FEBS Letters 581 (2007) 1233–1242

GTF2IRD1 regulates transcription by binding an evolutionarily conserved DNA motif 'GUCE'

P.D. Thompson[a], M. Webb[a], W. Beckett[a], T. Hinsley[a], T. Jowitt[b], A.D. Sharrocks[b], M. Tassabehji[a,*]

[a] *Academic Unit of Medical Genetics, The University of Manchester, St Mary's Hospital, Hathersage Road, Manchester M13 0JH, UK*
[b] *School of Life Sciences, The University of Manchester, Manchester, UK*

Received 26 October 2006; revised 14 February 2007; accepted 16 February 2007

Available online 28 February 2007

Edited by Ivan Sadowski

Abstract ***GTF2IRD1*** **is a member of a family of transcription factors whose defining characteristic is varying numbers of a helix–loop–helix like motif, the I-repeat. Here, we present functional analysis of human GTF2IRD1 in regulation of three genes (*HOXC8*, *GOOSECOID* and *TROPONIN* I_{SLOW}). We define a regulatory motif (GUCE–GTF2IRD1 Upstream Control Element) common to all three genes. GUCE is bound in vitro by domain I-4 of GTF2IRD1 and mediates transcriptional regulation by GTF2IRD1 in vivo. Definition of this site will assist in identification of other downstream targets of GTF2IRD1 and elucidation of its role in the human developmental disorder Williams–Beuren syndrome.**

Keywords: *GTF2IRD1*; TFII-I; GUCE; Transcription regulation; Williams–Beuren syndrome; DNA binding, isoforms

1. Introduction

The TFII-I family of transcription factors comprises three members (GTF2I (encoding TFII-I), *GTF2IRD1* and *GTF2IRD2*) all mapping to the region of chromosome 7 hemizygously deleted in the multisystem disorder Williams–Beuren syndrome (WBS) [1]. Our human and mouse studies have associated *GTF2IRD1* with craniofacial development [2] and evidence from patient studies implicates the other family members, particularly *GTF2I*, in the main clinical pathology of WBS [3–8].

Each of the TFII-I transcription factors has a putative leucine zipper (LZ) (a dimerisation motif) at the N-terminus and varying numbers of a repeated domain, known as an I-repeat [9,10]. The I-repeats are of particular interest since the domains appear to have DNA binding properties [11–14] and their predicted secondary structure indicates that the mechanism may be novel. Studies of the founder member, TFII-I, have shown that, although the I-repeats are capable of weak DNA binding [13], specificity of the protein for E-box sequences appears to be mediated through a basic region upstream of I-repeat 2, with the I-repeats involved in homomeric interactions [15]. Functional studies on GTF2IRD1 in different species have shown that it binds to a number of regulatory elements upstream of genes involved in tissue development/differentiation. *Xenopus* GTF2IRD1 (XWBSCR11) acts as a positive regulator of Xenopus goosecoid in response to activin [12]; mouse *Gtf2ird1* (BEN) interacts with a fragment of *Hoxc8* early enhancer [16] and plays a role in regulation of the mouse IgG variable heavy (*IgGVH*) promoter through a specific downstream regulatory element, DICE [17]; GTF2IRD1 binds to an element (USEB1), upstream of human *Troponin* I_{SLOW} (TnI_{SLOW}) and is necessary for high level expression of the gene in slow twitch muscle fibres [18]. GTF2IRD1 and TFII-I are thought to interact in regulation of transcription through the *c-fos*, *IgGVH*, *goosecoid* and *VFGFR2* promoters [17,19–21]. Both proteins also interact with HDAC3 and PIASxβ (a member of the E3 ligase family involved in the small ubiquitin-like modifier (SUMO) pathway), implicating them in transcription regulation through alteration of chromatin structure [22].

Defining the DNA binding sites for a transcription factor and understanding how a regulatory protein occupies its sites in vivo is central to understanding gene regulation. In this study we define a highly conserved DNA element, GUCE 'GTF2IRD1 Upstream Control Element', common to three genes regulated by GTF2IRD1 by a combination of comparative sequence analysis and both in vitro and in vivo binding assays. Binding of GTF2IRD1 to GUCE in vitro is mediated through protein domain I-4. Semi-quantitative analysis of the binding of I-4 to wild type and mutated GUCE sequences suggest the sequence is key for DNA recognition by I-4 in vitro. Recombinant reporter and CHIP assays demonstrate its importance for site recognition by GTF2IRD1 in vivo.

2. Materials and methods

2.1. Sequence alignments

Sequences upstream of *Goosecoid*, *Troponin* I_{SLOW} and *Hoxc8* coding regions were isolated from Genbank (where available), screened for the GUCE consensus and aligned using ClustalW (Supplementary methods A1).

2.2. Antibody production

Polyclonal antibodies were raised in rabbits to the region of human GTF2IRD1 between I-2 and I-3 and to a specific peptide within the N-terminal domain of GTF2IRD2 (Antibody Resource Centre, UK). GTF2IRD1 antibody was purified from crude serum by ammonium sulphate precipitation followed by affinity chromatography on a

*Corresponding author. Fax: +44 0 161 276 6606.
E-mail address: m.tassabehji@manchester.ac.uk (M. Tassabehji).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.02.040

thiopropyl sepharose 6B column (Amersham) conjugated with recombinant GTF2IRD1 peptide antigen.

2.3. Western blots and peptide blocking

Western blots of cell lysates were blocked for 16 h (4 °C in 10% Marvel in 1× TTBS) prior to incubation with antibodies to GTF2IRD1 (3 μg/ml) GTF2IRD2 (1:5000) or TFII-I (Transduction laboratories; 1:500) in 10% Marvel, 1 × TTBS, 2 h, room temperature. For peptide blocking the antibody was pre-incubated for 1 h in blocking solution with an excess (0.1 mg/ml) of the purified GTF2IRD1. For GAPDH antibody (Abcam) the antibody was diluted 1:10000. Secondary antibodies were goat anti-rabbit-HRP (Biosource) and sheep anti-mouse (Amersham Bioscience).

2.4. Chromatin immunoprecipitation

Experiments were carried out using 293HEK cells cultured in DMEM (Invitrogen) supplemented with 10% FBS, or human primary muscle cells cultured according to supplier's instructions (Promocell) and our GTF2IRD1 antibody or GAPDH antibody (Abeam), using published protocols [23].

2.5. PCR and QPCR

PCR was carried out using Reddy Mix PCR mix (Applied Biosystems) and Green JumpStart Taq Ready Mix (Finnzymes) was used for qPCR assays. Primers and cycling conditions are detailed in supplementary methods (A2).

2.6. DNA/protein expression constructs

The human cDNA clones *GTF2IRD*1β and γ (gi: 3913745 and gi: 18776294) were cloned into pCDNADEST40 (Invitrogen) for mammalian expression. *GTF2IRD*1α was engineered from *GTF2IRD*1β and γ. *GTF2IRD*1α*Δ*I4 was engineered from *GTF2IRD*1α and encodes a protein with amino acids 664–786 deleted and 662–663 altered ('LV' → 'VI'). Bacterial expression constructs were engineered using PCR products amplified from *GTF2IRD*1β and cloned into pDEST17 or pDEST15 (Invitrogen). Tagged peptides were expressed in *E. coli* and purified by affinity chromatography.

For the *HOXC8* reporter construct, sense – 5′-GATCTAAATCGGATTATAGGAAATCGGATTATAGGAAATCGGATTATAGGG-3′ and antisense-5-GATCCCCTATAATCCGATTTCCTATAATCCGATTTCCTATAATCCGATTT-3′ oligonucleotides were 5′ phosphorylated, annealed and two copies cloned into pGL3promoter (Promega). *HOXC8MUT* and *HOXC8LIMUT* reporter constructs contained mutations within the GUCE motifs (Fig. 6). pGL3*GSC*Prom was constructed by cloning a PCR amplified DNA fragment encompassing the human GSC promoter (nt-443 to +34) into the *Xho*I/*Hin*dIII sites of pGL3Basic (Promega).

2.7. Gel shift assays

Oligonucleotides used: *HOXC8*-Sense 5′-CTGGCACTTTCCTTTGAAATCGGATTATA-3′, *HOXC8*-Antisense 5′-GTGGTATAATCCGATTTCAAAGGAAAGTG-3′. Gel shifts were carried out at 50 mM KC1 essentially as described previously [24].

2.8. Circular dichroism

GTF2IRD1-I4-GST and GST peptides (at 0.3–0.5 mg/ml in buffer containing 50 mM NaCl and l0 mM phosphate) were loaded onto a 0.5 mm path-length cuvette (Starna). Circular dichroism spectra were analyzed at 20 °C between wavelengths of 260 and 190 nm on a Jasco J810 model circular dichroism spectrometer at 0.5 nm, 0.5 s response and are the average of 10 accumulations.

2.9. Quantitative gel shift assays

Wild type and mutated DNA fragments were prepared by annealing equimolar concentrations of sense and antisense oligonucleotides. DNA duplexes (50 pmol) were end-labelled with [γ-^{32}P] ATP and purified on Sephadex G50 columns (Amersham Biosciences). Following ethanol precipitation, the probes were resuspended in 10 mM Tris–HCl (pH 7.0), 50 mM NaCl. Binding was for 15 min at room temperature (in 10 mM Tris (pH 7.8), 50 mM NaCl, 5% glycerol and 1 mM DTT) with increasing amounts of protein (0, 0.1, 0.2, 0.5, 1, 2, 3, and 4 μM) incubated with a fixed concentration (250 nM) of labelled double stranded DNA in a final reaction volume of 10 μl. The reactions were resolved by electrophoresis on 6% non-denaturing polyacrylamide/TAE gels. The fraction of bound DNA was determined by quantitation of unbound probe in the dried gels [25] using electronic autoradiography (Packard Instant Imager).

2.10. Luciferase assays

293HEK cells were transfected with 6 μl Polyfect (Qiagen) per triplicate and a total of 800 ng DNA per well in 12 well plates. Luciferase levels were normalised for transfection efficiency by co-transfection of the control β-galactosidase pCH110 plasmid (Pharmacia) and activities of both were measured using the Dual-light luciferase assay system (Applied Biosystems) as described previously [26].

3. Results

3.1. Identification of a putative binding site for GTF2IRD1

Analysis of the sequence of promoters regulated by GTF2IRD1 and its homologues (*Xenopus Goosecoid* DE, mouse *Hoxc8* enhancer element, rat TnI_{SLOW} SURE and human TnI_{SLOW}USEB1) [12,16,18,27] identified a conserved 10 nucleotide motif of '5′-**AYMRGATTA**W-3′ present in all of them. Phylogenetic footprints were generated by aligning the motif in the context of flanking sequences upstream of the genes *Hoxc8*, *Goosecoid* and TnI_{SLOW} using sequences available in the databases (Fig. 1).

Overall, the evolutionary conservation data suggest a core conserved site of 5′-RGATTA-3′, with strong evidence of further conservation in flanking sequences. We called the putative 5′-AYNRGATTAWM-3′ DNA binding element, GUCE – 'GTF2IRD1 Upstream Control Element' (Fig. 1D).

3.2. Endogenous human GTF2IRD1 localises to GUCE-containing chromosomal regions in vivo

We performed chromatin immunoprecipitation (ChIP) to determine whether endogenous GTF2IRD1 is associated with the human regulatory elements containing GUCE in vivo.

Initially, we assessed the specificity of our antibody. Western blots from untransfected 293HEK and COS7 cell lysates identified a ∼115 kDa band (expected size), and three smaller bands of ∼75, 45 and 40 kDa, (Fig. 2Ai, lanes 1 and 3). 293HEK and COS7 cells transfected with a construct expressing full length, native GTF2IRD1 displayed more intense bands co-migrating with the ∼115 and 75 kDa endogenous proteins (Fig. 2Ai, lanes 2 and 4). Specificity of the bands was confirmed by incubating the primary antibody with the GTF2IRD1 peptide antigen prior to Western blotting. After this treatment no bands were detected in either transfected or untransfected cells (Fig. 2Ai, lanes 5–8). This indicates that the observed bands are isoforms of GTF2IRD1, with the 115 kDa band likely representing the full length protein and demonstrates both the presence of endogenous GTF2IRD1 in these cells and specificity of the antibody.

The GTF2IRD1 peptide antigen was designed to avoid the regions with greatest similarity to other members of the TFII-I family of proteins (i.e. the I-repeats and the leucine zipper). GTF2IRD1 antibody specificity was demonstrated on a western blot comprising of 3 identical lanes with equal amounts of 293HEK cell lysates; each was probed with an antibody against one of the family members. No TFII-I or GTF2IRD2 bands were detected with the GTF2IRD1 antibody (see Fig. 2Aii). Protein/DNA complexes were isolated from 293HEK cells and interrogated using ChIP assays.

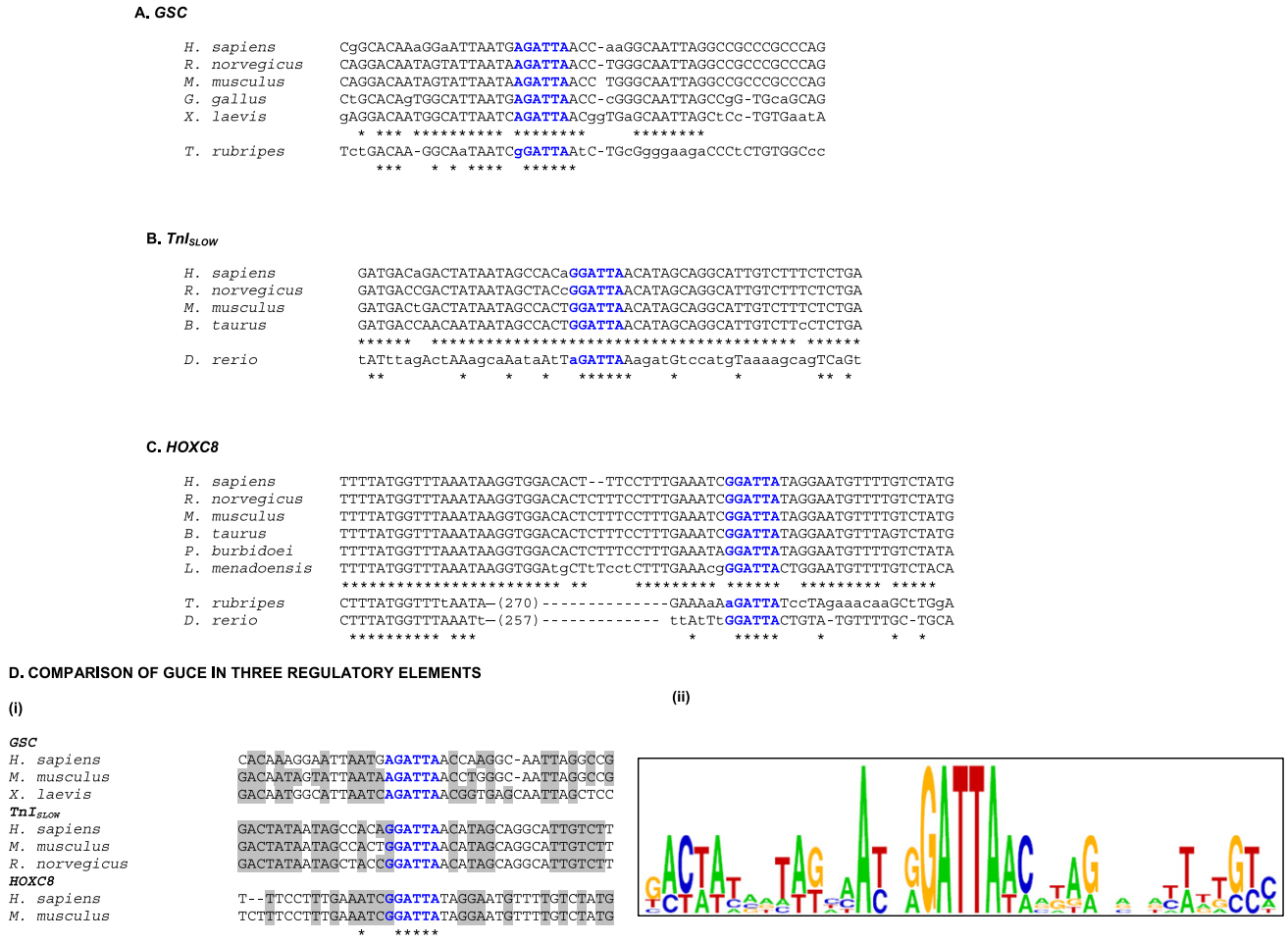

Fig. 1. Conservation of upstream elements containing the putative GTF2IRD1 binding site. Predicted GTF2IRD1 recognition sites located within the regulatory regions upstream of the genes are shown in blue. Identical nucleotides are starred and non-conserved bases shown in lower-case. Alignments of regulatory elements upstream of: (A) *Goosecoid* (B) TnI_{SLOW} and (C) *Hoxc8* – numbers in brackets indicate nucleotide gaps in two fish species between the 5′ Hoxc8 enhancer element and GUCE containing sequences. (D) Comparison of the putative GTF2IRD1 consensus across the different regulatory elements: (i) Nucleotides conserved between two of the three promoter regions are shaded. (ii) Sequence logo of the GUCE consensus using "Weblogo"[33]. Height of letters correlates with degree of conservation.

DNA was PCR amplified with specific primers designed from the human GUCE-containing regions and control primers to a genomic region flanking the *Calneuron* gene (*CALN*). Enrichment of DNA from the three GUCE-containing elements, but not the control, in DNA isolated with GTF2IRD1 was observed. Although the presence of each of the three elements before immunoprecipitation was confirmed (Input), ChIP assays using a control antibody (GAPDH) resulted in only very weak (background) amplification in each case (Fig. 2Bi).

In addition, complexes were isolated from cultured human primary skeletal muscle cells. Quantitative PCR demonstrated enrichment for a fragment containing the GUCE element upstream of TnI_{SLOW} when DNA isolated with the GTF2IRD1 antibody was used as a template, but not with the DNA associated with the GAPDH antibody. Again, no enrichment of the control genomic region (*CALN*) was observed (Fig. 2Bii).

Overall, these results suggest that endogenous human GTF2IRD1 is present at GUCE-containing sites in vivo under basal cell culture conditions.

3.3. Identification of a GTF2IRD1 GUCE-binding domain

In vitro gel shift assays were used to determine whether GTF2IRD1 interacts directly with GUCE and to identify the protein domain(s) involved. I-repeats (I-1–I-5) and the C-terminus region were purified as tagged peptides (Fig. 3A). Peptide I-4 was sufficient to bind to a human *HOXC8* GUCE 25 bp oligonucleotide probe (Fig. 3B). Peptide I-3 also displayed weak shifts on the *HOXC8* probe which may be due to non-specific DNA interactions since, unlike I-4, it also bound to a control non-specific probe tested alongside (data not shown). Sequence alignments of the I-repeats from all TFII-I family members highlight greater similarities between GTF2IRD1 repeat I-4, TFII-I repeat I-6 and GTF2IRD2 repeat I-2 compared to any of the other I-repeats [10], which may explain the selective binding properties of the I-4 domain for GUCE demonstrated here (Fig. 3C).

3.4. Semi-quantitative DNA binding analysis

To further delineate the interactions between the I-4 domain and the conserved GUCE motif, semi-quantitative DNA binding assays using purified GST-tagged I-4 were performed.

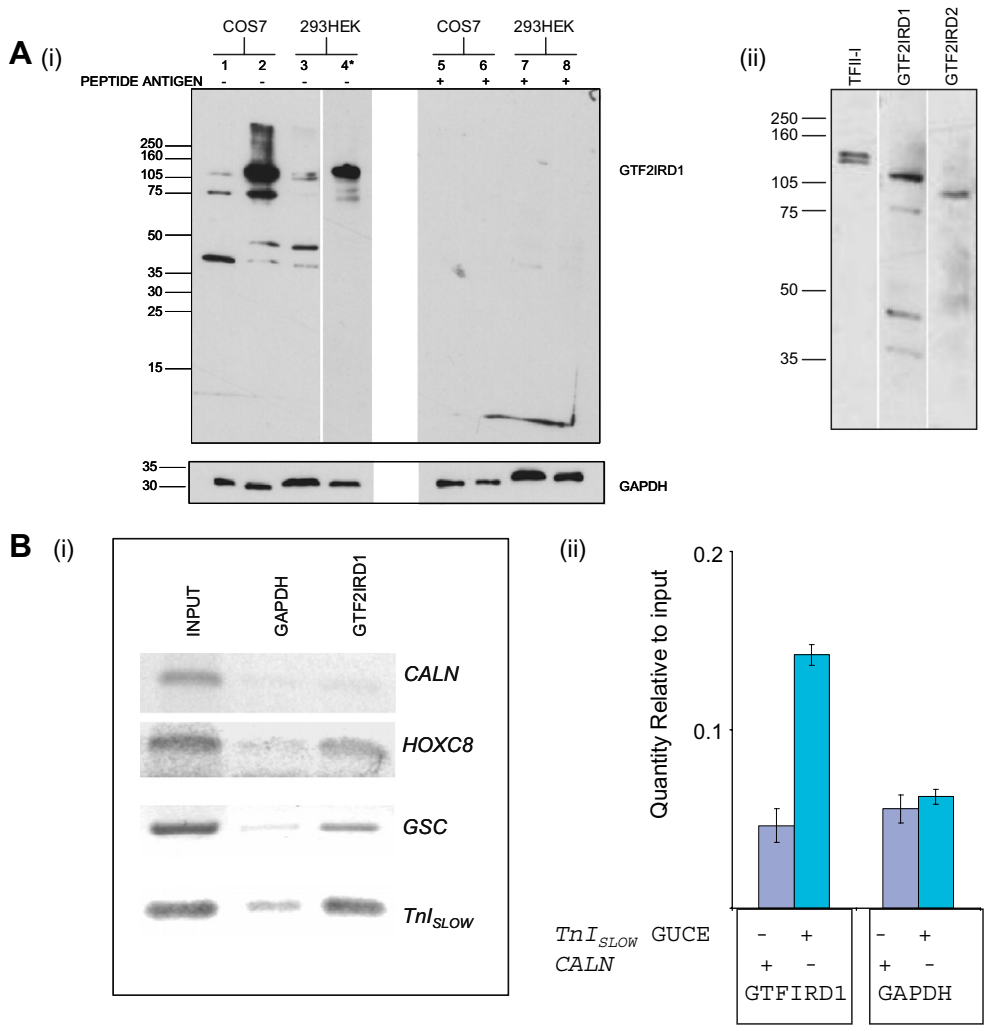

Fig. 2. Endogenous GTF2IRD1 localises to chromosomal sites containing GUCE. (A) Specificity of GTF2IRD1 antibody. (i) Western blots of whole COS7 (lanes 1–2 and 5–6) or 293HEK (lanes 3–4 and 7–8) cell lysates probed with GTF2IRD1 antibody, either unblocked (lanes 1–4) or blocked with excess peptide antigen (lanes 5–8); either untransfected (lanes 1, 3, 5, 7) or transfected with a construct expressing full length GTF2IRD1 (lanes 2, 4, 6, 8). * indicates a shorter exposure time for lane 4. The lower panel shows the blots probed with GAPDH antibody, (ii) Western blot demonstrating that antibody to GTF2IRD1 does not react to TFII-I family members. Whole cell lysates (30 μg total protein) from 293HEK cells probed with antibodies specific for TFII-I, GTF2IRD1 or GTF2IRD2. The bands detected by each antibody do not co-migrate with one another. (B) Chromatin immunoprecipitation assays, (i) Amplification products from control (*CALN*) or GUCE containing regulatory elements. Lanes: 1, input control; 2, GAPDH control antibody; 3, GTF2IRD1 specific antibody. (ii) Results of quantitative PCR of using template DNA co-immunoprecipitated from human primary skeletal muscle cells using GTF2IRD1 or a control (GAPDH) antibody. There is clear enrichment for the chromosomal region encompassing the GUCE motif upstream of TnI_{SLOW}, but not the control (*CALN*) region.

The structure of the GST-I-4 fusion protein was first analysed by circular dichroism (CD) spectroscopy to ensure that it adopted a folded conformation. To assess the relative content of secondary structure, deconvolution of the CD data for both GST-I4 and GST alone was performed using the CDSSTR algorithm [28] (Dichroweb). The GST protein alone showed a typical α-helical pattern with minima at 222 and 208 nm with evidence of some β-strand and random coil which is consistent with the known structure of the molecule. Comparison with the GST-I4 fusion protein showed a change in the overall structure. The structure of the GST-I4 fusion protein appears to have more β-sheet content, with a decrease in overall alpha helices and a similar fraction of random coil, indicating that the domain used in the binding assays has secondary structure elements consistent with a folded domain (Fig. 4A).

To define the important residues from the predicted consensus GUCE element, increasing amounts of I-4 were titrated against a fixed amount of each of 14 (one wild type GUCE and 13 mutated) radio-labelled oligonucleotide probes (13 bp) spanning the human *GSC* GUCE, and the binding affinity compared (Fig. 4B). As expected, the control GST peptide did not show any evidence of DNA binding.

The fraction of bound oligonucleotide was plotted against $\log_{10}$ protein concentration (nM) and the mean concentration at which 50% of the oligo was bound was used to estimate dissociation constants (K_ds) for each of the 14 DNA duplexes (Fig. 4C). The K_d for WT GUCE was estimated as 0.474 ± 0.06 μM. Other than mutation M1, which I-4 bound with a similar affinity to WT (K_d -0.434 ± 0.19 μM) all of the mutations within the predicted consensus resulted in some decrease in affinity of I-4 for the DNA. Mutation of the 'T' at

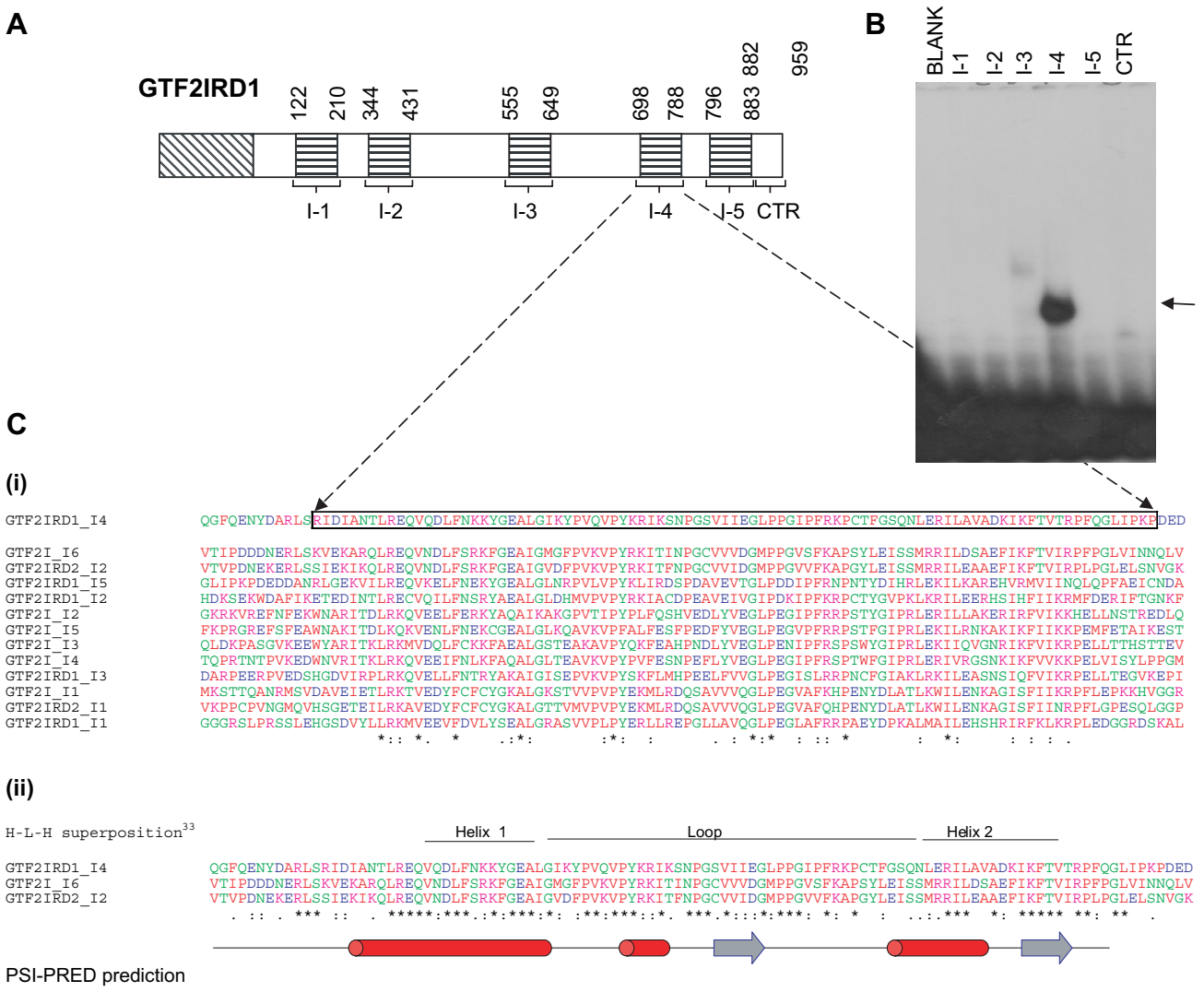

Fig. 3. GTF2IRD1 binds GUCE via I-4 in vitro. (A) Schematic of GTF2IRD1 (gi: 6635333) showing amino acid residues cloned and expressed in *E. coli* for band shift assays. I-1 = Aa residues 122–210; I-2 = residues 344–431; I-3 = residues 555–649; I-4 = residues 698–788; I-5 = residues 796–883; C-terminal region (CTR) = residues 882–959. (B) Gel mobility shift assays of ^{32}P-labelled oligonucleotide probes from the *HOXC8* upstream enhancer. Lanes: 1, no peptide control; 2–6, domains; 11–15, respectively; 7, GTF2IRD1-CTR.

position 8 (M10) is particularly detrimental, resulting in an approximately four-fold decrease in the affinity of I-4 for the sequence (K_d 2.109 ± 0.41 μM). Surprisingly, due to the lack of evolutionary conservation, changing the nucleotide at position 3 from 'G' to 'C' (M4) also has a marked effect on the DNA binding affinity of I-4. This may reflect species specificity of human GTF2IRD1 for its cognate *GSC* site. Mutation of the 'W' (A/T) at position 10 has only a marginal effect on affinity of I-4 (K_d −0.786 ± 0.42) which is unlikely to be significant. Therefore, based on these binding studies alongside evolutionary conservation data, we propose a consensus of 5′-**AYNRGATTANM**-3′ for I-4 of GTF2IRD1.

3.5. All known isoforms of GTF2IRD1 can repress transcription through GUCE

Two splice variants of human GTF2IRD1 have been characterised previously and are referred to here as the α (gi:4680483, [29]) and β (gi:6635332, [30]) isoforms (predicted M_w 105 and 106 kDa respectively), the latter having a 45bp extension of exon 19 (encoding 15 amino acids) relative to the former. We have identified another isoform, designated GTF2IRD1γ (gi: 49781271) through database searches. This human cDNA clone isolated from adult hippocampus (gi: 18776294) contains an additional 96 bp of sequence (an extension of exon 4) giving a predicted protein size of 108 kDa. Sequences of GTF2IRD1 orthologues demonstrate strong conservation in vertebrate species including mammals, birds, amphibians and fish. However, the peptide sequences encoded by alternative splicing are less well conserved. There is evidence for the presence of both extended exons in primates, and also of the extended exon 19 (*GTF2IRD1β*) in cattle (Fig. 5), however, neither exon is discernible in rodents (*M. musculus*, *R. norvegicus*) or other vertebrates.

Reporter assays were carried out to determine the role of the GUCE element in mediating transcriptional control by GTF2IRD1 in vivo. Luciferase reporter constructs containing 6 × 15 bp *HOXC8* GUCE elements were prepared (pGL3-*HOXC8*). Co-transfection of plasmids expressing the

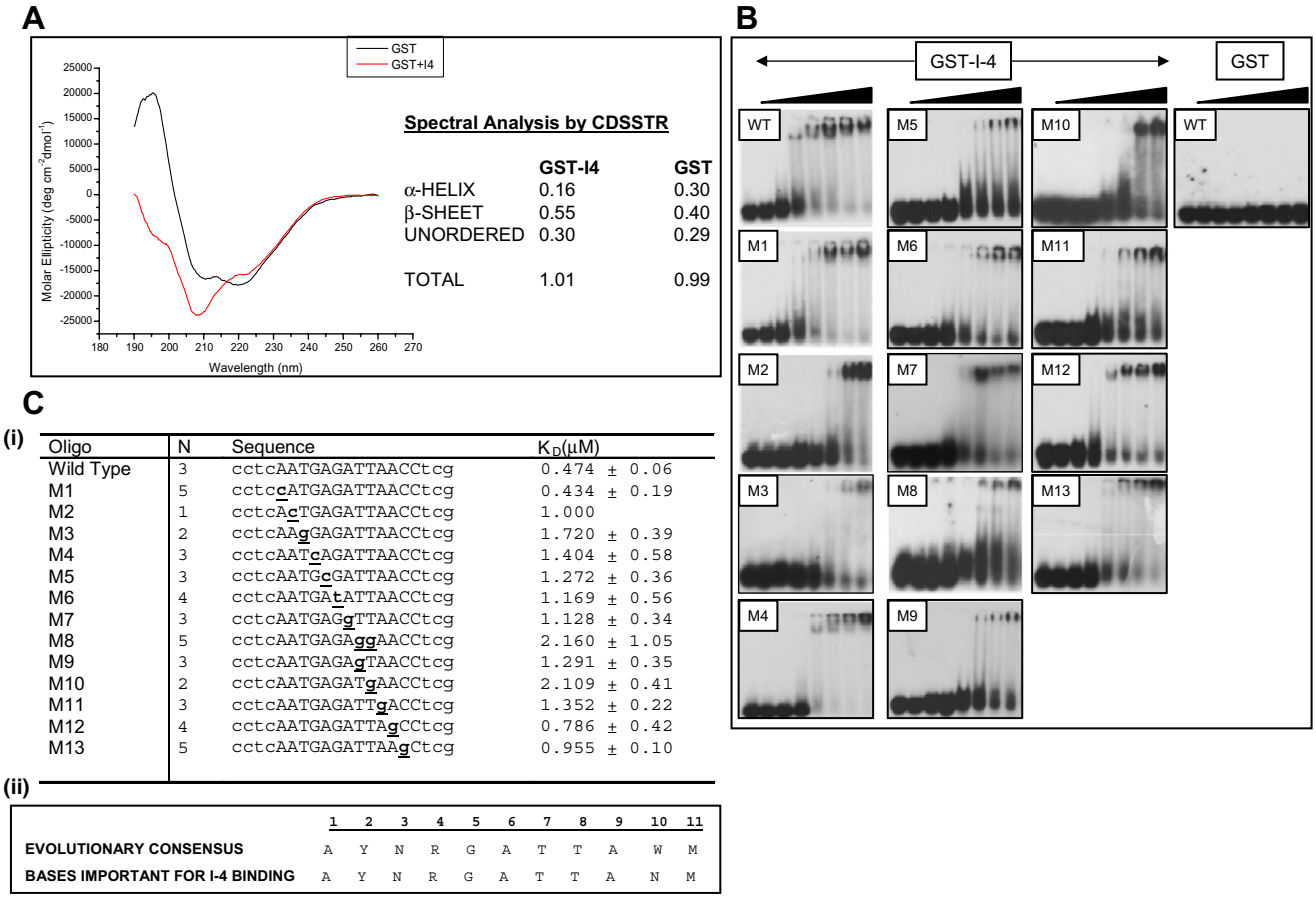

Fig. 4. (A) Circular dichroism spectra of GST alone (black line) and GST-I4 fusion protein (red line). Table shows CDSSTR analysis. (B) Representative gels used for semi-quantitative DNA binding analysis of I-4 to wild type GUCE (WT) and mutated (M1–M13) DNA probes (oligonucleotide sequences in C). The first lane is a no peptide control and the subsequent lanes show ^{32}P-labelled DNA incubated with increasing concentrations of GST or GST-I4 (0.1, 0.2, 0.5, 1.0, 2.0, 3.0 and 4.0 μM). (C) (i) Table showing K_d values calculated from semi-quantitative DNA binding analysis (*N* indicates the number of replicates). (ii) Comparison of the consensus sequences derived from the evolutionary and binding studies.

GTF2IRD1 isoforms with pGL3-*HOXC8* resulted in repression of luciferase activity in each case (Fig. 6A).

3.6. Deletion of domain I-4 diminishes, but does not abolish, transcription repression by GTF2IRD1

To determine the role of I-4 in GTF2IRD1 mediated transcriptional regulation a mutated form of GTF2IRD1α lacking amino acids 664–786 (GTF2IRD1αΔI4) was constructed and its ability to modulate transcription from a GUCE containing reporter construct (pGL3*GSC*Prom) compared to that of the wild type protein. GTF2IRD1αΔI4 retained the ability to mediate transcriptional repression from the reporter construct, however, in comparison to the wild type protein, the repression was significantly diminished (Fig. 6B). These results confirm that I-4 does contribute to transcription regulation by GTF2IRD1.

3.7. GUCE is necessary for transcription regulation by GTF2IRD1

To assess the importance of GUCE in transcription regulation by GTF2IRD1 the core element was mutated from 5′-GATTA-3′ to 5′-*CCCCC*-3′ or 5′-GA*AA*A-3′ (constructs pGL3-*HOXC8*MUT and *HOXC8*LIMUT). These mutations were sufficient to abrogate repression by GTF2IRD1 (Fig. 6C and D) demonstrating the importance of GUCE in mediating gene regulation by GTF2IRD1.

4. Discussion

GTF2IRD1 has recently been shown to play an important role in craniofacial development [2], however, little is known about its mechanism of action or downstream targets.

Phylogenetic footprinting, using sequence conservation as an indicator of functional significance, has been used to identify gene regulatory elements [31] and was adopted here as a preliminary *in silico* approach for GTF2IRD1. We identified a highly conserved DNA sequence in upstream regions of three genes previously linked with GTF2IRD1 [12,16,18,27]. Subsequent *in vitro* DNA binding studies using human protein confirmed and defined the DNA motif further, leading to the consensus 11 bp GUCE site 5′-AYNRGATTANM-3′ Domain I-4 of GTF2IRD1 (Aa 698–788) is involved in binding GUCE, agreeing with and refining previous studies on the TnI_{SLOW}

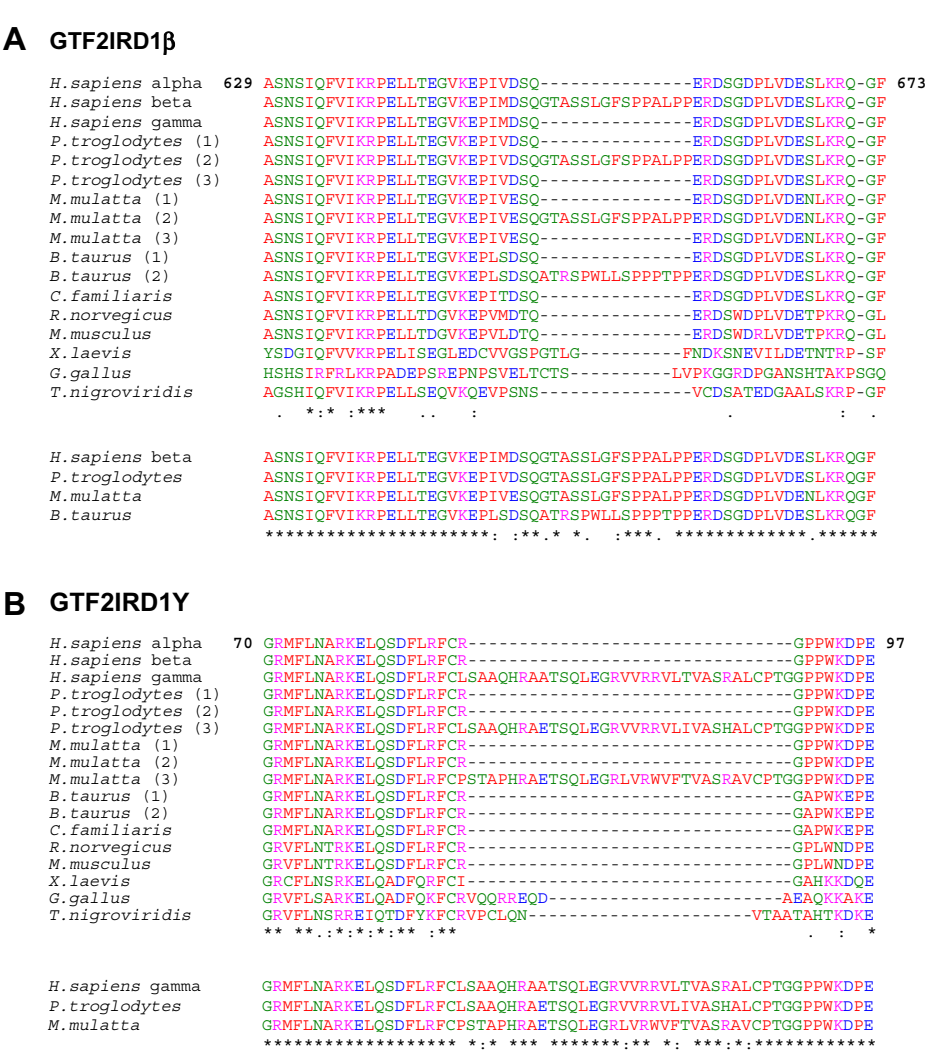

Fig. 5. Alignments of GTF2IRD1 peptide sequences encoded by alternatively spliced isoforms (GTF2IRD1β and γ). Sequences (Accession numbers in Supplementary methods A3) were compiled in FASTA format and aligned using ClustalW.

USEB1 sequence (an 18 bp region that contains a GUCE motif) which demonstrated the binding properties of I-4 (Aa 544–786) [11]. The demonstration that, of GTF2IRD1 I-repeats, only I-4 appears to interact specifically with GUCE is interesting in light of the fact that our analyses of I-repeats, from all TFII-I family members in both human and mouse, place I-4 in a separate clade (along with I-2 of GTF2IRD2 and I-6 of TFII-I) from other GTF2IRD1 I-repeats and highlight specific amino acid differences within I-4 which may play a functional role [10]. We have demonstrated the functional significance of I-4's DNA binding capabilities by showing that a mutated form of GTF2IRD1α lacking I-4 has a reduced ability to repress transcription from a reporter construct containing GUCE. This data confirms and extends work using an I-4-VP16 fusion construct to activate transcription from the TnI_{SLOW} GUCE containing element [13]. However, although both sets of data confirm the functional significance of I-4 associated DNA binding, our data also demonstrate that GTF2IRD1 is capable of transcription repression in the absence of this domain. This suggests that for full repressive function either GTF2IRD1 requires the involvement of other domains or alternatively the repression may be partially independent of DNA binding.

Our data also confirms and extends that obtained by *in vitro* SELEX methods [13] which proposes a binding site of 5′(G/A)GATT(G/A)3′ for mouse gtf2irdl I-4. In contrast, our semi-quantitative DNA binding analysis demonstrates that substituting the last guanine for an adenine in this consensus significantly decreases (~3-fold) the affinity of human GTF2IRD1 I-4 for the sequence, which is reinforced by the strong conservation of this nucleotide within known GTF2IRD1 target sites. In addition, residues flanking the core 'RGATTA' site influence the binding of I-4 to GUCE. In vivo chromatin immunoprecipitation (ChIP) confirmed the interaction of endogenous GTF2IRD1 with specific chromosomal sites and the GUCE motif was shown to be necessary for repression of reporter gene expression by GTF2IRD1. Transcription regulation through GUCE appears not to be directly affected by peptides encoded by the alternatively spliced exons of the human gene, as all three isoforms of GTF2IRD1 tested repressed expression of the pGL3-*HOXC8* reporter to a similar extent.

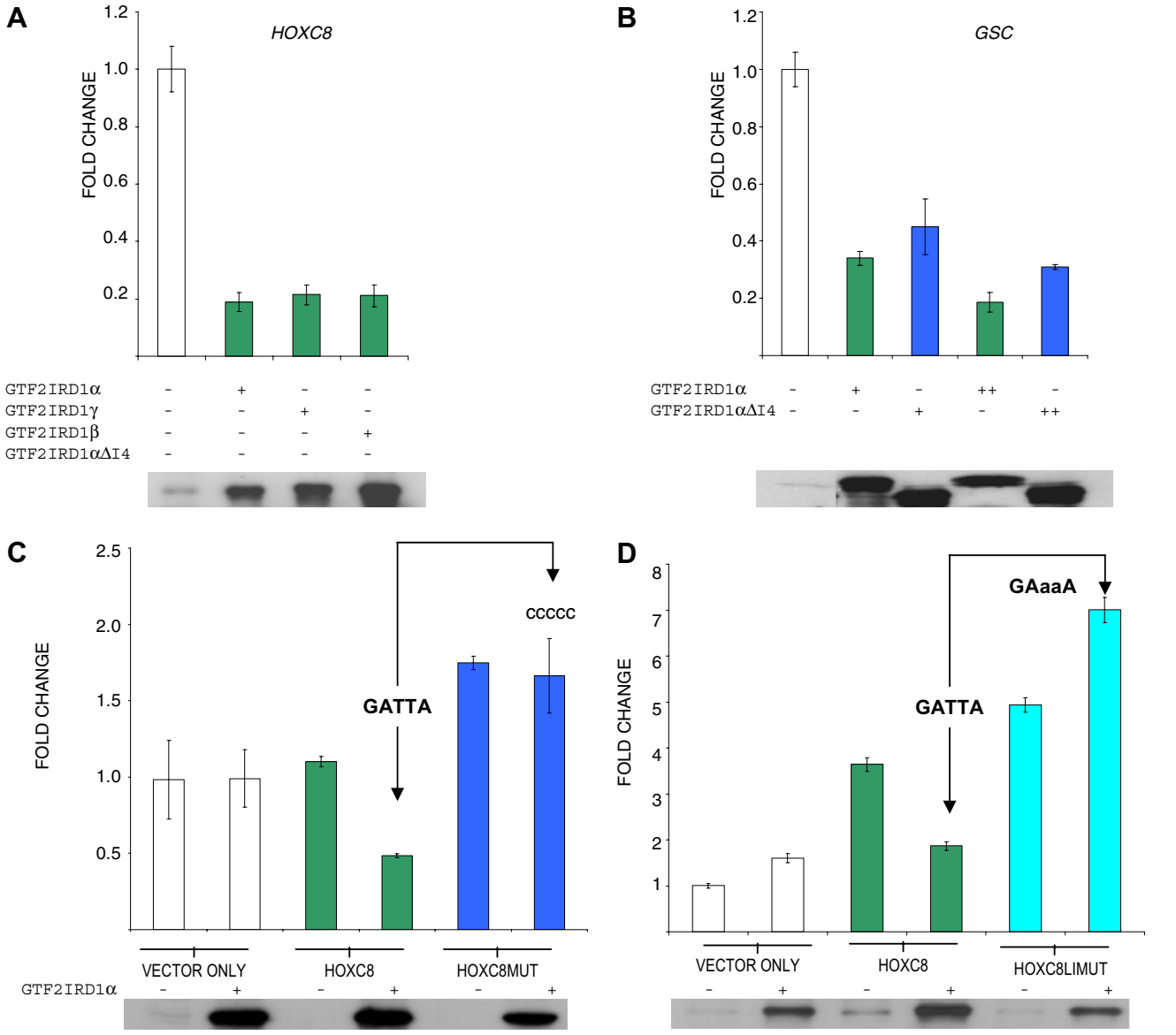

Fig. 6. Luciferase assays showing that; (A) all three GTF2IRD1 isoforms can repress through a reporter containing GUCE; (B) domain I-4 contributes to transcription regulation by GTF2IRD1. A reduction in the level of repression from pGL3*GSC*Prom is observed with increasing concentrations of GTF2IRD1αΔ-I4 compared with the wild type protein (GTF2IRD1α). (C and D) The GUCE sequence motif is necessary for repression of the reporter by GTF2IRD1. GTF2IRD1 is only able to repress expression from pGL3*HOXC8*, containing an intact GUCE motif. 293T HEK cells were transfected with, pGL3PROM vector, pGL3*HOXC8*, pGL3*GSC*Prom pGL3*HOXC8*MUT or pGL3*HOXC8*LIMUT, either alone or with DEST40*GTF2IRD1*α, β, γ or αΔ-I4. Results are expressed as mean(±S.E.M.) of triplicate samples, are representative of at least two independent experiments and are shown as fold change relative to control experiments ((A) pGL3*HOXC8* and (B) pGL3*GSC*Prom reporter constructs alone or (C) and (D) pGL3PR0M vector alone). Panels below show Western blots of cell lysates used for luciferase assays probed with GTF2IRD1 antibody.

The exact role of GTF2IRD1 in regulation of *goosecoid* is somewhat contentious. An initial study demonstrated activation of *Gsc* by *Xenopus* GTF2IRD1 in response to Activin [12]. In contrast, a second study suggested that TFII-I family proteins oppose one another in the regulation of *goosecoid*, with TFII-I activating in response to TGFβ/activin signalling and GTF2IRD1 able to repress this response [21]. The latter study also demonstrated the constitutive presence of Gtf2ird1 at the *Gsc* promoter in mouse P19 cells (in the absence of TGFR signalling), which corroborates our observation that GTF2IRD1 localises to the GUCE containing element upstream of this gene in human cells under standard culture conditions.

Mouse gtf2irdl (BEN) can bind to the EFG site within the early enhancer (EE) of mouse *Hoxc8* [16] and our data indicate that the human equivalent of this enhancer motif mediates transcriptional repression by GTF2IRD1 through the GUCE consensus. The importance of the *Hoxc8* EE for transcriptional control was demonstrated in mutant mouse models deleted for the 200 bp *Hoxc8 EE* (encompassing GUCE) [32] which showed significant delay in the temporal expression of *Hoxc8* and also shares aspects of its phenotype with our

Gtf2ird1 null mouse[2]; both display an abnormal clasping of the limbs when suspended by the tail. This leads us to speculate that aspects of the neurological phenotypes observed in WBS could be influenced by GTF2IRD1 induced misregulation of *HOXC8*.

In conclusion, GTF2IRD1 is an interesting and unusual transcription factor, particularly in light of its role in the pathogenesis of WBS, and may facilitate deeper understanding of the molecular mechanisms underlying human development and cognition. Defining the novel regulatory element, GUCE, will help identify other targets of GTF2IRD1 and aid the elucidation of its biological role.

Acknowledgements: The authors thank Anne Clancy and Irene Smeaton for excellent technical assistance and Lin Zhang and Helen Garside for helpful discussions. This work was funded by The Wellcome Trust (Grant #061183).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2007.02.040.

References

[1] Ewart, A.K., Morris, C.A., Ensing, G.J., Loker, J., Moore, C., Leppert, M. and Keating, M. (1993) A human vascular disorder, supravalvular aortic stenosis, maps to chromosome 7. Proc. Natl. Acad. Sci. USA 90, 3226–3230.
[2] Tassabehji, M., Hammond, P., Karmiloff-Smith, A., Thompson, P., Thorgeirsson, S.S., Durkin, M.E., Popescu, N.C., Hutton, T., Metcalfe, K., Rucka, A., Stewart, H., Read, A.P., Maconochie, M. and Donnai, D. (2005) GTF2IRD1 in craniofacial development of humans and mice. Science 310, 1184–1187.
[3] Gagliardi, C., Bonaglia, M.C., Selicorni, A., Borgatti, R. and Giorda, R. (2003) Unusual cognitive and behavioural profile in a Williams syndrome patient with atypical 7ql 1.23 deletion. J. Med. Genet. 40, 526–530.
[4] Korenberg, J.R., Chen, X.-N., Hirota, H., Lai, Z., Bellugi, U., Burian, D., Roe, B. and Matsuoka, R. (2000) Genome structure and cognitive map of Williams syndrome. J. Cognit. Neurosci. 12, 89S–107S.
[5] Hirota, H., Matsuoka, R., Chen, X.N., Salandanan, L.S., Lincoln, A., Rose, F.E., Sunahara, M., Osawa, M., Bellugi, U. and Korenberg, J.R. (2003) Williams syndrome deficits in visual spatial processing linked to GTF2IRD1 and GTF21 on chromosome 7qll.23. Genet. Med. 5, 311–321.
[6] Botta, A., Novelli, G., Mari, A., Novelli, A., Sabani, M., Korenberg, J., Osborne, L.R., Digilio, M.C., Giannotti, A. and Dallapiccola, B. (1999) Detection of an atypical 7ql 1.23 deletion in Williams syndrome patients which does not include the STX1A and FZD3 genes. J. Med. Genet. 36, 478–480.
[7] Karmiloff-Smith, A., Grant, J., Ewing, S., Carette, M.J., Metcalfe, K., Donnai, D., Read, A.P. and Tassabehji, M. (2003) Using case study comparisons to explore genotype-phenotype correlations in Williams–Beuren syndrome. J. Med. Genet. 40, 136–140.
[8] Danoff, S.K., Taylor, H.E., Blackshaw, S. and Desiderio, S. (2004) TFII-I, a candidate gene for Williams syndrome cognitive profile: parallels between regional expression in mouse brain and human phenotype. Neuroscience 123, 931–938.
[9] Tipney, H.J., Hinsley, T.A., Brass, A., Metcalfe, K., Donnai, D. and Tassabehji, M. (2004) Isolation and characterisation of GTF2IRD2, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams–Beuren syndrome. Eur. J. Hum. Genet. 12, 551–560.
[10] Hinsley, T.A., Cunliffe, P., Tipney, H.J., Brass, A. and Tassabehji, M. (2004) Comparison of TFII-I gene family members deleted in Williams–Beuren syndrome. Protein Sci. 13, 2588–2599.
[11] Polly, P., Haddadi, L.M., Issa, L.L., Subramaniam, N., Palmer, S.J., Tay, E.S. and Hardeman, E.C. (2003) hMusTRDlalphal represses MEF2 activation of the troponin I slow enhancer. J. Biol. Chem. 278, 36603–36610.
[12] Ring, C., Ogata, S., Meek, L., Song, J., Ohta, T., Miyazono, K. and Cho, K.W. (2002) The role of a Williams–Beuren syndrome-associated helix–loop–helix domain-containing transcription factor in activin/nodal signaling. Genes Dev. 16, 820–835.
[13] Vullhorst, D. and Buonanno, A. (2005) Multiple GTF2I-like repeats of general transcription factor 3 exhibit DNA binding properties. Evidence for a common origin as a sequence-specific DNA interaction module. J. Biol. Chem. 280, 31722–31731.
[14] Vullhorst, D. and Buonanno, A. (2003) Characterization of general transcription factor 3, a transcription factor involved in slow muscle-specific gene expression. J. Biol. Chem. 278, 8370–8379.
[15] Cheriyath, V. and Roy, A.L. (2001) Structure–function analysis of TFII-I. Roles of the N-terminal end, basic region, and I-repeats. J. Biol. Chem. 276, 8377–8383.
[16] Bayarsaihan, D. and Ruddle, F.H. (2000) Isolation and characterization of BEN, a member of the TFII-I family of DNA-binding proteins containing distinct helix–loop–helix domains. Proc. Natl. Acad. Sci. USA 97, 7342–7347.
[17] Tantin, D., Tussie-Luna, M.I., Roy, A.L. and Sharp, P.A. (2004) Regulation of immunoglobulin promoter activity by TFII-I class transcription factors. J. Biol. Chem. 279, 5460–5469.
[18] O'Mahoney, J.V., Guven, K.L., Lin, J., Joya, J.E., Robinson, C.S., Wade, R.P. and Hardeman, E.C. (1998) Identification of a novel slow-muscle-fiber enhancer binding protein, MusTRDl. Mol. Cell Biol. 18, 6641–6652.
[19] Tussie-Luna, M.I., Bayarsaihan, D., Ruddle, F.H. and Roy, A.L. (2001) Repression of TFII-I-dependent transcription by nuclear exclusion. Proc. Natl. Acad. Sci. USA 98, 7789–7794.
[20] Jackson, T.A., Taylor, H.E., Sharma, D., Desiderio, S. and Danoff, S.K. (2005) Vascular endothelial growth factor receptor-2: counter-regulation by the transcription factors, TFII-I and TFII-IRD1. J. Biol. Chem. 280, 29856–29863.
[21] Ku, M., Sokol, S.Y., Wu, J., Tussie-Luna, M.I., Roy, A.L. and Hata, A. (2005) Positive and negative regulation of the transforming growth factor beta/activin target gene goosecoid by the TFII-I family of transcription factors. Mol. Cell Biol. 25, 7144–7157.
[22] Tussie-Luna, M.I., Bayarsaihan, D., Seto, E., Ruddle, F.H. and Roy, A.L. (2002) Physical and functional interactions of histone deacetylase 3 with TFII-I family proteins and PIASxbeta. Proc. Natl. Acad. Sci. USA 99, 12807–12812.
[23] Garside, H., Stevens, A., Farrow, S., Normand, C., Houle, B., Berry, A., Maschera, B. and Ray, D. (2004) Glucocorticoid ligands specify different interactions with NF-{kappa}B by allosteric effects on the glucocorticoid receptor DNA binding domain. J. Biol. Chem. 279, 50050–50059.
[24] Roberts, E.C., Deed, R.W., Inoue, T., Norton, J.D. and Sharrocks, A.D. (2001) Id helix–loop–helix proteins antagonize Pax transcription factor activity by inhibiting DNA binding. Mol. Cell. Biol. 21, 524–533.
[25] Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H. and Murray, N.E. (1993) DNA recognition by the EcoK methyltransferase : the influence of DNA methylation and the cofactor *S*-adenosyl-methionine. J. Mol. Biol. 234, 60–71.
[26] Yang, S.-H., Yates, P.R., Whitmarsh, A.J., Davis, R.J. and Sharrocks, A.D. (1998) The Elk-1 ETS-domain transcription factor contains a mitogen-activated protein kinase targeting motif. Mol. Cell. Biol. 18, 710–720.
[27] Calvo, S., Vullhorst, D., Venepally, P., Cheng, J., Karavanova, I. and Buonanno, A. (2001) Molecular dissection of DNA sequences and factors involved in slow muscle-specific transcription. Mol. Cell Biol. 21, 8490–8503.
[28] Johnson, W.C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. Prot.: Struct. Funct. Genet. 35, 307–312.
[29] Osbome, L.R., Campbell, T., Daradich, A., Scherer, S.W. and Tsui, L.-C. (1999) Identification of a putative transcription factor

gene (WBSCR11) that is commonly deleted in Williams–Beuren syndrome. Genomics 57, 279–284.
[30] Tassabehji, M., Carette, M., Wilmot, C., Donnai, D., Read, A.P. and Metcalfe, K. (1999) A transcription factor involved in skeletal muscle gene expression is deleted in patients with Williams syndrome. Eur. J. Hum. Genet. 7, 737–747.
[31] Lenhard, B., Sandelin, A., Mendoza, L., Engstrom, P., Jareborg, N. and Wasserman, W. (2003) Identification of conserved regulatory elements by comparative genome analysis. J. Biol. 2, 13.
[32] Juan, A.H. and Ruddle, F.H. (2003) Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer. Development 130, 4823–4834.
[33] Crooks, G.E., Hon, G., Chandonia, J.-M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190.