Human Molecular Genetics, 1998, Vol. 7, No. 3 325–334

ARTICLE

A duplicated gene in the breakpoint regions of the 7q11.23 Williams–Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK

Luis A. Pérez Jurado[1,2,3], Yu-Ker Wang[4], Risa Peoples[3], Antonio Coloma[2], Jesús Cruces[2] and Uta Francke[3,4,*]

[1]Unidad de Genética, Hospital Niño Jesús and [2]Departimento de Bioquímica, Facultad de Medicina, Universidad Autónoma, Madrid, Spain and [3]Department of Genetics and [4]Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, CA 94305, USA

Received October 6, 1997; Revised and Accepted December 8, 1997 DDBJ/EMBL/GenBank accession nos AF036613, AF035737

Williams–Beuren syndrome (WBS) is a neurodevelopmental disorder with multisystemic manifestations caused by heterozygosity for a partial deletion of chromosome band 7q11.23. The breakpoints cluster within regions located ~1 cM either side of the elastin (ELN) locus. We have characterized a duplicated region near the common deletion breakpoints, which includes a transcribed gene. The centromeric (C) and telomeric (T) copies are almost identical in the duplicated 3′ portions but diverge at their 5′-ends. C-specific 4.3 kb mRNA and T-specific 5.4 kb mRNA are widely expressed in embryonic and adult tissues. The telomeric gene gives rise to several alternatively spliced forms and is deleted in all WBS individuals who have documented ELN deletions. Database searches revealed that this gene encodes BAP-135, a protein phosphorylated by Bruton's tyrosine kinase in B cells, as well as the multifunctional transcription factor TFII-I, hence the gene name *GTF2I*. The centromeric gene is not deleted in WBS and appears to be a partially truncated expressed pseudogene with no protein product (gene name *GTF2IP1*). Both loci map to different genomic clone contigs that also contain other deleted and non-deleted loci. A probe from the shared region recognizes a >3 Mb *Not*I junction fragment that is unique to individuals with the WBS deletion. Therefore, the duplicated region containing *GTF2I* and *GTF2IP1* respectively is located close to the deletion breakpoints and may predispose to unequal meiotic recombination between chromosome 7 homologs and/or to intrachromosomal rearrangements. Hemizygosity for *GTF2I* may also contribute to the WBS phenotype.

INTRODUCTION

Williams–Beuren syndrome (WBS) is a multisystemic neurodevelopmental disorder characterized by distinctive facial features, mental disability with unique cognitive and personality profiles, supravalvar aortic and other vascular stenoses, growth retardation and occasional infantile hypercalcemia, caused by haploinsufficiency for genes in the deletion at chromosome band 7q11.23 (1–4). Consistently included within the deleted interval are the loci encoding elastin (ELN), replication factor C subunit 2 (RFC2), LIM kinase-1 (LIMK1), the wnt receptor *Drosophila frizzled* homolog FZD3, the transcript WBSCR1 and syntaxin1A (STX1A), as well as other ESTs and transcription units of unknown function (4–11). This common deletion is present in the majority of WBS individuals studied, with only 4–5% of clinically diagnosed WBS cases not deleted for any of the available markers (5,6,12–14). Rarely, affected parents and children have been reported, but the great majority of WBS cases are sporadic, with an estimated incidence of 1 in 20 000 live births, indicating a very high mutation rate (~0.5×10^{-4}) (3,15,16).

Little is known about the mechanisms predisposing to the frequent mutational events. Genetic and physical mapping data are consistent with clustering of common deletion breakpoints in relatively small genomic regions on both sides of the ELN locus (13,14). Meiotic recombination between polymorphic markers proximal and distal to the deleted interval was documented in most informative families, suggesting unequal crossing over between misaligned homologous regions as the most likely

*To whom correspondence should be addressed at: Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, CA 94305-5323, USA. Tel: +1 650 725 8089; Fax: +1 650 725 8112; Email: francke@cmgm.stanford.edu

mutational mechanism (17,18). In one study no evidence for unequal interchromosomal exchange was found in some cases, indicating that intrachromosomal meiotic or mitotic rearrangements can also be responsible. The WBS deletions may be of paternal or maternal origin, with a slight but not significant increase in maternally derived deletions (5,13,17,18). The discovery of a dinucleotide repeat polymorphism that represents more than one locus (D7S489) in 7q11.23 suggested the presence of low copy number repeats in the region (13,14).

We report here the characterization of partially duplicated sequences from the common breakpoint regions that include a protein coding gene with multiple alternatively spliced potential products. The shortest splice product encodes BAP-135, a protein identified as a target for phosphorylation by Bruton's tyrosine kinase in B cells (19). The sequence is also identical to the recently published cDNAs for a multifunctional DNA binding component of transcription complexes, named SPIN (20) or TFII-I (21). While the genomic duplication may represent a hotspot for the unequal recombination events leading to the WBS deletions, hemizygosity for this ubiquitously expressed transcription factor may contribute to the WBS phenotype.

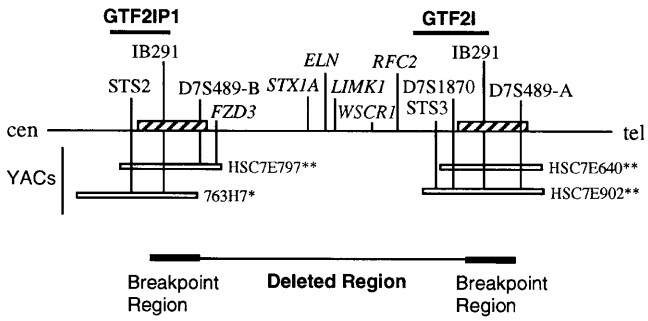


Figure 1. Physical map of the WBS deletion and flanking regions in 7q11.23 (not to scale). Two YAC contigs and corresponding STS markers encompass the deletion breakpoints on both sides. YAC identifiers are those of the original sources, CEPH* (56) and chromosome 7-specific** (57) YAC libraries. Single copy gene loci within the deletion are shown in italics. The estimated extent of the duplicated region is represented by hatched boxes. EST IB291 was mapped to both centromeric and telomeric YACs near the deletion breakpoints. STS2 and STS3 are described in the text and are more precisely mapped in Figure 2a.

RESULTS

A genomic duplication at the common WBS deletion breakpoint regions

As we have previously reported, the breakpoints leading to the WBS deletions cluster in genomic regions flanking the elastin locus. The commonly deleted interval is defined by loci D7S489B (centromeric) and D7S1870 (telomeric), that recombine in 2% of meioses (13). Yeast artificial chromosomes (YACs) containing these loci were identified by PCR typing and assigned to either side of the deletion by their sequence-tagged site (STS) content (Fig. 1). Inter-*Alu* PCR probe hybridizations and restriction mapping revealed high sequence homology between YAC clones from both sides of the deletion interval, suggesting a genomic duplication. An expressed sequence tag (EST), IB291 (22), was mapped to YAC 763H7 in the fifth release of the Whitehead Institute–MIT human genome database (23). By PCR analysis of a somatic cell hybrid panel we confirmed that IB291 maps to the long arm of chromosome 7 (not shown). Southern hybridization of an IB291 probe to somatic hybrid cell lines containing either a normal chromosome 7 or chromosome 7 with a WBS interstitial deletion on a Chinese hamster cell background (7) generated a signal of reduced intensity in the hybrid cell lines that contained the deleted chromosome 7 (not shown). The observed gene dosage difference indicates that IB291 recognizes at least two loci on chromosome 7 and that one of them is lost in the WBS deletion. By PCR and hybridization studies, IB291 was assigned to several YACs located at either side of the ELN locus, confirming its duplicated nature (Fig. 1). Each of these YAC groups also contains a different D7S489 locus that is, therefore, likely to be a part of the genomic duplication. Both YAC contigs contain deletion breakpoints, as previously shown by PCR typing of polymorphic markers in a series of WBS individuals (13).

Cloning, expression pattern and genomic structure of a partially duplicated gene in the breakpoint regions

The cDNA clone containing EST IB291 (22) (GenBank accession no. T03439) was obtained from the ATCC and sequenced. Sequence comparisons using the BLAST software (24) revealed sequence identity with many additional ESTs from dbEST. By sequence walking and computer assembling of multiple EST sequences (25), two composite cDNA variants diverging in the 5′-region were identified (Fig. 2a). IMAGE clone 147504 (GenBank accession nos. R81199 and R81308), containing the longest 5′-region, was purchased from Research Genetics and its full sequence verified. STSs specific for each divergent 5′-region (STS2 and STS3; Fig 2a) were developed and used to map each copy to either the centromeric (*GTF2IP1*) or the telomeric YAC clones (*GTF2I*) (Fig. 2b). Sequence comparisons indicate that the *GTF2IP1* and *GTF2I* cDNAs share 3.2 kb of their 3′-regions, while the 5′-regions of each gene are unique.

Northern blot analysis with the 1.3 kb IB291 cDNA insert as probe revealed two transcripts of ~4.3 and 5.4 kb in all tissues tested (Fig. 3a, left). Hybridization with a 840 bp cDNA probe generated by PCR from the *GTF2I*-specific 5′-region identified the 5.4 kb transcript only (Fig. 3a, right). Thus, *GTF2I* and *GTF2IP1* seem to be expressed as different sized mRNAs.

To further characterize the gene copy-specific transcripts, we used primers from the differential 5′-regions (primer A from *GTF2IP1* and primer C from *GTF2I*) (Fig. 2a) along with different reverse primers from the duplicated region to PCR amplify and sequence gene copy-specific cDNAs. By using a fetal brain cDNA library or reverse transcribed RNA as templates, a total sequence of 4.5 kb was obtained for *GTF2I* and of 3.4 kb for *GTF2IP1*. Only three nucleotide changes were observed in *GTF2IP1* cDNA compared with *GTF2I* cDNA in the 3.2 kb duplicated region: C1487T, G2217A and C3130T (*GTF2I* nucleotide numbers as in Fig. 4a). Therefore, these regions have 99.9% nucleotide identity within the cDNA sequence.

Reverse transcriptase (RT)-PCR analysis using primers C and M revealed four mRNA variants of *GTF2I*. Sequence analysis revealed that the different mRNAs resulted from alternative splicing of two exons located upstream of the duplicated region (Figs 3b and 5). The longest *GTF2I* transcript (band 1) is strongly present in fetal brain (not shown), weakly in adult brain, muscle and lymphoblasts and is almost undetectable in other adult tissues, while the other variant transcripts are equally expressed

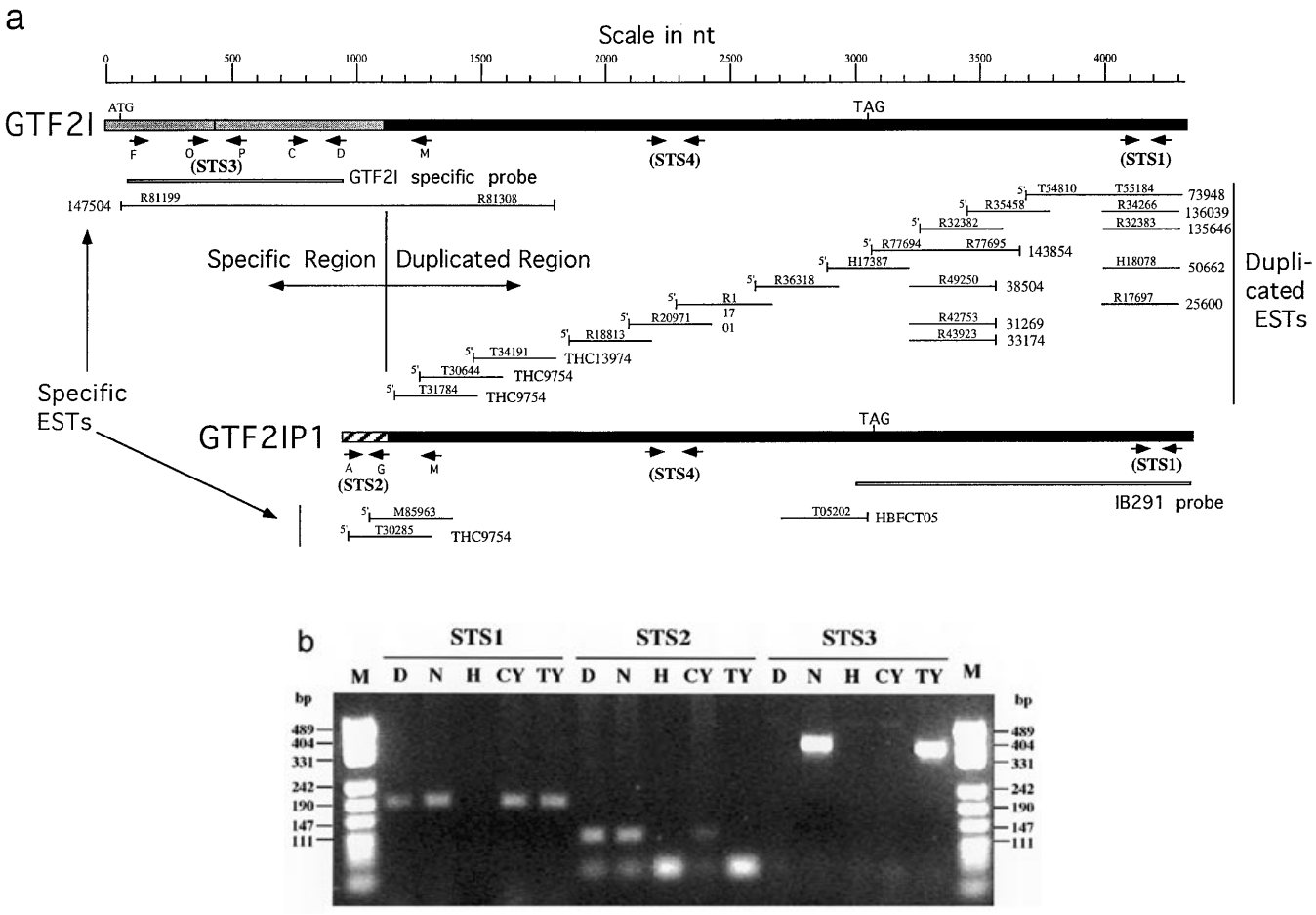


Figure 2. (**a**) Characterization of two highly related genes from the duplicated region, now named *GTF2I* and *GTF2IP1*. A contig was assembled of overlapping EST sequences, shown with their dbEST identifiers. *GTF2I* (telomeric) and *GTF2IP1* (centromeric) are two cDNAs that diverge in their 5′-regions. The part unique to *GTF2I* is shaded and the unique portion of *GTF2IP1* is hatched. Location and direction of PCR primers developed from cDNA sequences are indicated by arrows.Vertical bars between primers denote the presence of introns at STS3 and STS4, but not at STS1 and STS2. Probes used for hybridization are the *GTF2I*-specific probe, detecting only the larger telomeric gene, and the cDNA clone IB291, detecting both genes. (**b**) Mapping of the *GTF2I* and *GTF2IP1* loci with respect to the WBS deletion. Three different STSs were studied in Chinese hamster × human somatic cell hybrid lines, containing either the deleted (D) or non-deleted (N) chromosome 7, in Chinese hamster (H) control cells and in two YAC clones, containing either the centromeric (CY) or the telomeric (TY) deletion breakpoints. The results reveal the existence of at least two copies of the 3′-region (STS1), each located in different YACs. STS2, derived from the 5′-exon specific for *GTF2IP1*, is located in the centromeric YAC and outside the deletion, as it is present in both hybrid cell lines. STS3, derived from the 5′-region specific for *GTF2I*, is located in the telomeric YAC and is deleted from the D hybrid cell line. No amplification was observed in hamster DNA for any of the STSs.

in all adult tissues (Fig. 3b). With different primer combinations no variation in RT-PCR products was obtained for other regions of the available *GTF2I* cDNA. With the *GTF2IP1*-specific primer A and different reverse primers no alternative splicing products were seen, only a single product being expressed in most tissues (Fig. 3b). RT-PCR analysis of RNA from somatic hybrid cell lines containing the deleted (D) or non-deleted (N) chromosome 7 confirmed that *GTF2IP1* is not deleted (primer pair A+M) while *GTF2I* is deleted (primer pairs C+M and F+D) (Fig. 3c).

To establish the genomic structure of each of the duplicated loci, we identified several cosmid clones by hybridizing chromosome 7-specific cosmid library filters with two *GTF2I* probes (26). PCR analysis with locus-specific primers allowed us to assign the positive cosmids to either the centromeric (*GTF2IP1*) or the telomeric (*GTF2I*) locus. Exon–intron boundaries were identified by comparison of PCR amplifications from cDNA and genomic DNA with various combinations of primers designed from the cDNA sequence (primer sequences available upon request). The duplicated region contains 21 exons and spans ~30 kb of genomic DNA (Fig. 5). Exon and intron sizes are identical in *GTF2IP1* and *GTF2I* and the sequenced intron–exon junctions follow the canonical AG-GT rules in both genes. Only one exon that is specific for the 5′-region of *GTF2IP1* has been identified. The genomic structure of the non-duplicated 5′-region of *GTF2I* has been determined by primer walking because of the large size of several introns. It contains at least 12 exons and spans >70 kb of genomic DNA.

PCR amplification and Southern blot hybridization of genomic DNAs with several *GTF2I* primers and probes revealed the presence of cross-hybridizing signals corresponding to related sequences outside the WBS deletion. During BLAST searches we have identified ESTs with some sequence divergence from

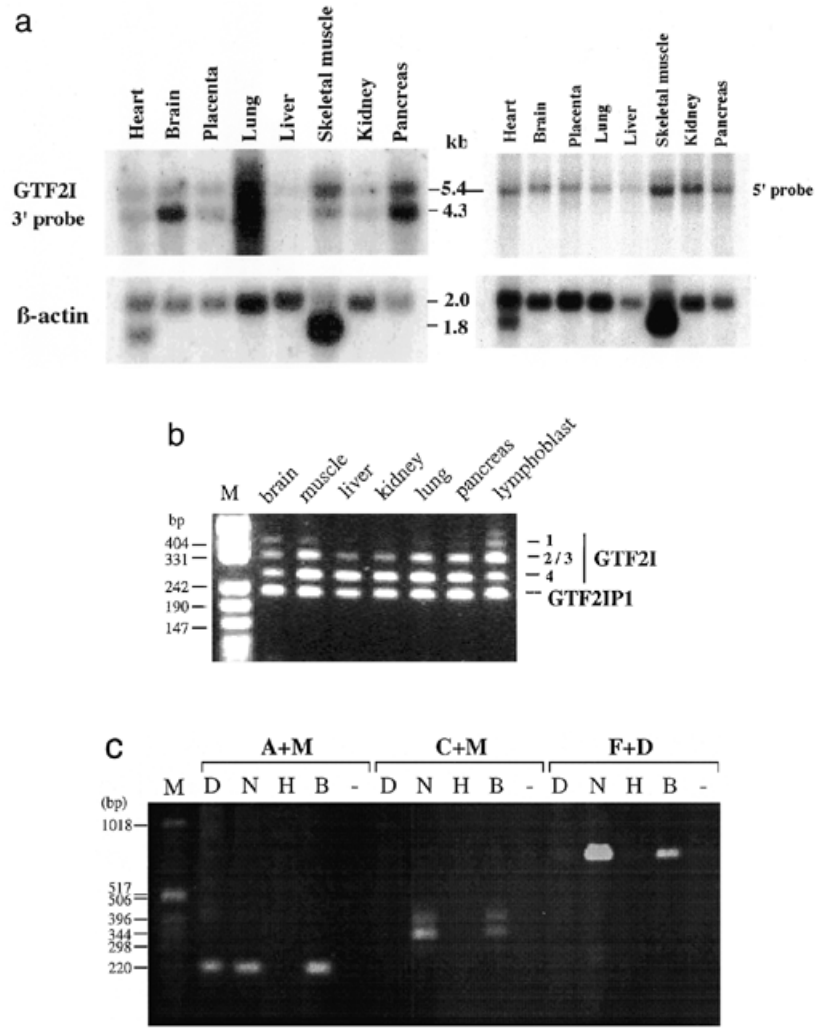


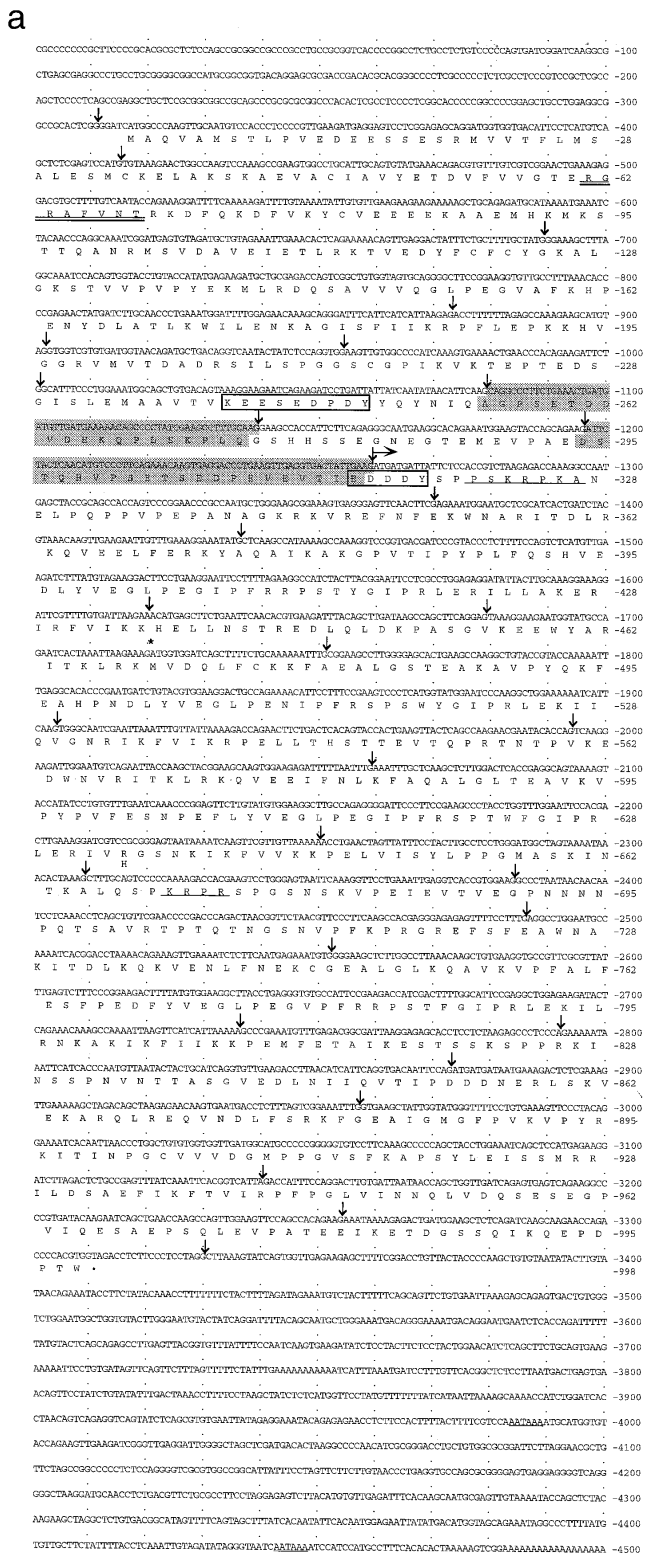
Figure 3. (**a**) Northern blot hybridization. With the 3′ probe (IB291) two transcripts of 4.3 and 5.4 kb were detected in all adult human tissues (left). A probe from the 5′-end of the *GTF2I* gene hybridized only to the 5.4 kb transcript (right). The filters were re-hybridized with β-actin cDNA as control. (**b**) RT-PCR analysis of *GTF2I* and *GTF2IP1* transcripts in embryonic and adult tissues. With forward primers A (*GTF2IP1*-specific) and C (*GTF2I*-specific) and reverse primer M (common to both cDNAs) the existence of both types of mRNAs was confirmed in all tissues tested. Alternative splicing of *GTF2I*-specific products was observed: band 1 contained both exons A and B and was most evident in brain and lymphoblasts, the co-migrating bands 2 and 3 contained either exon A or B respectively, by sequencing, and band 4 lacked both exons. Only a single amplification product corresponding to *GTF2IP1* mRNA was obtained from all tissues. (**c**) Mapping of *GTF2I* and *GTF2IP1* transcripts with respect to the WBS deletion by RT-PCR analysis of somatic cell hybrid lines. Primers A+M amplified the *GTF2IP1*-specific transcript from the deleted (D) and non-deleted (N) chromosomes 7, while *GTF2I* mRNAs were detected only in the hybrid cell line with the non-deleted (N) chromosome 7 with the gene-specific primer pairs C+M and F+D. Both types of RT-PCR products were amplified from brain tissue (B) and no amplification was observed from hamster RNA (H) or in control samples lacking RNA (–).

GTF2I (90–95% identity), which may correspond to these additional copies or related genes. Preliminary data reveal that copies of related genes exist on chromosome 7 as well as on several other chromosomes. Based on analysis of genomic DNA from several individuals we believe that the number of additional copies may be variable in the population (data not shown).

Transcripts and predicted proteins encoded by *GTF2I*

The longest *GTF2I* cDNA has a 2994 nt (nt 316–3309) open reading frame that predicts a 998 amino acid protein (predicted molecular mass 112.4 kDa) (Fig. 4a). The putative initiator methionine codon is within a context favorable for translation initiation (27). The discrepancy between the cloned cDNA (4.5 kb) and the transcript size determined by northern blot (5.4 kb) must be due to additional unidentified exons in the 5′-untranslated region. Attempts to extend the cDNA further by use of 5′-RACE were unsuccessful.

Alternative splicing of two exons immediately upstream of the duplicated region maintains the reading frame in all combinations and predicts the generation of four protein variants. The second and shortest mRNA variant, lacking exons A and B (Fig. 5), encodes BAP-135, a recently identified 957 amino acid novel protein (observed molecular mass 135 kDa, predicted molecular mass 107.8 kDa), which is a target for phosphorylation by Bruton's tyrosine kinase (Btk) in response to B cell receptor engagement (19). It has been proposed that the discrepancy between the calculated and observed molecular masses may be



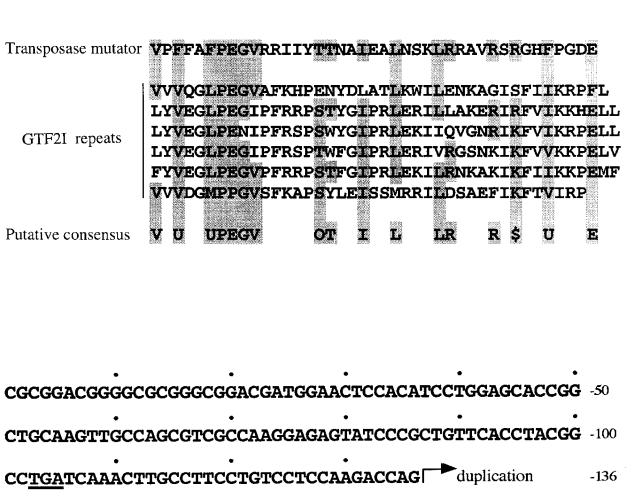


Figure 4. (**a**) Nucleotide sequence and predicted protein product of the longest *GTF2I* cDNA. Regions that are alternatively spliced out are shaded. The putative nuclear targeting signals are underlined, the putative RNP RNA binding domain is double underlined and the two tyrosine phosphorylation sites are boxed. Potential polyadenylation signals in the 3′-untranslated region are underlined. The vertical arrows indicate the exon–intron junctions. The start of the duplicated region that is also present in *GTF2IP1* cDNA is indicated by a horizontal arrow. The potential initiation codon of the *GTF2IP1* ORF is labeled with an asterisk and the single amino acid difference with respect to GTF2I (R634H substitution) is indicated. (**b**) Amino acid sequence alignment of a region from a mutator transposase (IsRm3 from *Rhizobium meliloti*) (top) and the C-terminal portions of the six GTF2I repeats as identified by BLOCK motif search (31). Identical or similar residues are highlighted. In the putative consensus an amino acid code letter indicates identity of the transposase residue with two or more of the GTF2I repeats. Conserved substitutions include: U, bulky hydrophobic residue (I, L, V, M, F, Y); O, small hydroxyl residue (S, T); $, positively charged basic residue (K, R). (**c**) Nucleotide sequence of the 5′ *GTF2IP1*-specific region. The in-frame stop codon is underlined.

due to the high percentage of acidic residues. *In vitro* expression of a cloned cDNA that used the predicted initiation codon as the initiator methionine resulted in a protein of the same molecular mass as endogenous BAP-135, suggesting that the proposed initiation site is the one used *in vivo* (19). Five amino acid differences between the published BAP-135 sequence and ours are likely due to sequencing errors, as our *GTF2I* sequence is identical to that for SPIN (20; GenBank accession no. Y14946). A third mRNA species (lacking exon B) encodes a 977 amino acid protein (predicted molecular mass 110.1 kDa) and a fourth mRNA (lacking exon A) encodes a 978 amino acid protein (predicted molecular mass 110.2 kDa). Interestingly, a protein of apparent 140 kDa was also found as a target for Btk phosphorylation and also reacts with antibodies raised against BAP-135 (19). This protein is likely to be the translation product of one of the larger *GTF2I* mRNA species produced by alternative splicing in lymphoblasts. Of note, there is a significant difference in the isoelectric points between the predicted GTF2I protein isoforms, dropping from 8.23 for BAP-135 to 6.09 for the largest 998 amino acid variant that includes both exons A and B.

Prediction of known sequence motifs using the PSORT program (28) suggests two putative nuclear targeting signals in all four potential *GTF2I*-encoded proteins. A PROSITE search revealed an RNA binding motif near the N-terminus as well as potential targets for tyrosine kinase phosphorylation (29; Fig. 4a). There are six amino acid repeats within the protein sequence from which a consensus sequence has been derived (19). With the BLOCKS algorithm (30) we detected a similarity between the six polypeptide repeats of the *GTF2I* gene and a putative transposase

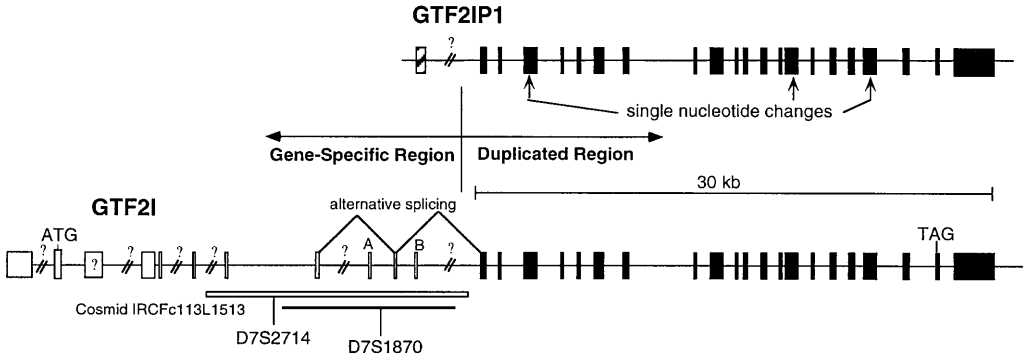


Figure 5. Comparative genomic structures of *GTF2IP1* (top) and *GTF2I* (bottom). Duplicated exons are represented as filled boxes, the *GTF2IP1*-specific exon as a hatched box and *GTF2I*-specific exons as open boxes. Intron sizes in the duplicated region are drawn to approximate the 30 kb scale, but exon sizes are enlarged to show their sizes relative to each other. The positions of the three nucleotide differences in the coding sequences of the duplicated region are marked by arrows. In the *GTF2I*-specific region question marks within introns indicate unknown size and a question mark within an exon means that an additional intron may exist. Exons A and B are alternatively spliced out, generating four different types of *GTF2I* transcripts. A single *GTF2I*-specific cosmid contained D7S1870, commonly deleted in WBS patients (32), as well as D7S2714 (sequence-tagged sites developed at the Stanford Human Genome Center, available on line at http://shgc-www.stanford.edu), indicating that these STS markers are within introns of *GTF2I*.

of the *Mutator* group of bacterial insertion sequences and transposable elements in maize (31; Fig. 4b).

The shorter *GTF2IP1*-specific cDNA has an in-frame stop codon within the unique 5′-region (Fig. 4c). The three nucleotide changes in the duplicated region, compared with *GTF2I*, cause an additional in-frame termination codon at amino acid 391 of the GTF2I sequence (Fig. 4a), a missense substitution (R634H) and a silent change in the wobble position. There is a potential translation initiation codon downstream of the second in-frame stop codon followed by an open reading frame that would encode a 530 amino acid protein almost identical to the C-terminus of GTF2I with the single R634H substitution (Fig. 4a). This potential initiator codon has a poor Kozak consensus (27). In addition, an antibody raised against a BAP-135 peptide that should be present in this hypothetical *GTF2IP1*-encoded protein product failed to detect any smaller proteins (19). This information suggests that *GTF2IP1* mRNA, despite being highly expressed and processed, does not get translated into a mature protein, at least not in the tissues examined.

GTF2I is deleted in WBS while *GTF2IP1* is not

PCR typing of several *GTF2I*-specific STSs in hybrid cells with chromosome 7 bearing the typical WBS deletion showed lack of amplification products, while *GTF2IP1*-specific products and fragments common to both loci did amplify (Fig. 2b). Furthermore, ICRF cosmid clone c113L1513, spanning several internal *GTF2I* exons (Fig. 5), was found to contain two STSs, D7S2714 (sequence-tagged sites developed at the Stanford Human Genome Center, available on line at http://shgc-www.stanford.edu) and D7S1870, which must, therefore, lie within introns of *GTF2I*. Since D7S1870 is usually deleted in WS patients (13,14,32), this region of *GTF2I* falls within the commonly deleted interval.

The few known sequence differences provide tools for further deletion analysis. The G→A change in *GTF2IP1* (R634H) causes the absence of a *Sac*II/*Bst*UI restriction site that is present in *GTF2I*. In the adjacent intron there is an *Rsa*I restriction site in

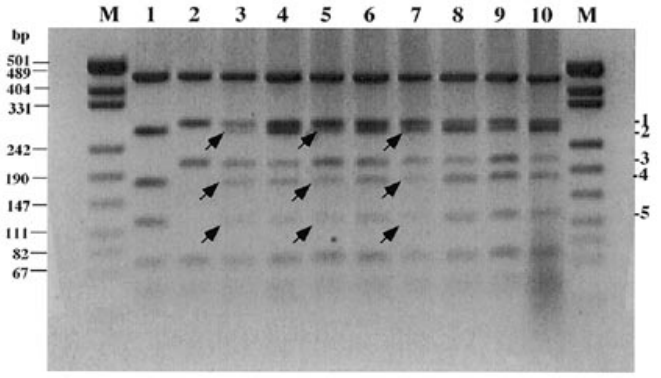


Figure 6. Deletion analysis of the *GTF2I* and *GTF2IP1* loci in WBS patients and controls. PCR products obtained with primers for STS4, which contain exon and intron sequences of the duplicated region, were cleaved with *Bst*UI and *Rsa*I and were separated in a 4% agarose gel. Lanes M contain pUC plasmid DNA digested with *Hpa*II as molecular size marker. Lane 1 contains PCR products from a *GTF2I*-specific cosmid and lane 2 contains PCR products from a *GTF2IP1*-specific cosmid. Lanes 3, 5, 7 and 9 represent WBS individuals and lanes 4, 6, 8 and 10 independent normal controls. A dosage reduction is evident for the *GTF2I*-specific fragments (arrows) in lanes 3, 5 and 7 compared with controls. The patient in lane 9 appears to have a normal dosage ratio in spite of being deleted at locus D7S1870.

GTF2IP1 that is absent in *GTF2I*. Using primers flanking this region for PCR and subsequent restriction digestion of the amplified products with *Bst*UI and *Rsa*I revealed a dosage reduction of the *GTF2I*-specific fragments in most WBS samples with respect to controls (Fig. 6). Two WBS individuals, known to be deleted at D7S1870 (13), did not show dosage reduction. This could be due to the presence of additional related sequences which also amplify and interfere with the analysis. Alternatively, the telomeric deletion breakpoints in these individuals could lie between D7S1870 and the region amplified by PCR near R638H, which are at least 20 kb apart (Fig. 5).

Search for junction fragments

Following digestion with *Not*I and pulsed field gel electrophoresis, novel deletion junction fragments >3 Mbp in size were detected with the IB291 cDNA probe in the hybrid cell lines with the WBS deletion chromosome 7, as well as in four WBS samples. The corresponding fragments in normal control samples and somatic cell hybrids retaining the non-deleted chromosome 7 were ~2.5 Mbp in size (data not shown). Frequently cutting enzymes and other rarely cutting enzymes, however, did not generate novel junction fragments on Southern blots probed with *GTF2I* cDNA, although reduced intensity of the fragments compared with controls suggested a dosage effect. We interpret these results to indicate that the breakpoints leading to the WBS deletions occur within or close to the *GTF2I* and *GTF2IP1* loci. However, because of the extremely high sequence identity, most restriction sites may be identical in both genes. If recombination events leading to WBS deletions occur between aligned highly similar but non-homologous sequences, most of the resulting junction fragments may be identical in size to the original fragments and, therefore, impossible to detect.

DISCUSSION

More than 95% of clinically defined WBS patients have *de novo* deletions, quite homogeneous in size, at chromosomal band 7q11.23 (13,14). We have identified a genomic duplication flanking the commonly deleted interval. The duplication includes ~30 kb of the 3′-region of a transcribed gene, called *GTF2I* (telomeric copy) and *GTF2IP1* (centromeric copy). The duplicated region is >30 kb, however, and encompasses D7S489 loci as well as genes related to EST IB1445, belonging to the PMS2 mismatch repair gene family (33), and other as yet uncharacterized sequences. The very high conservation (99.9%) of the coding sequences of *GTF2I* and *GTF2IP1*, as well as the preservation of restriction fragment patterns at both loci, indicate that the duplication is of recent evolutionary origin. Indeed, as reported elsewhere, there is a single locus, corresponding to *GTF2I*, in mouse (named *Gtf2i*) and probes from this gene also detect single transcripts in rat (34).

The duplicated genes in humans differ in their products and deduced functions. *GTF2I*, at the telomeric repeat, is transcribed and alternatively spliced into several mRNA species with long ORFs that encode various protein isoforms, one of which is identical to BAP-135 (19), SPIN (20) and TFII-I (21). *GTF2IP1*, in the centromeric repeat, is transcribed into a shorter mRNA which is apparently not translated. We conclude that the telomeric repeat contains a functional gene and the centromeric repeat a truncated pseudogene.

De novo germline deletions causing WBS occur at an approximate rate of 0.5×10^{-4}, however, there is little information addressing the mechanisms of these rearrangements. Haplotypes recombinant for segments flanking the deleted region have been demonstrated in most WBS chromosomes studied, indicating that these deletions result from unequal crossing over between the chromosome 7 homologs during gametogenesis (17,18). Recombination events between misaligned repetitive elements have been observed at the deletion and duplication breakpoints in a variety of DNA rearrangements resulting in human genetic disorders. Low copy number repeat sequences have been found dispersed within some chromosomal regions that are subject to spontaneous deletions, such as Xp22.3 (Kallman syndrome; 35), 22q11 (DiGeorge/velo-cardio-facial syndromes; 36), 15q11–13 (Prader–Willi/Angelman syndromes; 37,38) and 5q13 (spinal muscular atrophy; 36). Both the repetitive nature of these sequences and their clustering within a relatively small interval provide a mechanism for aberrant recombination events and, thus, the high frequency of interstitial deletions in these regions. In addition, there are large rearrangements between repeats on the X chromosome leading to steroid sulfatase gene deletions (39) and reciprocal duplication and deletion products of unequal crossing over events on chromosomal band 17p11.2 that are responsible respectively for Charcot–Marie–Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP) (40,41). In these two cases a precise mechanism for homologous unequal recombination has been found to occur between misaligned repeats that are duplicated in the regions flanking the rearrangements (42). After submission of this manuscript a similar mechanism has also been reported for deletion formation in Smith–Magenis syndrome (43). Furthermore, unequal crossing over at homologous genes has led to variation in the copy number of green pigment genes in humans (44) and X-linked color blindness is due to a similar mechanism resulting in complete deletion of the pigment genes or recombination leading to hybrid red and green genes (45).

Our data suggest a similar situation for WBS deletions. Typing of multiple polymorphic markers has previously indicated that deletion breakpoints cluster in regions which, as shown here, are part of a recent evolutionary genomic duplication. A probe from this duplicated region detected large junction fragments of similar size with long range restriction enzymes in five unrelated WBS individuals. Therefore, unequal recombination appeared to occur in a similar manner in multiple unrelated WBS chromosomes, further supporting the idea of a common mechanism for the deletions.

GTF2I is deleted in all informative WBS patients with a documented ELN deletion at 7q11.23. Therefore, these patients are hemizygous for the functional copy of the gene and are expected to produce half-normal dosage of its protein product(s). BAP-135, a *GTF2I*-encoded protein, has been identified as a target of phosphorylation by Btk during B cell receptor engagement (19). It occupies a position immediately downstream of Btk in the B cell receptor activating pathway. Since Btk mutations cause X-linked agammaglobulinemia, mutations or deletions in BAP-135 have been speculated to be associated with a similar immunodeficiency phenotype in the homozygous state (19). However, B cell function appears to be within normal limits in WBS individuals heterozygous for deletions of *GTF2I*/BAP-135. In addition, the expression pattern of *GTF2I* is much broader than that of Btk and includes multiple fetal tissues. Therefore, its functional role is likely to extend beyond signaling in B cells. As suggested, the BAP-135 tyrosine phosphorylation signals could also be targets for tyrosine kinases other than Btk whose expression is B cell specific (19). While Btk is a cytoplasmic protein and Btk–BAP-135 interactions would be expected to occur in the cytoplasm, the presence of putative nuclear localization signals in the GTF2I sequence suggested a function in the nucleus.

Also intriguing is a weak similarity of the six amino acid repeats in GTF2I/BAP-135 to putative transposases from plants and bacteria (31; http://www-leland.stanford.edu/~eisen/Mutator/Mutator.html). Transposases are implicated in some of

the mechanisms of meiotic recombination, such as reciprocal breakage, exchange of DNA segments and rejoining of chromatids (46). Interestingly, a *mariner* transposon-like element is part of the duplicated region that is the recombination hotspot leading to duplication, in CMT1A, and deletion, in HNPP, of a 1.5 Mb region of 17p12–p11.2 (47). This *mariner* element contains a transposase-like sequence with similarity to insect transposases. This analogy initially led us to speculate that *GTF2I* may have transposase-like function and that the duplicated genomic region could be a target for its own protein product which could, therefore, contribute to the unequal recombination events that cause the WBS deletions.

After our manuscript was submitted other functions of this gene were reported independently by two groups. Grueneberg *et al.* (20) cloned and functionally characterized the cDNA for a multifunctional DNA binding protein, termed SPIN, that binds to the c-*fos* promoter and interacts with serum response factor and with the homeobox protein *Phox1*. Studying a quite different experimental system, Roy *et al.* (21) succeeded in cloning the transcription factor TFII-I that binds core promoters at the pyrimidine-rich initiator element (*inr*) as well as at the upstream E box and is an essential component of a transcription factor complex. Both human SPIN and TFII-I cDNA sequences were found to be almost the same as BAP-135 and correspond to the shortest of the four different splice products of *GTF2I* that we identified.

In summary, we have characterized a genomic repeat which includes a duplicated gene, named *GTF2I* and *GTF2IP1*, flanking the interval at 7q11.23 commonly deleted in WBS. Unequal crossing over between chromosomes or chromatids paired at these misaligned repeats appears to be the major mechanism responsible for *de novo* WBS deletions. The recombination occurs in a similar manner in multiple unrelated individuals, resulting in a junction fragment that lacks the telomeric (and functional) locus *GTF2I*. Therefore, hemizygosity for *GTF2I*, that encodes a multifunctional member of a eukaryotic transcription factor complex and is tyrosine phosphorylated by Btk, is a consistent feature in WBS and haploinsufficiency for this protein may contribute to the WBS phenotype. There is a growing list of human multisystem developmental disorders that are caused by haploinsufficiency for transcriptional co-activators or DNA binding proteins with potential function as transcription factors. For example, Rubinstein–Taybi syndrome, characterized by broad thumbs and great toes, mental retardation, agenesis of corpus callosum, typical facial features, eye, skin, skeletal and cardiovascular abnormalities, is caused by deletion or disruption of the transcriptional co-activator CREB binding protein (48,49). Haploinsufficiency for the paired-box DNA binding protein PAX6 results in eye malformations, including aniridia (50), while deletion or inactivating mutations of *PAX3* give rise to a form of Waardenburg syndrome with pigmentary disturbance and hearing loss (51). The *GLI3* gene, a member of GLI–Kruppel zinc finger gene family that represents a potential sequence-specific DNA binding transcription factor, is deleted in patients with Greig cephalopolysyndactyly syndrome (52), while frameshift mutations cause Pallister–Hall syndrome (53). Therefore, when considering the contributions to the phenotype of candidate genes identified in the 2 Mb WBS deletion those with potential functions as transcription factors/co-activators, such as *GTF2I*, are likely to be dosage sensitive and should be taken seriously.

MATERIALS AND METHODS

Clinical samples

Criteria for the clinical diagnosis of WBS in the five individuals included in this study have been reported elsewhere (13). Genomic DNA was isolated from peripheral blood lymphocyte samples or from Epstein–Barr virus-transformed lymphoblastoid cell lines by standard techniques and RNA was isolated using RNA STAT-60 (Tel-Test'B Inc.) as described by the manufacturer.

Somatic cell hybrid mapping

Initial assignment of IB291 was performed by PCR amplification of our panel of rodent × human somatic cell hybrids whose content of human chromosomes has been previously documented by cytogenetic and molecular methods (54). Somatic cell hybrid lines in which the normal chromosome 7 has been separated from the chromosome 7 carrying the interstitial deletion of a WBS patient on a Chinese hamster cell background have been reported (7). Deletion mapping was performed by PCR typing or genomic Southern hybridization using DNA from the hybrid cell lines.

cDNA and genomic clone isolation and characterization

EST sequences were tested for similarity to known sequences in GenBank using the BLAST software (24). cDNA clones were purchased from Research Genetics (clone 147504) and ATCC (no. 86072, containing EST IB291). A 5′ *GTF2I*-specific cDNA probe was generated by PCR from clone 147504 with primers F (5′-ATGTCCACCCTCCCCGTTGA-3′) and D (5′-GGTGGCTT-CCTTGAATGTTA-3′). Larger cDNAs were obtained by PCR-based screening of a human fetal brain cDNA library cloned in λZAPII (Stratagene).

A chromosome 7-specific cosmid library spotted on high density filters (26) was kindly provided by Hans Lehrach (Resource Center, Max Planck Institute for Molecular Genetics, Berlin, Germany). Library screening was performed with random hexamer-labeled cDNA probes and positive clones were obtained from the same source. Multiple oligonucleotide primers were designed based on the cDNA sequence and were used for PCR of cDNA and genomic DNA. Comparison of cDNA and genomic sequences permitted estimation of intron sizes and identification of intron–exon junctions and exon boundaries.

STSs were developed to identify and distinguish *GTF2I* and *GTF2IP1* by PCR (Fig. 2). STS1 (both genes) = IB291 (22,23); STS2 (*GTF2IP1*-specific), 5′-CGGACGATGGAACTCCACAT-3′ and 5′-GACAGGAAGGCAAGTTTGAT-3′ to amplify 108 bp; STS3 (*GTF2I*-specific), 5′-AGGCAAATCCACAGTGGTAC-3′ and 5′-CTACATGCTTCTTTGGCTCT-3′ to amplify 380 bp; STS4 (both genes), 5′-GCTCAAGCTCTTGGACTCAC-3′ and 5′-CAG-GAGGCAAGTAGGAAATA-3′ to amplify 1.2 kb.

Restriction analysis was performed by digestion of 10 µl PCR-amplified products (STS4) from cosmid or genomic WBS and control DNAs with *Rsa*I and *Bst*UI (NEB) according to the manufacturers. The resulting fragments were analyzed by ethidium bromide staining following electrophoresis in 3% agarose gels.

Sequence analysis

Sequencing of cloned DNA was performed on an ABI prism 377 sequencer (Perkin Elmer). Double-strand sequencing of PCR-amplified products was performed using a PCR sequencing kit (US Biochemical). The cDNA sequences have been deposited in GenBank: *GTF2I*, accession no. AF035737; *GTF2IP1* accession no. AF036613. Protein sequence comparisons were carried out with the BLASTP (24), BEAUTY (55) and PROSITE (29) programs and the BLOCKS database version 9.1 (30). Prediction of protein motifs was with PSORT (28).

Expression analysis

A northern blot filter containing poly(A)$^+$ RNA from multiple human tissues was purchased from Clontech and probed according to the manufacturer's instructions. Hybridization probes were generated by PCR from plasmid clones or by RT-PCR from tissues. PCR products were gel purified and radiolabeled with random primers (Pharmacia). The probes were hybridized overnight at 65°C in Church buffer and washed at a final stringency of 0.1× SSC, 0.1% SDS at 65°C.

For RT-PCR studies tissues were obtained from an adult female cadaver, according to an IRB approved human subjects in research protocol and snap frozen in liquid nitrogen. Tissues were dispersed using a polytron homogenizer directly into RNA STAT-60 (Tel-Test'B Inc.) and total RNA was extracted as per the manufacturer. One microgram of total RNA was used for first strand cDNA synthesis, with 200 U SuperScriptII reverse transcriptase (Gibco BRL) and random hexamers, in 20 µl total volume and incubated for 60 min at 42°C. One microliter of the reaction was used for 35 cycles of PCR at 94°C for 1 min, 60°C for 1 min and 72°C for 1 min and a final extension at 72°C for 7 min. Primer sequences were: *GTF2IP1*-specific, A, 5′-CGGACGATGGAACTCCA-CA-3′; *GTF2I*-specific, C, 5′-CCCCATCAAAGTGAAAAC-TG-3′; common to both genes, M, 5′-AGTTGAACTCCCTCA-CTTTC-3′

ACKNOWLEDGEMENTS

We thank Paige Kaplan for WBS samples, Hans Lehrach for cosmid filters and clones, M.Josh Cisco and Pilar Casas for technical assistance, Stephen Scherer and Ananda Roy for prepublication manuscripts and Jonathan Eisen and Virginia Walbot for helpful discussions. The work was supported by grants from the US National Institutes of Health, the Spanish Fondo de Investigaciones de la S.S. and by the Howard Hughes Medical Institute.

REFERENCES

1. Burn,J. (1986) Williams syndrome. *J. Med. Genet*., **23**, 389–395.
2. Morris,C.A., Demsey,S.A., Leonard,C.O., Dilts,C. and Blackburn,B.L. (1988) The natural history of Williams syndrome; physical characteristics. *J. Pediat*., **113**, 318–326.
3. Grimm,T. and Wesselhoeft,H. (1980) The genetic aspects of Williams–Beuren syndrome and the isolated form of the supravalvular aortic stenosis: investigation of 128 families. *Z. Kardiol*., **69**, 168–172.
4. Ewart,A.K., Morris,C.A., Atkinson,D., Jin,W., Sternes,K., Spallone,P., Stock,A.D., Leppert,M. and Keating,M.T. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nature Genet*., **5**, 11–16.
5. Nickerson,E., Greenberg,F., Keating,M.T., McCaskill,C. and Shaffer,L.G. (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.
6. Lowery,M.C., Morris,C.A., Ewart,A., Brothman,L.J., Zhu,X.L., Leonard,C.O., Carey,J.C., Keating,M. and Brothman,A.R. (1995) Strong correlation of elastin deletions, detected by FISH, with Williams syndrome: evaluation of 235 patients. *Am. J. Hum. Genet*., **57**, 49–53.
7. Peoples,R., Pérez Jurado,L.A., Wang,Y.-K., Kaplan,P. and 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.
8. Tassabehji,M., Metcalfe,K., Fergusson,W.D., Carette,M.J.A., Dore,J.K., Donnai,D. and Read,A.P. (1996) LIM-kinase deleted in Williams syndrome. *Nature Genet*., **13,** 272–273.
9. Wang,Y.-K., Harryman Samos,C., Peoples,R., Pérez Jurado,L.A., Nusse,R. and Francke,U. (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.
10. Osborne,L.R., Martindale,D., Scherer,S., Shi,X.M., Huizenga,J., Heng,H.H.Q., Costa,T., Pober,B., Lew,L., Brinkman,J., Rommens,J., Koop,B. and Tsui,L.-C. (1996) Identification of a 500 kb region that is commonly deleted in Williams syndrome patients. *Genomics*, **36**, 328–336.
11. Osborne,L.R., Soder,S., Shi,X.-M., Pober,B., Costa,T, Scherer,S.W. and Tsui,L.-C. (1997) Hemizygous deletion of the syntaxin 1A gene in individuals with Williams syndrome. *Am. J. Hum. Genet*., **61**,444–448.
12. Mari,A., Amati,F., Mingarelli,R., Gianotti,A., Sebastio,G., Colloridi,V., Novelli,G. and Dallapiccola,B. (1995) Analysis of the elastin gene in 60 patients with clinical diagnosis of Williams syndrome. *Hum. Genet*., **96**, 444–448.
13. Pérez Jurado,L.A., Peoples,R., Kaplan,P., Hamel,B.C.M. and 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.
14. Robinson,W.P., Waslynka,J., Bernasconi,F., Wang,M., Clark,S., Kotzot,D. and Schinzel,A. (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.
15. Morris,C.A., Thomas,I.T. and Greenberg,F. (1993) Williams syndrome: autosomal dominant inheritance. *Am. J. Med. Genet*., **47**, 478–481.
16. Sadler,L.S., Robertson,L.K., Verdaasdonk,K.R. and Gingell,R. (1993) The Williams syndrome: evidence of possible autosomal dominant inheritance. *Am. J. Med. Genet*., **47**, 468–470.
17. Dutly,F. and Schinzel,A. (1996) Unequal interchromosomal rearrangements may result in elastin gene deletions causing the Williams–Beuren syndrome. *Hum. Mol. Genet*., **5**, 1893–1898.
18. Urbán,Z., Helms,C., Fekete,G., Csiszar,K., Bonnet,D., Munnich,A., Donis-Keller,H. and Boyd,C.D. (1996) 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. *Am. J. Hum. Genet*., **59**, 958–962.
19. Yang,W. and Desiderio,S. (1997) BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. *Proc. Natl. Acad. Sci. USA*, **94**, 604–609.
20. Grueneberg,D.A., Henry,R.W., Brauer,A., Novina,C.D., Cheriyath,V., Roy,A.L. and Gilman,M. (1997) A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes Dev*., **11**, 2482–2493.
21. Roy,A.L., Du,H., Gregor,P.D., Novina,C.D., Martinez,E. and Roeder,R.G. (1997) Cloning of an Inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. *EMBO J*., **16**, 7091–7104.
22. Khan,A.K., Wilcox,A.S., Polymeropoulos,M.H., Hopkins,J.A., Stevens,T.J., Robinson,M., Orpana,A.K. and Sikela,J.M. (1996) Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet*., **2**, 180–185.
23. Hudson,T.J. *et al.* (1995) An STS-based map of the human genome. *Science*, **270**, 1945–1954.
24. 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.
25. Hillier,L., Clark,N. and Dubuque,T. The WashU–Merck EST project. Available online at http://genome.wustl.edu/est/esthmpg.html.
26. Lehrach,H. *et al.* (1990) In Davies,K.E. and Tilghman,S.M. (eds), *Genome Analysis*, Vol. 1: *Genetic and Physical Mapping*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY, pp 39–81.
27. Kozak,M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome*, **6**, 563–574.

28. Nakai,K. and Kanehisa,M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, **14**, 897–911.
29. Bairoch,A., Bucher,P. and Hofmann,K. (1996) The PROSITE database, its status in 1995. *Nucleic Acids Res.*, **24**, 189–196.
30. Henikoff,S. and Henikoff,J.G. (1991) Automated assembly of protein blocks for database searching. *Nucleic Acids Res.*, **19**, 6565–6572.
31. Eisen,J.A., Benito,M.-I. and Walbot,V. (1994) Sequence similarity of putative transposases links the maize *Mutator* autonomous element and a group of bacterial insertion sequences. *Nucleic Acids Res.*, **22**, 2634–2636.
32. Gilbert-Dussardier,B., Bonneau,D., Gigarel,N., Le Merrer,M., Bonnet,D., Philip,N., Serville,F., Verloes,A., Rossi,A., Aymè,S. and Munnich,A. (1995) A novel microsatellite DNA marker at locus D7S1870 detects hemizygosity in 75% of patients with Williams syndrome. *Am J. Hum. Genet.*, **56**, 542–544.
33. Osborne,L.R., Herbrick,J.-A., Greavette,T., Heng,H.H.Q., Tsui,L.-C, Scherer,S.W. (1997) PMS2-related genes (*hPMSR*) flank the rearrangement breakpoints associated with Williams syndrome and other diseases on human chromosome 7. *Genomics*, **45**, 402–406.
34. Wang,Y.-K., Pérez Jurado,L.A. and Francke,U. (1998) A mouse single copy gene, *Gtf2i*, the homolog of human GTF2I that is duplicated in the Williams–Beuren syndrome deletion region. *Genomics*, in press.
35. Yen,P.H., Li,X.M., Tsai,S.P., Johnson,C., Mohandas,T. and Shapiro,L.J. (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell*, **61**, 603–610.
36. Theodosiou,A.M., Morrison,K.E., Nesbit,A.M., Daniels,R.J., Campbell,L., Francis,M.J., Christodoulou,Z. and Davies,K.E. (1994) Complex repetitive arrangements of gene sequence in the candidate region of the spinal muscular atrophy gene in 5q13. *Am. J. Hum. Genet.*, **55**, 1209–1217.
37. Buiting,K., Gregor,V., Brownstein,B.H., Mohr,R.M., Voiculescu,I., Winterpacht,A., Zabel,B. and Horsthemke,B. (1992) A putative gene family in 15q11–q13 and 16p11.2: possible implications for Prader–Willi and Angelman syndromes. *Proc. Natl. Acad. Sci. USA*, **89**, 5457–5461.
38. Carrozo,R., Rossi,E., Christian,S.L., Kittikamron,K., Livieri,C., Corrias,A., Pucci,L., Fois,A., Simi,P., Boiso,L., Beccaria,L., Zuffardi,O. and Ledbetter,D.H. (1997) Inter- and intrachromosomal rearrangements are both involved in the origin of 15q11–q13 deletions in Prader–Willi syndrome. *Am. J. Hum. Genet.*, **61**, 228–231.
39. Li,X.-M., Yen,P.H. and Shapiro,L. (1992) Characterization of a low copy repetitive element S232 involved in the generation of frequent deletions of the distal short arm of the human X chromosome. *Nucleic Acids Res.*, **20**, 1117–1122.
40. Patel,P.I. and Lupski,J.R. (1994) Charcot–Marie–Tooth disease: a new paradigm for the mechanism of inherited disease. *Trends Genet.*, **10**, 128–133.
41. Chance,P.F., Abbas,N., Lensch,M.W., Pentao,L., Roa,B.B., Patel,P.I. and Lupski,J.R. (1994) Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.*, **3**, 223–228.
42. Pentao,L., Wise,C.A., Chinault,A.C., Patel,P.I. and Lupski,J.R. (1992) Charcot–Marie–Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nature Genet.*, **2**, 292–300.
43. Chen,K.-S., Manian,P., Koeuth,T., Potocki,L., Zhao,Q., Chinault,A.C., Lee,C.C. and Lupski,J.R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nature Genet.*, **17**, 154–163.
44. Nathans,J., Davenport,C.M., Maumenee,I.H., Lewis,R.A., Hejtmancik,J.F., Litt,M., Lovrien,E., Weleber,R., Bachynski,B., Zwas,F., Klingaman,R. and Fishman,G. (1989) Molecular genetics of human blue cone monochromacy. *Science*, **245**, 831–838.
45. Vollrath,D., Nathans,J. and Davis,R.W. (1988) Tandem array of human visual pigment genes at Xq28. *Science*, **240**, 1669–1672.
46. Craig,N.L. (1995) Unity in transposition action. *Science*, **270**, 253–254.
47. Reiter,L.T., Murakami,T., Koeuth,T., Pentao,L., Muzny,D.N., Gibbs,R.A. and Lupski,J.R. (1996) A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nature Genet.*, **12**, 288–297.
48. Petrij,F., Giles,R.H., Dauwerse,H.G., Saris,J.J., Hennekam,R.C.M., Masuno,M., Tommerup,N., van Ommen,G.-J.B., Goodman,R.H., Peters,D.J.M. and Breuning,M.H. (1995) Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature*, **376**, 348–351.
49. Wallerstein,R., Anderson,C.E., Hay,B., Gupta,P., Gibas,L., Ansari,K., Cowchock,F.S., Weinblatt,V., Reid,C., Levitas,A. and Jackson,L. (1997) Submicroscopic deletions at 16p13.3 in Rubinstein–Taybi syndrome: frequency and clinical manifestations in a North American population. *J. Med. Genet.*, **34**, 203–206.
50. Glaser,T., Jepeal,L., Edwards,J.G., Young,S.R., Favor,J. and Maas,R.L. (1994) PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nature Genet.*, **7**, 463–471.
51. Kirkpatrick,S.J., Kent,C.M., Laxova,R. and Sekhon,G.S. (1992) Waardenburg syndrome type I in a child with deletion (2)(q35q36.2). *Am. J Med. Genet.*, **44**, 699–700.
52. Pettigrew,A.L., Greenberg,F., Caskey,C.T. and Ledbetter,D.H. (1991) Greig syndrome associated with an interstitial deletion of 7p: confirmation of the localization of Greig syndrome to 7p13. *Hum. Genet.*, **87**, 452–456.
53. Kang,S., Graham,J.M.,Jr, Olney,A.H. and Biesecker,L.G. (1997) GLI3 frameshift mutations cause autosomal dominant Pallister–Hall syndrome. *Nature Genet.*, **15**, 266–268.
54. Francke,U., Yang-Feng,T.L., Brissenden,J.E. and Ullrich,A. (1986) Chromosomal mapping of genes involved in growth control. *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 855–866.
55. Worley,K.C., Wiese,B.A. and Randall,F. (1995) BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Res.*, **5**, 173–184.
56. Cohen,D., Chumakov,I. and Weissenbach,J. (1993) A first-generation physical map of the human genome. *Nature*, **366**, 698–701.
57. Kunz,J., Scherer,S.W., Klawitz,I., Soder,S., Du,Y.Z., Speich,N., Kalff-Suske,M., Heng,H.H.Q., Tsui,L.-C. and Grzeschik,K.H. (1994) Regional localization of 725 human chromosome 7-specific yeast artificial chromosome clones. *Genomics*, **22**, 439–448.