSCR1, and RFC2) and the novel gene, WBSCR5. We have not found evidence for WSCR2, identified through cDNA selection and described as being between LIMK1 and WBSCR1 (Osborne et al. 1996), although there is evidence for background transcription of repeat elements in this region (see Results). Furthermore, WSCR6 is not a separate gene, but part of WBSCR5. Continued mapping and sequencing of the mouse WS region will allow the generation of mouse models of WS by creating mice lacking one or more of the genes located in the commonly deleted region.

Acknowledgments. We thank Ute Rink, Linda McKinnell, and Joanne Whitehead for their expert technical assistance in these studies. We would also like to acknowledge Drs. Lucy Osborne, Lap-Chee Tsui, and Stephen Scherer (University of Toronto) for supplying the human cosmids sequenced in this study and for initially directing us to this region. This work was supported in part by grants to B.F. Koop from the BC Health Research Foundation and the Medical Research Council of Canada, a grant to V. Duronio from the Medical Research Council of Canada, and by a grant to R.D. Burke from the Heart and Stroke Foundation of BC and Yukon.

References

Ansari-Lari MA, Oeltjen JC, Schwartz S, Zhang Z, Muzny DM et al. (1998) Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. Genome Res 8, 29–40

Bellugi U, Bihrle A, Jernigan T, Trauner D, Doherty S (1990) Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. Am J Med Genet Suppl 6, 115–125

Bellugi U, Adolphs R, Cassady C, Chiles M (1999a) Towards the neural basis for hypersociability in a genetic syndrome. Neuroreport 10, 1653–1657

Bellugi U, Lichtenberger L, Mills D, Galaburda A, Korenberg JR (1999b) Bridging cognition, the brain and molecular genetics: evidence from Williams syndrome. Trends Neurosci 22, 197–207

Bernardi G (1995) The human genome: organization and evolutionary history. Annu Rev Genet 29, 445–476

Botta A, Novelli G, Mari A, Novelli A, Sabani M et al. (1999) Detection of an atypical 7q11.23 deletion in Williams syndrome patients which does not include the STX1A and FZD3 genes. J Med Genet 36, 478–480

Chen EY, Zollo M, Mazzarella R, Ciccodicola A, Chen CN et al. (1996) Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. Hum Mol Genet 5, 659–668

Cullmann G, Fien K, Kobayashi R, Stillman B (1995) Characterization of the five replication factor C genes of *Saccharomyces cerevisiae.* Mol Cell Biol 15, 4551–4671

DeBry RW, Seldin MF (1996) Human/mouse homology relationships. Genomics 33, 337–351

DeSilva U, Massa H, Trask BJ, Green ED (1999) Comparative mapping of the region of human chromosome 7 deleted in Williams syndrome. Genome Res 9, 428–436

Duret L, Mouchiroud D, Gautier C (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores. J Mol Evol 40, 308–317

Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K et al. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nat Genet 5, 11–16

Flint J, Thomas K, Micklem G, Raynham H, Clark K et al. (1997) The relationship between chromosome structure and function at a human telomeric region. Nat Genet 15, 252–257

Frangiskakis JM, Ewart AK, Morris CA, Mervis CB, Bertrand J et al. (1996) LIM-kinase1 hemizygosity implicated in impaired visuospatial constructive cognition. Cell 86, 59–69

Gardiner K (1996) Base composition and gene distribution: critical patterns in mammalian genome organization. Trends Genet 12, 519–524

Harrison SD, Solomon N, Rubin GM (1995) A genetic analysis of the 63E-64A genomic region of *Drosophila melanogaster:* identification of mutations in a replication factor C subunit. Genetics 139, 1701–1709

Hockenhull EL, Carette MJ, Metcalfe K, Donnai D, Read AP et al. (1999) A complete physical contig and partial transcript map of the Williams syndrome critical region. Genomics 58, 138–145

Keating MT (1997) On the trail of genetic culprits in Williams syndrome. Cardiovasc Res 36, 134–137

Mari A, Amati F, Conti E, Bengala M, Novelli G (1998) A highly polymorphic CA/GT repeat (LIMK1GT) within the Williams syndrome critical region. Clin Genet 53, 226–227

Mazzarella R, Schlessinger D (1997) Duplication and distribution of repetitive elements and non-unique regions in the human genome. Gene 205, 29–38

Meng X, Lu X, Li Z, Green ED, Massa H et al. (1998) Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. Hum Genet 103, 590–599

Mizuno K, Okano I, Ohashi K, Nunoue K, Kuma K et al. (1994) Identification of a human cDNA encoding a novel protein kinase with two repeats of the LIM/double zinc finger motif. Oncogene 9, 1605–1612

Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL (1988) Natural history of Williams syndrome: physical characteristics. J Pediatr 113, 318–326

Nickerson E, Greenberg F, Keating MT, McCaskill C, Shaffer LG (1995) Deletions of the elastin gene at 7q11.23 occur in approximately 90% of patients with Williams syndrome. Am J Hum Genet 56, 1156–1161

Oberholtzer JC, Cohen EL, Davis JG (1994) Molecular cloning of a chick cochlea cDNA encoding a subunit of DNA replication factor C/activator 1. DNA Cell Biol 13, 857–863

Okano I, Hiraoka J, Otera H, Nunoue K, Ohashi K et al. (1995) Identification and characterization of a novel family of serine/threonine kinases containing two N-terminal LIM motifs. J Biol Chem 270, 31321–31330

Olson TM, Michels VV, Urban Z, Csiszar K, Christiano AM et al. (1995) A 30 kb deletion within the elastin gene results in familial supravalvular aortic stenosis. Hum Mol Genet 4, 1677–1679

Osborne LR, Martindale D, Scherer SW, Shi XM, Huizenga J et al. (1996) Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. Genomics 36, 328–336

Osborne LR, Herbrick JA, Greavette T, Heng HH, Tsui LC et al. (1997a) PMS2-related genes flank the rearrangement breakpoints associated with Williams syndrome and other diseases on human chromosome 7. Genomics 45, 402–406

Osborne LR, Soder S, Shi XM, Pober B, Costa T et al. (1997b) Hemizy-

gous deletion of the syntaxin 1A gene in individuals with Williams syndrome. Am J Hum Genet 61, 449–452

Pan ZQ, Chen M, Hurwitz J (1993) The subunits of activator 1 (replication factor C) carry out multiple functions essential for proliferating-cell nuclear antigen-dependent DNA synthesis. Proc Natl Acad Sci USA 90, 6–10

Peoples R, Perez-Jurado L, Wang YK, Kaplan P, Francke U (1996) The gene for replication factor C subunit 2 (RFC2) is within the 7q11.23 Williams syndrome deletion. Am J Hum Genet 58, 1370–1373

Perez Jurado LA, Peoples R, Kaplan P, Hamel BC, Francke U (1996) Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. Am J Hum Genet 59, 781–792

Richter-Cook NJ, Dever TE, Hensold JO, Merrick WC (1998) Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H. J Biol Chem 273, 7579–7587

Robinson WP, Waslynka J, Bernasconi F, Wang M, Clark S et al. (1996) Delineation of 7q11.2 deletions associated with Williams-Beuren syndrome and mapping of a repetitive sequence to within and to either side of the common deletion. Genomics 34, 17–23

Rogers GW, Jr., Richter NJ, Merrick WC (1999) Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. J Biol Chem 274, 12236–12244

Rosenblatt J, Mitchison TJ (1998) Actin, cofilin and cognition [news; comment]. Nature 393, 739–740

Rowen L, Koop BF (1994) In: Adams MD, Fields C, Venter JC (eds) Automated DNA sequencing and analysis. (London, San Diego: Academic Press)

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C et al. (2000) PipMaker—a web server for aligning two genomic DNA sequences. Genome Res 10, 577–586

Tassabehji M, Metcalfe K, Fergusson WD, Carette MJ, Dore JK et al. (1996) LIM-kinase deleted in Williams syndrome. Nat Genet 13, 272–273

Tassabehji M, Metcalfe K, Donnai D, Hurst J, Reardon W et al. (1997) Elastin: genomic structure and point mutations in patients with supravalvular aortic stenosis. Hum Mol Genet 6, 1029–1036

Tassabehji M, Metcalfe K, Karmiloff-Smith A, Carette MJ, Grant J et al. (1999) Williams syndrome: use of chromosomal microdeletions as a tool to dissect cognitive and physical phenotypes. Am J Hum Genet 64, 118–125

Urban Z, Helms C, Fekete G, Csiszar K, Bonnet D et al. (1996) 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. Am J Hum Genet 59, 958–962

Wang YK, Samos CH, Peoples R, Perez-Jurado LA, Nusse R et al. (1997) A novel human homologue of the Drosophila frizzled wnt receptor gene binds wingless protein and is in the Williams syndrome deletion at 7q11.23. Hum Mol Genet 6, 465–472

Wu YQ, Sutton VR, Nickerson E, Lupski JR, Potocki L et al. (1998) Delineation of the common critical region in Williams syndrome and clinical correlation of growth, heart defects, ethnicity, and parental origin. Am J Med Genet 78, 82–89