ARTICLE

A transcription factor involved in skeletal muscle gene expression is deleted in patients with Williams syndrome

Mayada Tassabehji[1], Martin Carette[1], Carrie Wilmot[2], Dian Donnai[1], Andrew P Read[1] and Kay Metcalfe[1]

[1]University Department of Medical Genetics and Regional Genetics Service, St Mary's Hospital, Manchester
[2]School of Biochemistry and Molecular Biology, University of Leeds, Leeds, UK

Williams-Beuren syndrome (WS) is a developmental disorder caused by a hemizygous microdeletion of approximately 1.4 MB at chromosomal location 7q11.23. The transcription map of the WS critical region is not yet complete. We have isolated and characterised a 3.4 kb gene, *GTF3*, which occupies about 140 kb of the deleted region. Northern blot analysis showed that the gene is expressed in skeletal muscle and heart, and RT-PCR analysis showed expression in a range of adult tissues with stronger expression in foetal tissues. Part of the conceptual *GTF3* protein sequence is almost identical to a recently reported slow muscle-fibre enhancer binding protein *MusTRD1*, and shows significant homology to the 90 amino-acid putative helix–loop–helix repeat (HLH) domains of the transcription factor *TFII–I* (encoded for by the gene *GTF2I*). These genes may be members of a new family of transcription factors containing this HLH-like repeated motif. Both *GTF3* and *GTF2I* map within the WS deleted region, with *GTF2I* being positioned distal to *GTF3*. *GTF3* is deleted in patients with classic WS, but not in patients we have studied with partial deletions of the WS critical region who have only supravalvular aortic stenosis. A feature of WS is abnormal muscle fatiguability, and we suggest that haploinsufficiency of the *GTF3* gene may be the cause of this.

Keywords: Williams-Beuren syndrome (WS); *GTF3*; transcription factor; *TFII–I/GTF2I*; HLH-like repeated motif; fatigue-resistant slow fibres; skeletal muscle; skeletal muscle gene expression

Introduction

Williams syndrome (WS: MIM 194050) is a rare condition (1/20 000) with striking physical and behavioural features.[1] People with WS have an unusual pattern of abilities (Williams Syndrome Cognitive Profile, WSCP), which include relatively good verbal abilities but deficient visuo-spatial abilities,[2–5] and characteristic personality traits.[6,7] Physically, there is growth retardation, a dysmorphic face, heart abnormalities (typically supravalvular aortic stenosis, SVAS), and sometimes severe infantile hypercalcaemia and hyperacusis.[8]

WS is caused by a chromosomal microdeletion at 7q11.23.[9] This chromosomal region is highly repetitive and the deletion arises from recombination between misaligned repeat sequences flanking the WS region.[10]

Table 1 Summary of genes mapped to the Williams Syndrome deleted region

Gene	Ref	Possible function of product
LIMK1	13,14	protein tyrosine kinase expressed in the developing brain
STX1A	15	syntaxin 1A, a component of the synaptic apparatus
RFC2	16	subunit of the DNA replication factor C complex
FZD3	17	homologous to the Drosophila tissue polarity gene *frizzled*
GTF21	18	encodes TFII-I (transcription factor) and BAP-135 (phosphorylated by Bruton's tyrosine kinase (*Btk*))
CPETR1,CPETR2	11,19	integral membrane proteins which may be involved in cell-cell signaling
FKBP6	20	FK-506 binding protein, possible role in hypercalcaemia and growth delay in WS
WSTF	21	transcription factor containing a PHD-type zinc finger and a bromodomain
WS-βTRP	22	homology to b-subunits of G proteins and other WD40 proteins. May play a role in signal transduction
BCL7B	22	may have cellular or developmental functions
WS-bHLH	22	basic helix–loop–helix leucine zipper protein; may act as a transcription factor

The deletion breakpoints cluster within these repeat regions, so that most WS patients have similar deletions of approximately 1.4 Mb.[11] Intensive transcription mapping has allowed identification of a number of genes localised to the WS critical region. Elastin (*ELN*) was the first deleted gene identified[12] and studies of patients with deletions or point mutations confined to this gene showed that hemizygosity for *ELN* causes SVAS but not the other typical features of WS.[12] Other genes since identified are shown in Table 1.[13–22]

All these genes appear to be deleted in most WS patients. No phenotype has been clearly assigned to haploinsufficiency for any of them and some are deleted in patients ascertained through SVAS who do not have WS.[13] This suggests that other genes within the deleted region may be responsible for the main features of WS apart from SVAS.

In our efforts to build a complete transcription map of the WS critical region, we have identified and characterised a potential transcription factor gene, *GTF3*. *MusTRD1*, a recently described slow muscle-fibre enhancer binding protein,[23] appears to be a partial transcript of this gene. There is significant homology to the putative helix–loop–helix repeat domains identified in the transcription factor *TFII–I*[24] (encoded for by the gene *GTF2–I*[18]). We have shown that *GFT3* is deleted in patients with classic WS but not in SVAS patients with partial deletions within the critical region who lack other features of WS. We suggest that haploinsufficiency of *GTF3* causes the muscle fatigue and lack of stamina seen in patients with WS.

Materials and Methods

GTF3 *cDNA Isolation*

A human 18-week foetal brain cDNA library (Gibco, Paisley, UK) was screened using the BAC clone 175E7 from within the WS deleted region. The DNA insert was isolated from the

vector by Not1 digestion and separated by PFGE in 1% low-melting TAE buffer with linear ramping of 5–15 s at a field intensity of 6 V/cm on a Bio-Rad Chef-DRIII PFGE apparatus. The inserts were sized against a low-size range marker (Bio-Rad, Herts, UK).

The DNA insert was then cut out of the gel and purified using a Geneclean resin. Hybridisation conditions were used according to the manufacturer's instructions and the repeats were competed out using placental DNA at a 100-fold excess. Stringent washing was used to avoid false positive signals. Isolated cDNA plasmid clones were miniprepped (Promega Wizard kit, Southampton, UK) and cycle sequenced using the Thermosequenase II sequencing kit (Amersham, Buckinghamshire, UK) and visualised on an ABI 373 fluorescent sequencer.

5' and 3' RACE This was carried out as described by Frohman *et al*[25] using the following primers:

5'RACE: GAACGCGGAGTTCCAGCGGTCG

3'RACE: GCCGGTGCTGGTCCCTTATAAAC

Gene Expression Studies

A northern blot of poly(A)$^+$ RNA from various human tissues (Clontech, Hampshire, UK) was probed with the *GTF3* cDNA clone and a β-actin control, and hybridised according to the manufacturer's instructions.

RT-PCR Analysis Total RNA (1 μg) from each RNA sample was reverse transcribed using a first-strand cDNA synthesis kit (Promega, Southampton, UK). Primers made from unique *GTF3* sequence were used to PCR amplify the cDNA samples. PCR conditions used were: 50 ng of cDNA, 10 pmol of each primer and 0.5 U Taq polymerase (BCL, Cheshire, UK) used with the manufacturer's buffer, in a 20 μl reaction. PCR conditions were: 2 min denaturation at 94°C, followed by 30 cycles of 94°C 1 min, 60°C 1 min, 72°C 1 min, with a 5 min extension at 72°C. The integrity of the RNA was assessed by PCR analysis of the human β-actin gene (330 bp product) using the same cycling conditions. Primer sequences:

GTF3 splice isoforms F:
GCCCGAGCTGCTCACTGAGGGAGT

GFT3 splice isoforms R:
ATGATGACCATGCGGACATGCTC

GTF3F: AACTGCCTCCTCACTTGGCTTCT

GTF3R: ATGATGACCATGCGGACATGCTC

β-actin F: 5' GGC CGT CTT CCC CTC CAT C

β-actin R: 5' TAG CAA CGT ACA TGG CTG GGG

Fluorescence in situ Hybridisation

Metaphase spreads of chromosomes from patients were prepared from EBV transformed lymphoblastoid cell lines using standard techniques. FISH was carried out as described previously[13] using BAC probes containing the *GTF3* gene. A cosmid mapping to chromosome 7 p was used to identify the chromosome 7 homologues and as an internal control.

Somatic Cell Hybrids

Hybrid cell lines were isolated after fusion of lymphoblastoid cells from the patients with mouse BW5147 cells as previously described.[14] The presence of a single chromosome 7 in all cell lines was verified by FISH analysis of interphase nuclei using a chromosome 7 specific centromeric probe (Oncor, Middlesex, UK) and on metaphase chromosomes using a human chromosome 7 specific paint (Cambio, Cambridge, UK). Hybrids containing normal and deleted chromosomes were distinguished by FISH (WSCR elastin probe, Oncor, Middlesex, UK), and microsatellite typing using D7S1870 with D7S653 as a control.

The preparation and analysis of hybrid cell lines from the WS and SVAS patients has been previously described.[13]

Deletion Mapping of Hybrids by PCR

Primers were designed for PCR analysis of somatic cell hybrids from the 5' and 3' *GTF3* cDNA sequence. 50 ng of DNA, 10 pmol of each primer and 0.5 U Taq polymerase (BCL, Cheshire, UK) were used with the manufacturer's buffer, in a 20 µl reaction. PCR conditions were: 2 min denaturation at 94°C, followed by 30 cycles of 94°C 1 min, 60°C 1 min, 72°C 1 min, with a 5 min extension at 72°C.

5' *GTF3* F: GCGACCATGGCCTTGCTGGGTAA

5' *GTF3* R: CATGGAGTCTAAGGCAGACACGA

3' *GTF3* F: CTTGGTGGCAGGAATATAGTGTA

3' *GTF3* R: AATGGCCAATGTACATGGTGGAC

Results

Isolation and Characterisation of the GTF3 Gene

A recombinant pCMV.SPORT plasmid containing a novel human gene was isolated by screening an 18-week human foetal brain cDNA library with BAC 175E7, which lies within the telomeric region of the WS microdeletion.[11] Three overlapping clones were obtained, consisting of partial transcripts of the same gene. The 5' and 3' termini were obtained by RACE extension and the nucleotide sequence of the gene was determined from both strands (Figure 1).

Searches in the EMBL database identified regions of homology in three BAC/PAC clones, RG228D17, DJ665P05 and DJ1186P10. Sequence comparisons between the cDNA clones and these genomic clones allowed some of the exon–intron boundaries to be determined. The rest were identified by sequencing directly from the BAC/PAC clones using cDNA specific primers. In this way we were able to define the genomic organisation of the gene, which we named *GTF3* (Figure 2).

The *GTF3* cDNA is 3421 bp in size with a coding region of 2835 bp (starting from the first methionine residue). It is composed of 27 exons ranging in size from 26 bp to 311 bp. The sizes of the introns were estimated from the sequence of the genomic BAC/PAC clones. All the introns have the consensus splice donor and acceptor sequences.

Sequence analysis of all the cDNA clones isolated demonstrated the existence of two isoforms differing by an in-frame insertion of 15 aminoacids, produced by the use of two alternative acceptor splice sites in intron 18. This was confirmed by RT-PCR analysis (Figure 5B).

The 5' UTR is interrupted by a large intron of approximately 53880 bp which contains the microsatellite repeat, D7S2472, commonly used for deletion mapping in WS patients. The presumed ATG initiation codon is located in exon 2 of the gene, although computer analysis identified an in-frame CTG, 63 bp further upstream in the same exon, as a possible alternative start codon. This CTG lies in an optimal primary Kozak consensus sequence (A or G in position –3 and G in position +4) and has a number of downstream stem-loop (hairpin) structures due to the GC-rich sequence, as predicted by a secondary structure programme (RNA Draw). Kozak[26] has shown the variable ability of mammalian ribosomes to initiate translation at non-AUG codons in cases where stem-loop structures are formed downstream from the alternative initiator codon. Following protein analyses were carried out using the conventional methionine start site.

We estimate that the *GTF3* gene spans a genomic distance of approximately 140 kb. PCR mapping using the clones in our contig of the WS critical region[11] shows that the gene lies in a 5' to 3' orientation from the centromere to the telomere and lies proximal to the

Figure 1 Nucleotide and amino-acid sequence of GTF3 cDNA. The two potential translation start sites (CTG and (ATG) are in boldface and underlined. Kozak consensus sequences surrounding the potential CTG alternative start codon are shaded. The five TFII-I repeat motifs are underlined. The alternatively spliced out sequence is in bold. The run of serine residues is shaded. cAMP- and cGMP-dependent protein kinase phosphorylation sites (RKRVS) are in bold and italics. (Genbank accession number: AF151354)

```
5'UTR                              Exon 1 (57bp)
ttaattaaggaggcagcggcaggaggctgagtcCTGGCCGCGGGCCGGGGCCGGGGCGCCGCTGGCAGGAGCGCTTGGGGATCCTCCAA
Ggtaggagaaacttttgcggg Intron 1 (approx 53880bp)

tgcctcctccctccccacagGCGACCATGGCC Exon 2 (129bp)  TTAGACTCCATGgtgagtgtccccacccaccca
Intron 2 (>4560bp)
acagcctgcccttcccacagTGCTCAGCGCTG Exon 3 (142bp)  TCAGGTTCTGCCgtgagtaccccagggc
Intron 3(2466bp)
gcgtgctctgtgtcccacagGAGGGCCCCCGT Exon 4 (156bp)  ATGTTCTTTATAgtaagatccttcctcattcc
Intron 4(2542bp)
tcctggactctcccctacagGCGAGGCCCTGG Exon 5 (184bp)  CAAGCTCAAGAGgtgagtgaggtagccggccc
Intron 5(1084 bp)
gtgtcctctcctttactcagGCCACTTGAGGA Exon 6 (311bp)  CCGACTGTTGTGgtaacattgctgctgggatc
Intron 6(1488bp)
tccttctctcccttgtccagGACAGAAGCCCA Exon 7 (90bp)   ATGACAAGTCAGgtaggacagcgcccacgaag
Intron 7(2773bp)
tgcctcctgtgttttgatagAGAAGTGGGACG Exon 8 (84bp)   ACAGCAGATATGgtgagtgggcggcgcggccc
Intron 8(>5920bp)
ctgtttggttgtctttccagCGGAAGCCCTGG Exon 9 (184bp)  CATCATTAAGAGgtgcgggtggggctgggcgc
Intron 9(5195bp)
tgtctctcttctctccccagGATGTTTGATGA Exon 10 (26bp)  GAATTTTCACAGgtatgtggggaccatctagt
Intron 10(1028bp)
ctccctgactctccccacagGGAACAAGTTTA Exon 11 (109bp)TGGGAATGCTCGgtgaggccccgcccctggcc
Intron 11(1860bp)
tttcctgcttcccttcacagGTCAGACAAGGG Exon 12 (38bp)  ACTGTGGGCCAGgtgagaaggaacagggcccg
Intron 12(506bp)
tcttcctttcctccttgcagGAACCTCCGGGG Exon 13 (81bp)  TCACCGTCCCAGgtaagggacgggcatctgac
Intron 13(1118bp)
attcggagttttgtcttcagACCCCTCGCCAA Exon 14 (90bp)  AGATGCTGACAGgtaagaaatggacctgggct
Intron 14(3011bp)
catcctctgttctctttagACAAAGGTCTGA Exon 15 (48bp)  GGCCCGTGGAGGgtgaggccctgtctacccct
Intron 15(941bp)
gactgccttttgccttccagACAGCCACGGTG Exon 16 (66bp)  ACACACGATACGgtgagcaagaagtgggacag
Intron 16(1293bp)
acagcagccagtctctgcagCCAAGGCCATTG Exon 17 (184bp)TGTCATCAAGAGgtaaggcccaaccaggtcca
Intron 17(7886bp)
gcctcgcttgtgttttccagGCCCGAGCTGCT Exon 18 (50bp)  TGGATAGTCAAGgtacccagcgcggggtcgg
Intron 18(169bp)
gatgccatcttgggtcccagGAACTGCCTCCT Exon 19 (102bp)AGGGCTTTCAAGgtaaggttgagctcacgggg
gatgccatcttgggtcccagAGAGGGATTCCG Exon 19A (57bp)AGGGCTTTCAAGgtaaggttgagctcacgggg
Intron 19(1153bp)
aatcgtttgcgctttgcagAAAATTATGACG Exon 20 (84bp)  ATAAGAAATACGgtaagcagtgcagaaccccc
Intron 20(1137bp)
accctgcctctcgcccacagGGGAAGCCTTGG Exon 21 (184bp)CACAGTCACCAGgtactcagtgggaagggtga
Intron 21 (581bp)
tgcctctgtttctcttctagGCCTTTCCAAGG Exon 22 (29bp)  TCCCAAAGCCTGgtaagaggcactggctgtgg
Intron 22 (>18321bp)
cccttgtcttcctcttgtagATGAAGATGACG Exon 23 (78bp)  CGAGAAATACGgtcagtgcctgtggtcaggg
Intron 23(896bp)
tgacttcctgtgccctctagGTGAGGCCCTGG Exon 24 (193bp)TAACCAGCTCCAgtgagtgcccggcctctgga
Intron 24(3994bp)
ctcatttctttcttttcagACCCTTTGCAGA Exon 25 (37bp)  GCCAAGTGCCAGgtgagtcagaggtgaagaag
Intron 25(5991bp)
gccttttcccctccttccagCCAAAGACAGCA Exon 26 (137bp)ATCTCACTCGTGgtaaagttgcaccgatttgg
Intron 26(1184bp)                                           3'UTR
ccctttgtccctctctctagCAATGGCCAATG Exon 27 (69 bp)CCTCTTAATTACTAGggaaatgtaatttatgta
```

Figure 2 *Genomic organisation of GTF3. Exon and intron sizes are given in brackets. Coding sequence is in capital letters, intron sequence is in lower case letters, and concensus acceptor (ag) and donor (gt) splice sites are in bold.*

GTF2I gene. Within the 2 Mb contig there is only a single copy of GTF3. Screening the cDNA sequence for the presence of repeat sequences (Alu, LINE, SINE etc), showed that it was not rich in repeats. Only 226 bp were masked, representing 6.5% of the sequence.

Sequence Analysis and Homology with MusTRDI and TFII–I

BLAST analysis[27] of GTF3 (from the first ATG codon) with a non-redundant protein database showed that aminoacids 1 to 434 (out of 959 aminoacid residues) are

identical to residues 1 to 434 (out of 458 amino-acid residues) of a recently submitted gene *MusTRDI* (human muscle *TFII–I* repeat domain containing protein 1),[23] except at two positions. There were sequence discrepancies G132S and V411G. Only the C-terminal 24 amino acid sequence of *MusTRD1* does not align with *GTF3*.

The BLAST analysis also revealed significant homology to the family of transcription factors *TFII–I* (also known as *SPIN*). The major regions of homology involve a repeated amino-acid motif previously described in *TFII–I*[24] (Figure 3). Sequence alignment between *TFII–I* and *GTF3* shows that 20/79 amino-acids possess a match in all the repeats in both genes, with no more than one residue variation; and 31/79 residues possess the same amino-acids in at least three repeats in one gene and at least two repeats in the other. Further homology is evident on inspection of Figure 3B. Significant homology (34/66 identity, 17 conservative changes) was also detected at the N-terminus between TFII–I residues 21–86 and *GTF3* residues 29–94 (Figure 3A). This sequence was not repeated

elsewhere in the coding sequence of either gene, and did not show homology to any motifs in the PROSITE database.[28] The amino acid sequence that is variable between the two *GTF3* isoforms did not show significant homology to *TFII–I* and lay outside the conserved repeated motifs.

Predicted Properties of GTF3 Protein

Both *GTF3* isoforms are predicted to be soluble proteins with pI 6.45[29] and molecular weights 106,057 Da (long isoform, 959 aminoacids) and 104,734 Da (short isoform, 945 aminoacids). PROSITE[28] reported phosphorylation sites. The presence of 13 [ST]- $\times$ -[RK] sequences suggest possible phosphorylation by protein kinase C, and the sequence RKRVS (902–906, long isoform; 888–892, short isoform) suggests possible cAMP and cGMP dependent protein kinase phosphorylation (PROSITE[28]). They appear to be localised within the nucleus, based on Euclid distances between 11 normalised methods for determining localisation to different compartments of the cell (74% probability[30]). A run of 12 serine residues
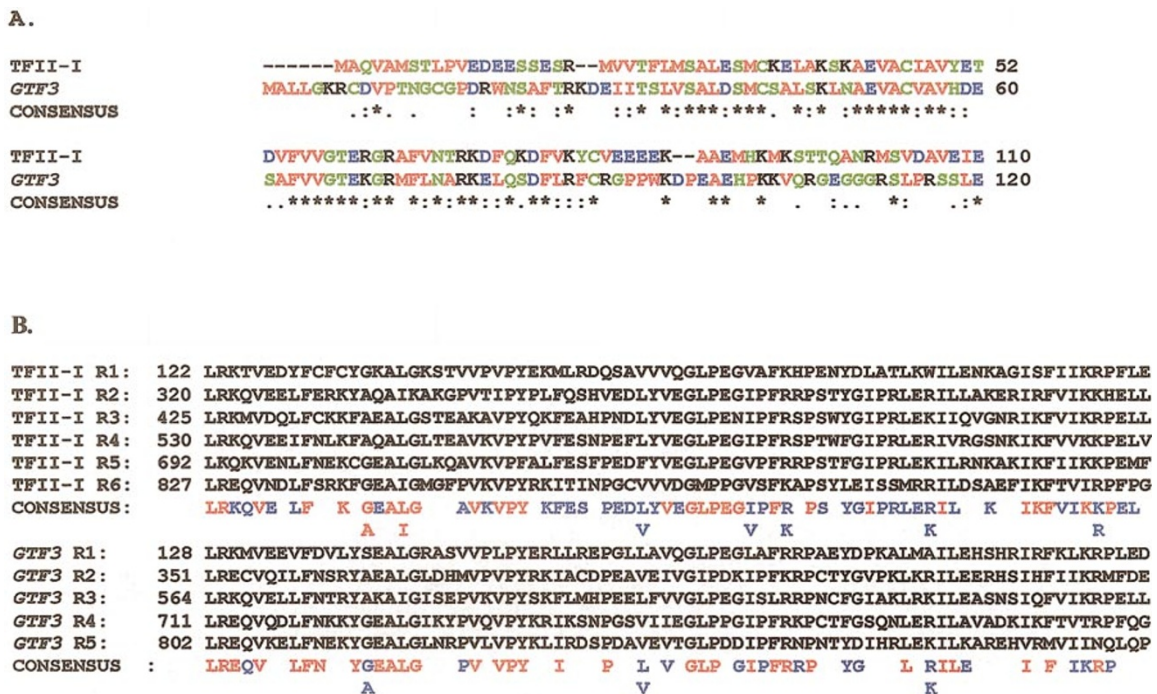
Figure 3A *Sequence alignment between at the start of GTF3 and TFII–I. The consensus line: "*"identical or conserved residues in all sequences in the alignment; ":" = indicates conserved subsitutions; "." = indicates semi-conserved substitutions. Amino-acid codes: Green = hydroxyl, amine, basic-Q. Blue = acidic, Red = small and hydrophobic including aromatic- Y, black = basic. Significant homology between TFII–I residues 21–86 and GTF3 residues 29–94 with 34/66 identity and 17 conservative changes.* **3B** *Alignment of the repeat domains in TFII–I and GTF3. Bases in red: conserved; Bases in blue: semi-conserved. 20/79 amino acids possess a match in all the repeats in both genes (no more than one residue variation); 31/79 residues possess the same amino acids in at least three repeats in one gene and at least two repeats in the other.*

occurs at the C-terminus of the protein. Multi-serine runs appear to be a feature of many soluble proteins. Many of these are nuclear (eg *Humhoxy1*, a zinc finger DNA-binding protein AC:D45132), and some bind DNA and RNA. There does not appear to be any other relationship between *GTF3* protein and the proteins in the database with multi-serine runs. PSORT II, PRO-SITE, BLOCKS,[31] and PRINTS[32] did not detect any strong sequence homology to known DNA or RNA binding motifs.

Protein secondary structure predictions (GORB-TURN[33,34]) suggest a highly helical structure in the N-terminal half of the sequence, whilst the C-terminal half appears to be a mix of helix and beta-sheet.

Analysis of Expression Pattern

Northern blot analysis using an almost complete cDNA probe (exon 2 to the 3'UTR) showed expression mainly in skeletal muscle and the heart, but the presence of stronger β-actin control bands in these tissues suggests that the blot used is unequally loaded (Figure 4A). Under the stringent hybridisation conditions employed to avoid cross-hybridisation with *GTF2–I*, we were able to identify one main mRNA band of approximately 3.5 kb, and a weak band of approximately 5 kb. Our cDNA construct for *GTF3* represents the 3.5 kb transcript, and the larger transcript may be a larger isoform or another gene with regions of homology.

RT-PCR analysis of a number of human adult and foetal tissues was carried out to detect low level expression of *GTF3*. Primers used for this analysis were designed in the unique regions of the gene to avoid PCR amplification of *GTF2–I*. The levels of *GTF3* expression were estimated by comparison with an
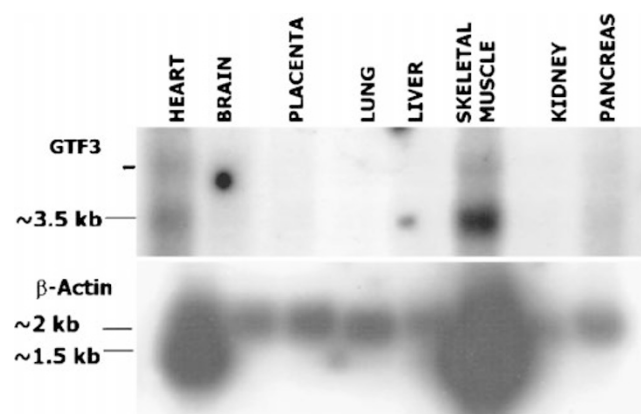

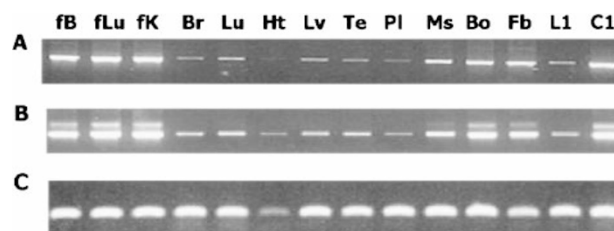**Figure 4B** *RT-PCR analysis of GTF3 expression. Panel A: GTF3 exon 19 long isoform; B: exon 19 alternative isoforms; C: β-actin control. fB = foetal brain, fLu = foetal lung, fK = foetal kidney; Br = brain, Lu = lung, Ht = heart, Lv = liver, Te = testis, Pl = placenta, Ms = skeletal muscle, Bo = bone, Fb = fibroblast cell line, L1 = lymphoblastoid celline, C1 = osteoblast-like bone cell line.*

internal β-actin control. *GTF3* is highly expressed in all the foetal tissues tested, with lower levels of expression in the adult tissues (Figure 4B). Of the adult tissue tested, skeletal muscle, bone, a fibroblast cell line and an osteoblast-like bone cell line showed significantly higher *GTF3* expression. Primers designed to amplify the individual isoforms of *GTF3* showed that they are both expressed in all the tissues tested.

FISH Analysis of GTF3 in WS Patients

FISH analyses using a clone containing the *GTF3* gene were carried out on five patients with classic WS (MB, DB, MC, GU, KE) and on three patients with SVAS only, who have partial deletions of the WS critical region (TM, CS and HG, previously described[11,13]). One copy of *GTF3* was deleted in all of the WS patients, but not in any of the three SVAS patients, showing that their partial deletions did not extend as far as *GTF3* (data not shown). There did not appear to be multiple copies of *GTF3* on chromosome 7, and no signals were detected on any other chromosome.

GTF3 Deletion Analysis in WS Patients by Somatic Hybrid PCR

Because partially deleted genes can give detectable FISH signals, somatic cell hybrid lines were made from cells of the WS and SVAS patients to allow their deletion breakpoints to be more finely mapped. The normal and deleted copies of chromosome 7 were segregated, and stable cell lines established. Hybrids containing the deleted chromosome 7 from each patient were tested for the presence of *GTF3* by PCR using primers designed from the 5' and 3' regions of the gene. A non-deleted marker was included as a positive control in the PCR reaction. Mouse DNA was also
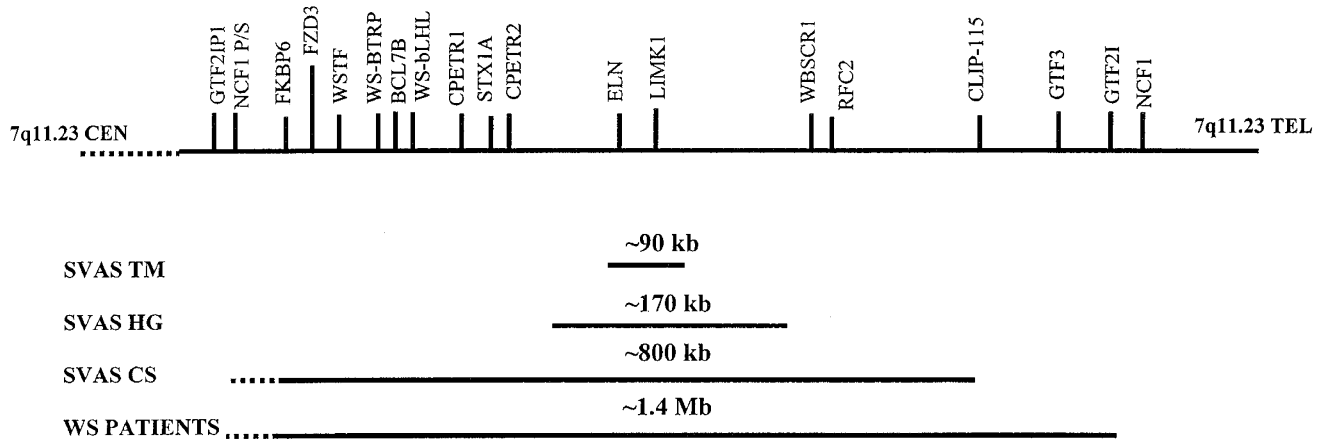

Figure 4A *Northern blot analysis using GTF3 cDNA as a probe. A transcript of about 3.5 kb is seen in adult skeletal muscle and heart tissue.*

Figure 5A *Map of Williams syndrome critical region showing the approximate size and the extent of the deletions in the SVAS patients (TM, HG, CS) and WS patients (MB, DB, MC, GU, KE).*
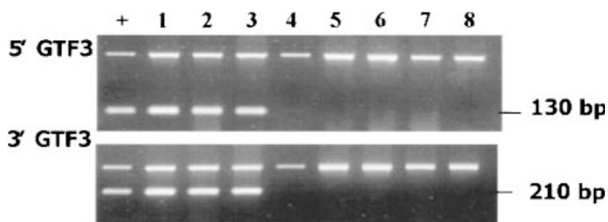

Figure 5B *GTF3 PCR analysis of hybrids. The 5' and 3' ends of GTF3 were amplified in the deleted 7 somatic cell hybrid. An internal control was included in each reaction (top band). +: normal control. Lanes 1–3: SVAS patients (TM, HG, CS) were not deleted for GTF3; Lanes 4–8: Williams syndrome patients (MB, DB, MC, GU, KE) were all deleted for GTF3.*

included as a control to show that the PCR amplification products were human specific. The results (Figure 5) show that the entire *GTF3* gene is deleted in each WS patient but is not deleted in the three SVAS patients.

Discussion

We have isolated a single copy gene, named *GTF3*, mapping within the WS critical region, which is deleted in patients with WS but not in three patients who have smaller deletions and have no features of WS apart from supravalvular aortic stenosis. Structural analysis of *GTF3* suggests that it may encode a transcription factor. The putative *GTF3* protein is predicted to be soluble and localised to the nucleus, and shows homology to the transcription factor, *TFII–I*. Since submitting this manuscript, the gene WBSCR11 mapping to the

WS critical region has been published[35] and sequence comparisons with *GTF3* shows that there are some sequence and genomic organisation discrepancies, but the two genes have significant identity. Northern blot analysis of WBSCR11 shows transcripts of 1.8, 2 and 5 kb, in contrast to our main *GTF3* transcript of 3.4 kb.

The homology to *TFII–I* involves two parts of the sequence. At the N-terminus there is strong homology over a 66 aminoacid run that does not correspond to any recognised motif in the databases. A second region of homology is to a 90-aminoacid potential helix–loop–helix repeat motif, present in six copies in *TFII–I* and five in *GTF3*. *TFII–I* is reported to bind both proteins and DNA. DNA targets include Inr, Inr-like (SIE, SRE) and E-box elements, whilst the proteins include serum-response factor (a MADS box family transcription factor), Phox (a homeodomain protein) and USF1 (a bHLH protein).[36,37] It is thought that the six potential HLH, repeat domains of *TFII–I* are important for its role as a co-ordinator of diverse cell signaling responses and the basal transcription machinery.[24] The conservation of such domains in *GTF3* suggests that these genes belong to a new family of transcription factors which may play similar roles *in vivo*.

A recently reported gene, *MusTRD1*[23] is predicted to encode a protein identical, except at two positions, to aminoacids 24–455 of *GTF3*. The predicted *MusTRD1* protein has a 21-aminoacid C-terminus that differs from the 510 aminoacids of the *GTF3* C-terminus. Inspection of the *MusTRD1* cDNA sequence shows the *GTF3* downstream sequence is present, with a few

minor discrepancies, but is interpreted as a 3' untranslated region. PCR did not amplify any sequence having the stop codon as reported for *MusTRD1*, and our FISH analysis confirmed that *GTF3* is present in only a single copy in the genome. We conclude that *MusTRD1* is *GTF3*, and that this gene encodes a protein with isoforms of 965 or 980 aminoacids.

MusTRD1 was identified by O'Mahoney *et al*[23] as a protein that binds to an enhancer involved in the regulation of skeletal muscle gene expression. They identified the sequence elements (E-box, MEF2, USE B1 and USE C1) of the human troponin I slow upstream enhancer (USE) that bind muscle nuclear proteins. They showed that the TnIs USE B1 sequence element appears to be essential for high-level expression in slow-twitch muscles, and cloned *MusTRD1* as a gene encoding a 51 kDa USE B1 binding protein that is expressed predominantly in adult skeletal muscle. No chromosomal location was assigned. Our data on the expression profile of *GTF3* show that it too is highly expressed in adult skeletal muscle, fibroblasts and bone, but foetal expression was much more pronounced and was seen at similar levels in all the tissues tested.

Skeletal muscle commitment, differentiation and maturation are largely controlled by genes encoding channels, receptors, metabolic enzymes and muscle-specific contractile proteins,[38,39] but not much is known about the regulatory elements that restrict transcription of genes encoding contractile proteins specifically to either slow or fast-twitch skeletal muscles. Troponin I (TnI), a member of the troponin complex that regulates muscle contraction in response to calcium flux, comprises three isoforms, fast (TnIf), slow (TnIs) and cardiac (TnIc). These isoforms are differentially regulated during skeletal muscle development. Most skeletal muscles contain a mixture of fibres with different characteristics, and their slow-twitch units are well known to possess properties suitable for postural tasks: they are highly fatigue-resistant, well equipped for oxidative metabolism, and their slowness makes them energetically cheap in (semi-)isometric contractions.

WS patients are reported to lack stamina[38] (KM, clinical observations 1999). Routine manual tasks considered suitable for people with learning difficulties, such as stacking shelves, packing or assembly-line work, are unsuitable for many patients with WS, partly because of their visuomotor difficulties, but also because of a lack of stamina and a tendency to tire easily.[40] Many young and old WS patients appear to suffer from muscle problems, ranging from slight aches

like cramp, to noticeable pain from walking long distances (UK WS Foundation committee, personal communication 1999). Voit *et al*[41] have described myopathy in infancy in WS patients with clinical manifestations including walking delay, joint contractures and increased exhaustion on exertion. If the skeletal muscles of WS patients contained fewer slow-twitch fibres, they might be expected to tire more easily. There is evidence to suggest that a decrease in the proportion of fatigue-resistant slow fibres in skeletal muscle affects stamina.[42] In patients with chronic obstructive pulmonary disease the scarce oxygen availability and/or the inactivity and physical deconditioning are thought to cause fibre transformation, with an increase in the proportion of the fast fatiguable fibres that are responsible for lactate production and fatigue at low work rates. Investigations into the fibre composition in skeletal muscle biopsies from WS patients, as well as stamina tests, would be required to determine if this is also the case in Williams syndrome.

Both *TFII–I* and *GTF3* appear to share similar properties and may play similar roles in regulating transcription of a number of genes. They map next to each other within the WS critical region, just outside the tandemly repeated region. We have previously shown[11] that *GTF2I* (*TFII–I*) is also deleted in patients with Williams syndrome. Transcription factors are often dosage-sensitive, and haploinsufficiency for transcription factors is a cause of a growing list of developmental defects, such as Waardenburg syndrome Types I and II and aniridia. Although *GTF3* appears to be involved in controlling expression of a muscle gene, and shows significant expression in adult skeletal muscle, it is more highly and more widely expressed in fetal than in adult tissues. *GTF3* occupies 10% of the entire WS critical region, and studies of rare patients with partial deletions suggests that this part of the critical region is important for determining the WS phenotype.[13,43] *GTF3* and/or *TFII–I* must be strong candidates for the critical WS genes.

Acknowledgements

References

1 Williams JCP, Barratt-Boyes BG, Lowe JB: Supravalvular aortic stenosis. *Circulation* 1961; **4**: 1311–1318.

2 Udwin O, Yule W: A cognitive and behavioural phenotype in Williams Syndrome. *J Clin Exp Neuropsychol* 1991; **13**: 232–244.

3 Frangiskakis JM, Ewart AK, Morris CA *et al*: LIM-kinase 1 hemizygosity implicated in impaired visuospatial constructive cognition. *Cell* 1996; **86**: 59–69.

4 Bellugi U, Wang P, Jernigan TL: Williams syndrome: An unusual neuropsychological profile. In: S Broman, J Graffman (eds). *Atypical Cognitive Defects in Developmental Disorders: Implications for Brain Function*. Lawrence Erlbaum Associates: Hillsdale, NJ, 1994, pp 23–56.

5 Jarrold C, Baddeley AD, Hewes AK: Verbal and non-verbal abilities in the Williams syndrome phenotype: evidence for diverging developmental trajectories. *J Child Psychol Psychiatr* 1998; **39**: 511–523.

6 Karmiloff-Smith A, Klima E, Bellugi U, Grant J, Baron-Cohen S: Is there a social module? Language, face processing and theory of mind in individuals with Williams syndrome. *J Cognit Neurosci* 1995; **7**: 196–208.

7 Einfeld SL, Tonge BJ, Florio T: Behavioral and emotional disturbance in individuals with Williams syndrome. *Am J Mental Retard* 1997; **7**: 45–53.

8 Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL: The natural history of Williams syndrome: physical characteristics. *J Pediatr* 1988; **113**: 318–326.

9 Nickerson E, Greenberg F, Keating M, McCaskill C, Shaffer LG: Deletions in the elastin gene at 7q11.23 occur in ~ 90% of patients with Williams syndrome. *Am J Hum Genet* 1995; **56**: 1156–1161.

10 Baumer A, Dutly F, Balmer D *et al*: High level of unequal meiotic crossovers at the origin of the 22q11.2 and 7q11.23 deletions. *Hum Mol Genet* 1998; **7**: 887–894.

11 Hockenhull EL, Carette MJ, Metcalfe K, Donnai D, Read AP, Tassabehji M: A complete physical contig and partial transcript map of the Williams Syndrome critical region. *Genomics* 1999; **58**(2): 138–45.

12 Ewart AK, Morris CA, Atkinson D *et al*: Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat Genet* 1993; **5**: 11–16.

13 Tassabehji M, Metcalfe K, Karmiloff-Smith A *et al*: Williams syndrome: Using chromosomal microdeletions as a tool to dissect cognitive and physical phenotypes. *Am J Hum Genet* 1999; **64**(1): 118–125.

14 Tassabehji M, Metcalfe K, Fergusson WD *et al*: LIM-kinase deleted in Williams syndrome. *Nat Genet* 1996; **13**: 272–273.

15 Osborne LR, Soder S, Shi XM *et al*: Hemizygous deletion of the syntaxin 1A gene in individuals with Williams syndrome. *Am J Hum Genet* 1997; **61**: 449–452.

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

17 Wang YK, Samos CH, Peoples R, Perez-Jurado LA, Nusse R, Francke U: 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* 1997; **6**: 465–472.

18 Perez Jurado LA, Wang YK, Peoples R, Coloma A, Cruces J, Francke U: 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. *Hum Mol Genet* 1998; **7**(3): 325–334.

19 Katahira J, Sugiyama H, Inoue N, Horiguchi Y, Matsuda M, Sugimoto N: Clostridium perfringens enterotoxin utilizes two structurally related membrane proteins as functional receptors in vivo. *J Biol Chem* 1997; **272**: 26652–26658.

20 Meng X, Lu X, Morris CA, Keating MT: A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* 1998; **52**(2): 130–137.

21 Lu X, Meng X, Morris CA, Keating MT: A novel human gene, WSTF, is deleted in Williams Syndrome. *Genomics* 1998; **54**: 241–249.

22 Meng X, Lu X, Li Z *et al*: Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel gene. *Hum Genet* 1998; **103**: 590–599.

23 O'Mahoney JV, Guven KL, Lin J *et al*: Identification of a novel slow-muscle-fiber enhancer binding protein, *MusTRD1*. *Mol Cell Biol* 1998; **18**(11): 6641–6652.

24 Roy AL, Du H, Gregor PD, Novina CD, Martinez E, Roeder RG: Cloning of an inr- and E-box-binding protein, *TFII–I*, that interacts physically and functionally with USF1. *EMBO J* 1997; **16**(23): 7091–7104.

25 Frohman MA, Dush MK, Martin GR: Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 1988; **85**(23): 8998–9002.

26 Kozak M: Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA* 1990; **87**: 8301–8305.

27 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.

28 Bairoch A, Bucher P, Hofmann K: The PROSITE database, its status in 1997. *Nucleic Acids Res* 1997; **25**(1): 217–221.

29 Bjellqvist B, Hughes GJ, Pasquali Ch *et al*: The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* 1993; **14**: 1023–1031.

30 Horton P, Nakai K: Better prediction of protein cellular localization sites. Intelligent systems for molecular biology, Ismb. 1997; **5**: 147–152.

31 Henikoff S, Henikoff JG: Protein family classification based on searching a database of blocks. *Genomics* 1994; **19**: 97–107.

32 Attwood TK, Flower DR, Lewis AP *et al*: PRINTS prepares for the new millennium. *Nucleic Acids Res* 1999; **27**(1): 220–225.

33 Wilmot CM, Thornton JM: Analysis and prediction of beta-turn types in proteins. *J Mol Biol* 1988; **203**: 221–232.

34 Wilmot CM, Thornton JM: Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng* 1990; **3**: 479–493.

35 Osborne LR, Campbell T, Daradich A, Scherer SW, Tsui LC: Identification of a putative transcription factor gene (WBSCR11) that is commonly deleted in Williams-Beuren syndrome. *Genomics* 1999; **57**(2): 279–284.

36 Clark MP, Chow CW, Rinaldo JE, Chalkley R: Multiple domains for initiator binding proteins *TFII–I* and YY-1 are present in the initiator and upstream regions of the rat XDH/XO TATA-less promoter. *Nucleic Acids Res* 1998; **26**(11): 2813–2820.

37 Kim DW, Cheriyath V, Roy AL, Cochran BH: *TFII–I* enhances activation of the c-fos promoter through interactions with upstream elements. *Mol Cell Biol* 1998; **18**(6): 3310–3320.

38 Buonanno A, Rosenthal N: Molecular control of muscle diversity and plasticity. *Dev Genet* 1996; **19**: 95–107.

39 Schiaaffino S, Reggiani C: Molecular diversity of myofibrillar proteins: gene regulation and functional significance. *Physiol Rev* 1996; **76**: 371–423.

40 Davies M, Howlin P, Udwin O: Independence and adaptive behavior in adults with Williams Syndrome. *Am J Med Genet* 1997; **70**: 188–195.

41 Voit T, Kramer H, Thomas C, Wechsler W, Reichmann H, Lenard HG: Myopathy in Williams-Beuren syndrome. *Eur J Pediatr* 1991; **150**(7): 521–526.

42 Satta A, Migliori GB, Spanevello A *et al*: Fibre types in skeletal muscles of chronic obstructive pulmonary disease patients related to respiratory function and exercise tolerance. *Eur Respir J* 1997; **10**(12): 2853–2860.

43 Botta A, Novelli G, Mari A *et al*: Detection of an atypical 7q11.23 deletion in Williams Syndrome patients which does not include STX1A and FZD3 genes. *Am J Hum Genet* 1998; **63**(4): (abstract supplement) 439.