

# *GTF2IRD2* from the Williams–Beuren critical region encodes a mobile-element-derived fusion protein that antagonizes the action of its related family members

Stephen J. Palmer<sup>1,\*</sup>, Kylie M. Taylor<sup>1</sup>, Nicole Santucci<sup>2</sup>, Jocelyn Widagdo<sup>1</sup>, Yee-Ka Agnes Chan<sup>2</sup>, Jen-Li Yeo<sup>2</sup>, Merritt Adams<sup>2</sup>, Peter W. Gunning<sup>3</sup> and Edna C. Hardeman<sup>1</sup>

<sup>1</sup>Neuromuscular and Regenerative Medicine Unit, School of Medical Sciences, The University of New South Wales, Sydney 2052, Australia

<sup>2</sup>Muscle Development Unit, Children's Medical Research Institute, Westmead, NSW, 2145, Australia

<sup>3</sup>Oncology Research Unit, School of Medical Sciences, The University of New South Wales, Sydney 2052, Australia

\*Author for correspondence (s.palmer@unsw.edu.au)

Accepted 9 July 2012

Journal of Cell Science 125, 5040–5050

© 2012. Published by The Company of Biologists Ltd

doi: 10.1242/jcs.102798

## Summary

*GTF2IRD2* belongs to a family of transcriptional regulators (including TFII-I and *GTF2IRD1*) that are responsible for many of the key features of Williams–Beuren syndrome (WBS). Sequence evidence suggests that *GTF2IRD2* arose in eutherian mammals by duplication and divergence from the gene encoding TFII-I. However, in *GTF2IRD2*, most of the C-terminal domain has been lost and replaced by the domesticated remnant of an in-frame hAT-transposon mobile element. In this first experimental analysis of function, we show that transgenic expression of each of the three family members in skeletal muscle causes significant fiber type shifts, but the *GTF2IRD2* protein causes an extreme shift in the opposite direction to the two other family members. Mating of *GTF2IRD1* and *GTF2IRD2* mice restores the fiber type balance, indicating an antagonistic relationship between these two paralogs. In cells, *GTF2IRD2* localizes to cytoplasmic microtubules and discrete speckles in the nuclear periphery. We show that it can interact directly with TFII-I $\beta$  and *GTF2IRD1*, and upon co-transfection changes the normal distribution of these two proteins into a punctate nuclear pattern typical of *GTF2IRD2*. These data suggest that *GTF2IRD2* has evolved as a regulator of *GTF2IRD1* and TFII-I; inhibiting their function by direct interaction and sequestration into inactive nuclear zones.

**Key words:** TFII-I, Transposon, Transgenic, Muscle, Nuclear speckles

## Introduction

*GTF2IRD2* belongs to a small family of related genes – *GTF2I*, *GTF2IRD1* and *GTF2IRD2* – that are clustered within the mammalian genome. The *GTF2IRD2* genes in the human lie within blocks of highly homologous low copy repeats (LCRs) that flank a domain containing 28 genes within 7q11.23 that is hemizygotically deleted in WBS (Osborne and Mervis, 2007). Haploinsufficiency of a subset of dosage-sensitive genes from within the deletion region is thought to be the cause of the most prominent features of this neurodevelopmental disorder. The disease is typified by a set of characteristic physical, cognitive and behavioral abnormalities. Of the 28 genes in the deleted contig, the only proven genotype-phenotype link is that haploinsufficiency of elastin (*ELN*) causes a range of connective tissue problems including supravalvular aortic stenosis (Morris and Mervis, 2000). Careful analysis of the features of patients carrying atypical deletions of the region has led to the conclusion that loss of *GTF2IRD1* and *GTF2I* can explain most of the other principal features including craniofacial dysmorphism, hypersociability and visuospatial deficits (Antonell et al., 2010). Recent work studying mice with mutations of the orthologous *Gtf2ird1* and *Gtf2i* genes supports this conclusion and phenotypes of craniofacial dysmorphism (Tassabehji et al., 2005; Lucena et al., 2010; Howard et al., 2012), reduced fear and aggression (Young et al., 2008) and increased social interactions (Sakurai et al., 2011) have been reported.

At least 2 copies of *GTF2IRD2* have been identified in the human as well as a *GTF2IRD2* pseudogene in the centromeric LCR containing a frameshift mutation. Due to the highly homologous nature of the duplicated regions, some doubt remains concerning the existence of a possible third copy (Tipney et al., 2004). It is currently unknown whether all/both of these genes are active at equal rates or whether only one is responsible for generating the majority of transcripts. Since the *GTF2IRD2* genes lie within the sites where illegitimate recombination occurs during the meiotic events that create the WBS deletion, it is possible that *GTF2IRD2* total expression is disrupted, leading to a variable reduction of *GTF2IRD2* protein in WBS patients. On these grounds, *GTF2IRD2* should be considered as a potential causative agent in the characteristic features of WBS.

TFII-I, *GTF2IRD1* and *GTF2IRD2* have an N-terminal leucine zipper and a series of highly conserved repeat domains (RDs or I-repeats) of unclear function. These RDs, which adopt a previously unknown fold according to NMR analysis (Doi-Katayama et al., 2007), are unique to this family. Conservation of the RDs between orthologs or between family members is much higher than any other regions of the protein, suggesting that these domains constitute an essential functional element. Some of the RDs within *GTF2IRD1* have been shown to have sequence-specific DNA binding properties (Polly et al., 2003; Vullhorst and Buonanno, 2005) and *GTF2IRD1* protein can bind to its own promoter region via simultaneous usage of two RDs making

contact with two independent DNA binding sites in the *GTF2IRD1* promoter (Palmer et al., 2010). RD2 of TFII-I has also been associated with sequence-specific DNA binding properties (Cheriyath and Roy, 2001; Roy, 2001). A large body of evidence implicates TFII-I as a signal-induced transcription factor with a variety of gene targets including *c-fos* (Roy, 2007). Other evidence indicates that phosphorylated tyrosines within RD2 and RD3 regulate a cytoplasmic interaction with PLC- $\gamma$  that controls agonist-induced calcium entry (Caraveo et al., 2006). A number of other interactions both in the cytoplasm and the nucleus suggest a complex set of potential roles that are yet to be refined (Casteel et al., 2002; Sacristán et al., 2004; Jiang et al., 2005; Roy, 2006; Tapia-Páez et al., 2008; Ren et al., 2011).

GTF2IRD2 shares strong homology with TFII-I in the N-terminal region (Tipney et al., 2004), but it deviates from the structure of TFII-I and GTF2IRD1 in two main ways. Firstly, it only has two RDs, which are most similar in sequence to RD1 and RD6 of TFII-I (Makeyev et al., 2004). Secondly, the C-terminal half of the protein is encoded by a single large exon that has derived from the insertion of a CHARLIE8 transposon element (Tipney et al., 2004) belonging to the hAT class (Rubin et al., 2001), which is in-frame with the N-terminal truncated TFII-I-like portion. This process has, therefore, removed the equivalents of RDs 2–5, containing the presumed DNA-binding domain, the phosphorylation sites in RD2 and 3 and the nuclear localization signal. It is therefore difficult to predict what function GTF2IRD2 may perform based on evolutionary history and sequence homology alone.

In this study we examine the hypothesis that GTF2IRD2 has retained some functional overlap with GTF2IRD1 and TFII-I, but has acquired new properties bestowed by the acquisition and adaptation of a randomly inserted mobile DNA element. To examine GTF2IRD2 function, we considered using the assays that have been previously applied to GTF2IRD1 and TFII-I, but most are based on specific assumptions concerning known gene targets and known mechanisms of transcriptional regulation (Manzano-Winkler et al., 1996; Kim et al., 1998; Tantin et al., 2004; Tapia-Páez et al., 2008), the nature and sequence specificity of DNA binding (O'Mahoney et al., 1998; Bayarsaihan and Ruddle, 2000; Calvo et al., 2001; Ring et al., 2002; Polly et al., 2003; Vullhorst and Buonanno, 2003; Vullhorst and Buonanno, 2005; Palmer et al., 2010) or specifically identified protein–protein interactions (Roy et al., 1993; Novina et al., 1999; Casteel et al., 2002; Sacristán et al., 2004; Jiang et al., 2005; Caraveo et al., 2006). In this case, it was not possible to make specific assumptions about mechanism of GTF2IRD2 action based on sequence similarity alone as the protein has diverged significantly. Therefore, we elected to utilize a skeletal-muscle-specific transgenic expression model system to compare the regulatory relationship between the members of this gene family since we previously published regulation of skeletal muscle fiber type by GTF2IRD1 (Issa et al., 2006). We show that GTF2IRD2 does impact on fiber-specific gene expression and has an antagonistic effect on skeletal muscle fiber type conversion compared to GTF2IRD1 and TFII-I $\beta$ . In cell culture, we find that GTF2IRD2 localizes to microtubules and to distinct punctate nuclear domains. When GTF2IRD1 and TFII-I $\beta$  are also present, these proteins colocalize with the GTF2IRD2 punctate nuclear domains. We show that GTF2IRD2 can interact directly with both GTF2IRD1 and TFII-I $\beta$ . Our findings are consistent with a mechanism in which GTF2IRD2

regulates the activity of GTF2IRD1 and TFII-I proteins by direct interaction and sequestration of the proteins to a novel nuclear compartment.

## Results

### Evolutionary history and conservation of the *GTF2IRD2* gene

Previous work has shown that GTF2IRD2 shares strong sequence homology with the protein TFII-I (Hinsley et al., 2004; Tipney et al., 2004) and their encoding genes (*GTF2IRD2* and *GTF2I*) are located near to each other on human chromosome 7 (Fig. 1A).

Analysis of current genome sequence from a variety of vertebrate species indicates that *GTF2IRD2* arose after the emergence of mammals, but prior to the radiation of the eutherian lineages. This is based on comparison of conservation of synteny in the region surrounding the *GTF2I* gene in various species (Fig. 1A). *GTF2IRD2* is not present in non-mammalian vertebrates or in the opossum as a representative of the marsupials. However, it is present in all of the eutherian mammals for which there is good sequence data in this region. It is likely that *GTF2IRD2* has arisen as a result of duplication of the *GTF2I* gene, insertion of an in-frame CHARLIE8 transposon and subsequent sequence divergence.

Strong conservation of the GTF2IRD2 protein sequence across all eutherian mammals suggests that this newly evolved gene has stabilized in all placental mammals and retains a functional importance that has been maintained (Fig. 1B). Sequence conservation within the CHARLIE8 transposon fossil domain (81% identity between human and mouse) is equivalent to the N-terminal half of the protein (76% identity human/mouse), suggesting that this region plays an equally important role in the function of GTF2IRD2.

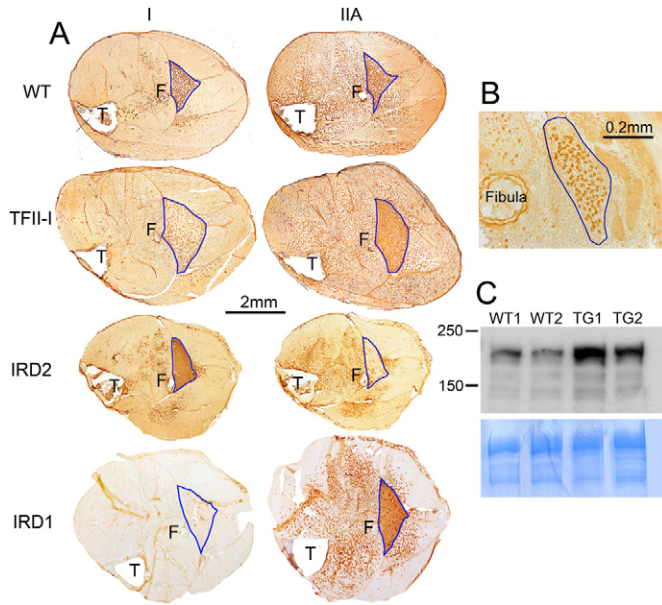
### Expression of TFII-I and GTF2IRD2 in muscle causes opposite fiber type shifts

Our previous work involving the transgenic expression of the GTF2IRD1 protein in muscle tissue proved a powerful means to show the repressive properties of GTF2IRD1 on slow fiber-specific gene expression and revealed a potential role for GTF2IRD1 in the development of skeletal muscle tissue (Issa et al., 2006). *Gtf2ird2* is ubiquitously expressed (Tipney et al., 2004) and thus a potential role in muscle tissue could also emerge from this assay. The transgenic muscle-specific expression model was used as a means to assess GTF2IRD2 function alongside the  $\beta$ -isoform of TFII-I for comparative purposes. The mouse *Gtf2ird2* ORF was cloned into an expression plasmid driven by the human skeletal actin (HSA) promoter/enhancer from the *ACTA1* gene that drives high levels of expression exclusively in skeletal muscle tissue (Brennan and Hardeman, 1993). An ORF (*Gtf2i $\beta$* ) encoding the mouse  $\beta$ -isoform of TFII-I (Roy, 2001) was cloned into the same vector so that direct phenotypic comparisons could be made between the activity of the two ancestral derivatives. The  $\beta$ -isoform of TFII-I was chosen because it is abundant in mouse tissues and has been shown to bind to target genes under basal conditions whereas other isoforms are known to require growth factor signaling-induced activation to achieve nuclear localization and DNA binding (Hakre et al., 2006).

Transgenic founder lines were analyzed by northern blot for levels of transgene expression and Southern blot genome





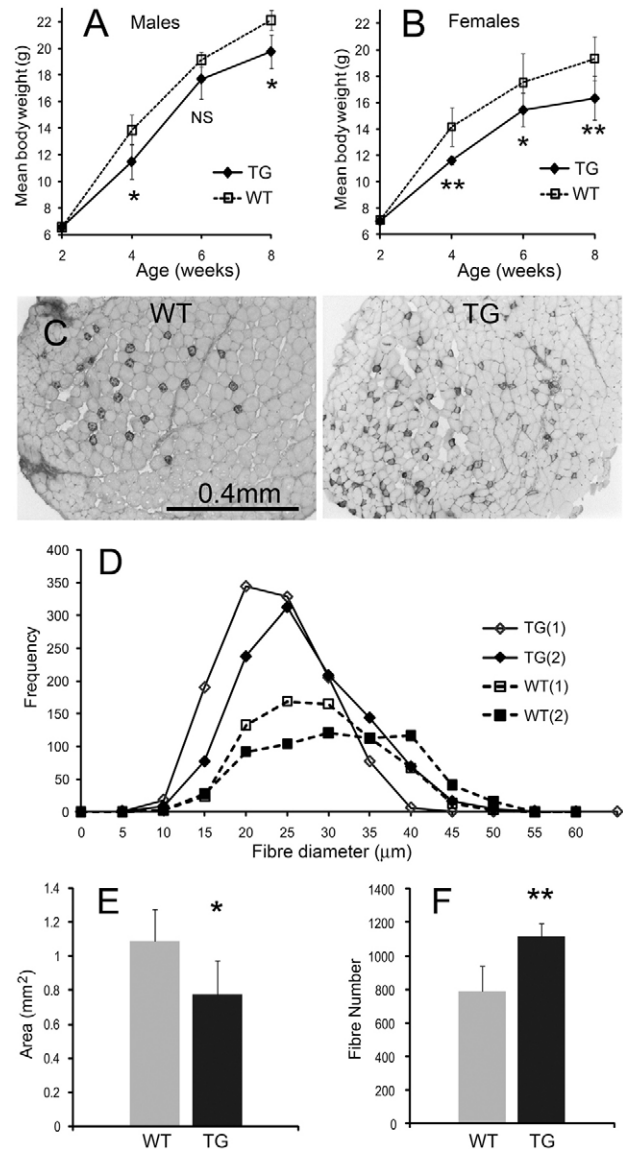


**Fig. 2. Transgenic expression of GTF2IRD2 and TFII-I $\beta$  in muscle tissue causes fiber type shifts in opposing directions.** (A) Whole sections through the lower hind limbs of wild-type (WT), TFII-I $\beta$  (TFII-I), GTF2IRD2 (IRD2) and GTF2IRD1 (IRD1) transgenic mice. The near-adjacent sections were stained for MyHC type I (slow) and fast type MyHC IIa. The soleus muscle is outlined in blue. T, tibia; F, fibula. (B) Whole section through the lower hind limb of a 2-day-old GTF2IRD2 transgenic mouse stained for MyHC type I showing normal distribution and numbers of stained and unstained muscle fibers. Outline indicates soleus muscle. (C) Western blot of soleus muscle protein extracts from two wild-type (WT) and two GTF2IRD2 transgenic (TG) mice probed with the anti-MyHC type I antibody (top panel). An identically loaded gel was stained with Coomassie Blue to show equivalent loading (bottom panel).

suggesting that these fiber type shifts are also due to post-natal conversion. Western blots of soleus muscle extracts from individual GTF2IRD2 transgenic animals, probed with anti-MyHC type I antibody, supported the immunohistochemistry results and illustrated the degree of MyHC type I upregulation associated with GTF2IRD2 expression (Fig. 2C).

#### GTF2IRD2 expression in muscle causes reduced fiber diameter and increased fiber numbers

Apart from the fiber type shifts, analysis of body weight and muscle histology revealed no other changes in muscle as a result of GTF2IRD1 (Issa et al., 2006) and TFII-I $\beta$  expression. In contrast, GTF2IRD2 led to a post-natal decline in body weight. Longitudinal body weight data (Fig. 3A,B) showed that weight differences were not significantly different at 2 weeks of age, but from 4 weeks onwards, male and female transgenic mice showed an 8–18% reduction in body weight compared with wild-type controls. In the highest expressing GTF2IRD2 line, it was possible to distinguish the adult transgenic animals from their wild-type siblings by external appearance alone, based on reduced muscle size and slow movement. Since the HSA promoter restricts transgene expression to skeletal muscle tissue, it is expected that this effect is attributable to loss of muscle mass; although, an impact on muscle function that affects movement and behavior could feed back on overall welfare. To analyze muscle fiber size in detail, individual extensor digitorum



**Fig. 3. GTF2IRD2 transgenic mice have reduced body weight, reduced muscle fiber size and increased fiber number.** (A,B) Body weight data from three litters of GTF2IRD2 transgenic and wild-type mice of 2–8 weeks of age (\* $P$ <0.05, \*\* $P$ <0.01, NS, not significant). (C) Transverse sections of extensor digitorum longus (EDL) muscles from a wild-type (WT) and a GTF2IRD2 transgenic (TG) mouse stained for MyHC type I showing the reduced average fiber diameter in the TG muscle and the increased incidence of type I fibers. (D) Analysis of all muscle fiber diameters in transverse sections of EDL muscle from two female transgenic (TG) and two female wild-type (WT) mice showing the frequency of fibers falling into 5  $\mu$ m bins. (E,F) Analysis of differences in total area (E) and total fiber number (F) in transverse sections of EDL muscles from five transgenic and five wild-type mice show that total transverse sectional area is reduced by ~29% (\* $P$ =0.0167), whereas fiber number is increased by approximately 41% (\*\* $P$ =0.00146).

longus (EDL) samples were collected from GTF2IRD2 transgenic mice and wild-type controls. Transverse sections were cut at the midpoint of the muscle to reduce the complication of muscle orientation and depth for quantitative analysis (Fig. 3C). Analysis of fiber diameter revealed an overall shift

in the entire EDL fiber population towards decreased thickness (Fig. 3D) and mean fiber diameter shows an approximate reduction of 18% from 27.2  $\mu\text{m}$  in the wild-type mice to 22.4  $\mu\text{m}$  in the GTF2IRD2 transgenic mice. This results in a significant reduction in mean whole muscle cross-sectional area (Fig. 3E). However, the muscle fiber size reduction was offset to some extent by an increase in muscle fiber number (Fig. 3F).

Qualitative examination of other muscles in the crural block (Fig. 2A) revealed that these effects are common to all muscles in the lower hind limb except for the soleus. Transverse sections of dissected soleus muscles revealed no significant alteration in the number of fibers with a mean total of 891 in the GTF2IRD2 transgenic mice compared with 888 in the wild-type mice and no significant difference in the mean fiber diameter or the range of sizes ( $n=4$  per genotype). However, analysis of fiber type revealed the same large increase in MyHC type I fibers seen in the crural sections with a variability that correlated approximately with the level of transgene expression in each line (data not shown).

#### GTF2IRD2 antagonizes the soleus fiber type shift caused by GTF2IRD1

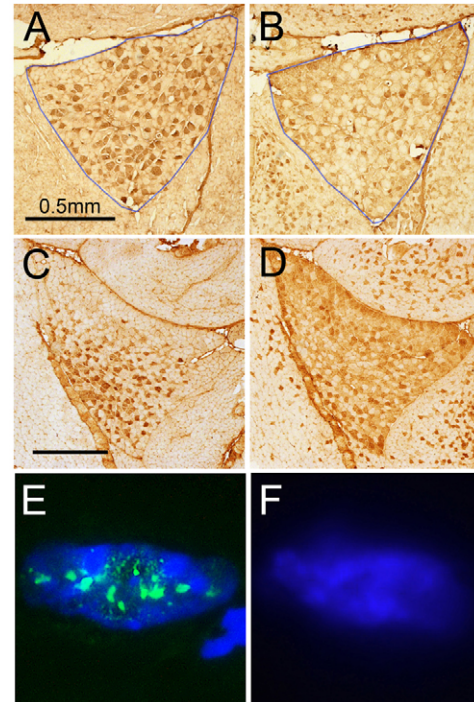
Since expression of human GTF2IRD1 protein in muscle leads to an extreme fiber type shift from slow (MyHC I) to fast (MyHC IIa) type (Issa et al., 2006) and TFII-I $\beta$  expression causes a similar, but less extreme shift in the same direction, whereas GTF2IRD2 causes the reverse shift from fast to slow (Fig. 2A), it seemed likely that these proteins operate on the same mechanism, but with antagonistic activities. To test this hypothesis in more detail, the GTF2IRD1 transgenic mice were crossed with the GTF2IRD2 line in order to generate double heterozygous *GTF2IRD1/Gtf2ird2* mice. Crural block sections were analyzed for fiber type distribution using immunohistochemistry (Fig. 4A–D). Although these soleus muscles showed irregular fiber size and unusual fiber type patterning, it was clear that the balance of slow and fast IIa fibers had been restored to some extent, suggesting a degree of functional complementation (Fig. 4A–D). In contrast, the reduction of fiber diameter in muscles outside the soleus was even more pronounced than in the GTF2IRD2 line alone and, unlike the GTF2IRD2 line, the soleus itself was also subject to reductions of fiber size (data not shown).

#### GTF2IRD2 localizes to punctate nuclear bodies in skeletal muscle fibers

It is assumed, based on transcript analysis, that endogenous GTF2IRD2 protein abundance is very low (Tipney et al., 2004). Immunofluorescent detection of the transgenic GTF2IRD2 protein was achieved using an anti-GTF2IRD2 antibody on longitudinal muscle sections from GTF2IRD2 transgenic mice and wild-type controls. Large and small punctate nuclear bodies could be seen (Fig. 4E) in the GTF2IRD2 transgenic samples. Although this antibody is known to cross-react with TFII-I (data not shown), the wild-type control muscle samples showed no signal (Fig. 4F) illustrating that the nuclear punctate bodies can be entirely attributed to the transgenic GTF2IRD2 protein.

#### GTF2IRD2 protein localizes to punctate nuclear domains and microtubules in cells in culture

As a means to assess the distribution of GTF2IRD2 protein in mammalian cells, we constructed GFP-tagged and Myc-tagged expression constructs containing the *Gtf2ird2* mouse cDNA. These

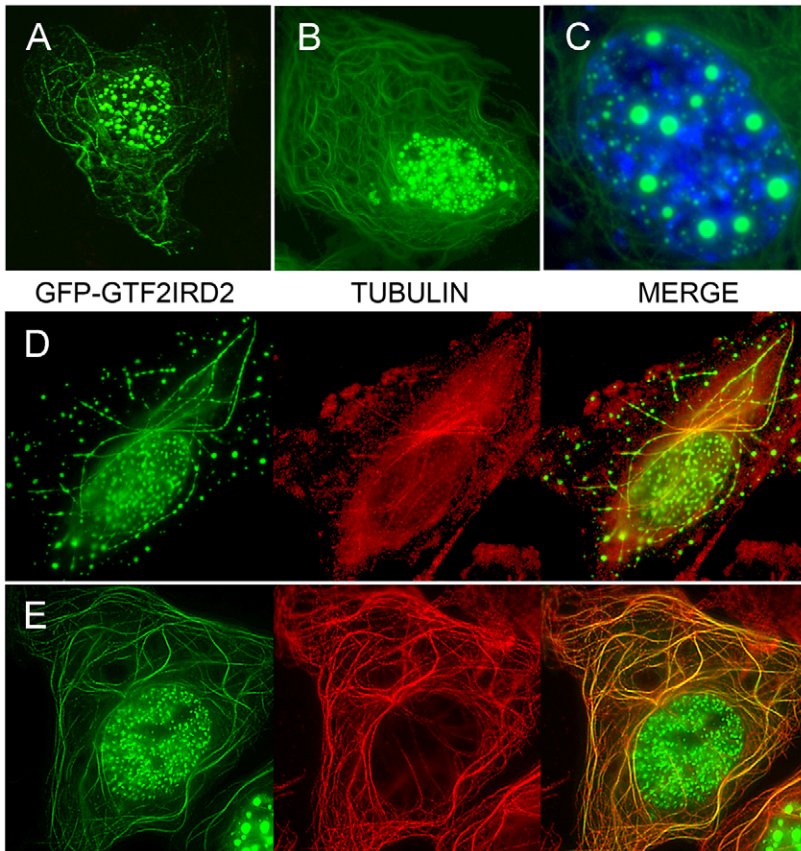


**Fig. 4. GTF2IRD1 and GTF2IRD2 antagonism in soleus muscles and distribution of GTF2IRD2 transgenic protein in muscle nuclei.** (A–D) Detail of soleus muscles in whole hind limb transverse sections from two mice carrying both the *GTF2IRD1* and the *Gtf2ird2* transgene. (A,C) Sections stained with anti-MyHC I; (B,D) near-adjacent sections stained with anti-MyHC IIa. (E,F) High power images of DAPI-stained nuclei from longitudinal muscle sections treated with anti-GTF2IRD2 antibodies and secondary immunofluorescent detection. GTF2IRD2 protein (green foci) is detectable in the GTF2IRD2 TG section (E) but not in the wild-type control section (F).

constructs were transfected into several cell lines including C2C12, COS-7, NIH3T3 and HeLa. The results were similar in all cell lines in that GTF2IRD2 protein was found to localize to nuclear foci of variable size and density as well as displaying cytoplasmic localization to a cytoskeletal element suspected to be microtubules (Fig. 5A–C). The microtubule association was confirmed using double immunofluorescence analysis with transfected GFP–GTF2IRD2 and an anti- $\alpha$ -tubulin antibody during (Fig. 5D) and after (Fig. 5E) nocodazole treatment, which blocks microtubule polymerization. Cytoplasmic GFP–GTF2IRD2 colocalized with fluorescently labeled  $\alpha$ -tubulin-containing microtubules, but during nocodazole treatment, the cytoplasmic GFP–GTF2IRD2 was scattered in vesicular-like structures (Fig. 5D).

In the nucleus, GFP–GTF2IRD2 and Myc–GTF2IRD2 localized to foci of two distinct size categories; large spherical inclusions and smaller, more numerous nuclear bodies. To identify the nature of the small nuclear bodies, GFP–GTF2IRD2-transfected cells were used in colocalization studies using antibodies against a variety of known punctate nuclear proteins (Fig. 6A–D). However, none of these proteins including SC35 (nuclear speckles/splicing bodies), Coilin (Cajal bodies), Pml (Pml bodies), Sam68 [SAM68 nuclear body (SNB)], Bmi1 (polycomb bodies) and Sp1 (OPT domains – data not shown) were found to colocalize consistently with GFP–GTF2IRD2. The same was also true of the large nuclear bodies. In order to test



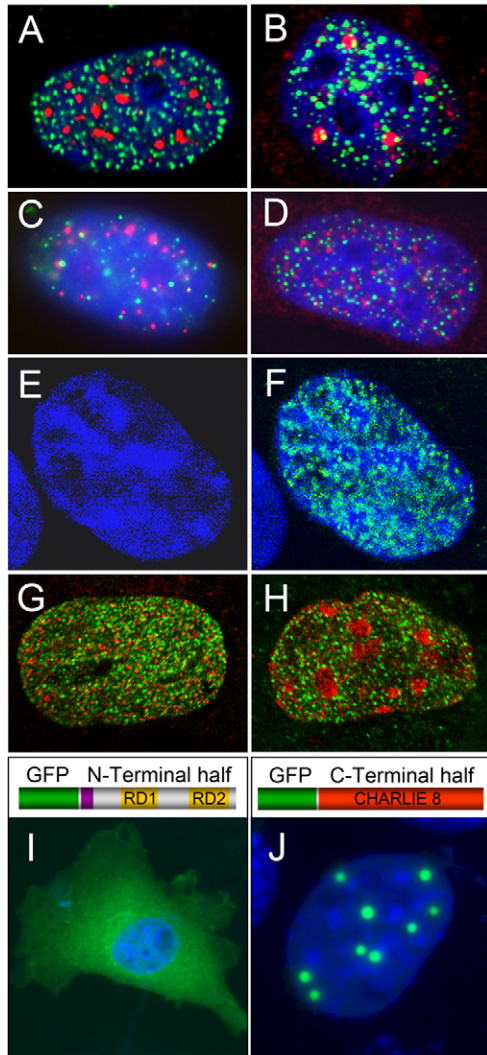


**Fig. 5. GTF2IRD2 localizes to punctate nuclear domains and microtubules.** (A–C) GFP-tagged GTF2IRD2 protein localizes to cytoplasmic microtubules and punctate nuclear bodies in NIH 3T3 cells (A) and COS-7 cells (B). (C) High power image of a DAPI-stained nucleus showing that the GFP–GTF2IRD2 bodies (green) can form both large spheres and smaller more numerous speckles. (D) GFP–GTF2IRD2-expressing COS-7 cells treated for 2 hours with nocodazole show disruption of the tubulin polymers and disruption of the associated GFP–GTF2IRD2, whereas the punctate nuclear bodies are unaffected. (E) Withdrawal of nocodazole for 1 hour permits re-formation of the tubulin cytoskeleton and re-association of GFP–GTF2IRD2 with the microtubules.

whether the GTF2IRD2 foci associate with inactive or active chromatin regions, GFP–GTF2IRD2 distribution was assessed in association with DAPI-stained chromatin using confocal imaging (Fig. 6E,F), which suggested some exclusion of GFP–GTF2IRD2 from dense heterochromatin. These results were explored further by examining the distribution of GFP–GTF2IRD2 in relation to active chromatin marked by trimethylated histone H3 lysine 4 (Fig. 6G) and inactive chromatin marked by trimethylated histone H3 lysine 9 (Fig. 6H). These data indicated that while the GFP–GTF2IRD2 foci appear to be excluded from regions of inactive heterochromatin, there is not a strong association with active chromatin either and it is unclear whether chromatin tethering plays any role in GTF2IRD2 distribution. However, by dividing the mouse *Gtf2ird2* cDNA into two halves at the junction formed by the beginning of the CHARLIE8 domain and cloning these into GFP expression plasmids, it was possible to examine the behavior of the N-terminal (aa1–411) and C-terminal (aa412–936) halves separately. The GFP–N-terminal domain was generally distributed in a broad cytoplasmic pattern (Fig. 6I) supporting the view that this region lacks a nuclear localization signal. When the GFP–C-terminal half was transfected into cells, all of the protein was localized exclusively to a small number of large spherical nuclear bodies (Fig. 6J) and none of the small nuclear bodies or microtubule associations were seen, thus suggesting that the large nuclear bodies can be attributed to associations formed by the CHARLIE8 domain alone and that both regions of the protein work in conjunction to create the association with microtubules and the small nuclear bodies.

#### Recombinant GTF2IRD1 and TFII-I colocalize with the GTF2IRD2 punctate nuclear domains located at the nuclear periphery

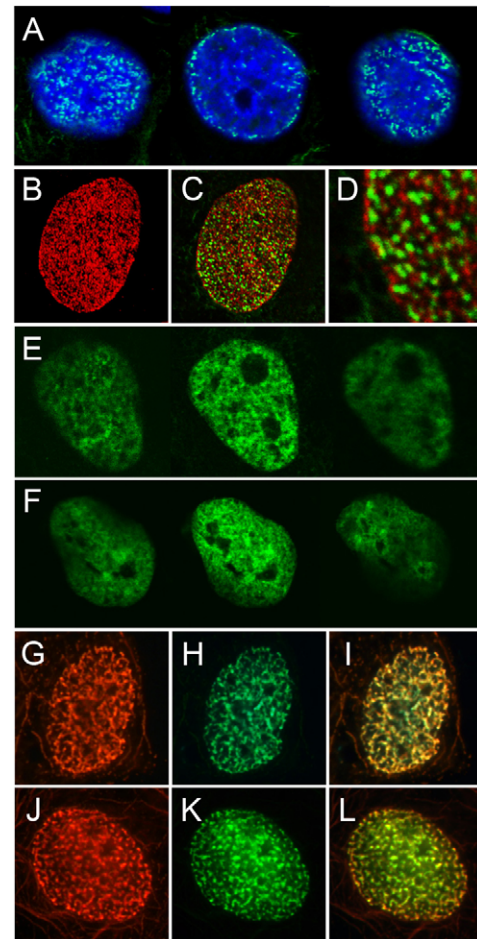
Confocal analysis of the distribution of Myc and GFP-tagged GTF2IRD2 protein in COS-7 cells revealed that the punctate nuclear domains were preferentially located at the periphery of the nucleus (Fig. 7A). Analysis of the location of the endogenous nuclear pore complexes (NPCs) revealed that there was no overlap in location and the GTF2IRD2 bodies occupied a slightly more interior zone than the NPCs (Fig. 7B–D). Analysis of the relative location of the lamin-associated protein LAP2B supported this conclusion (data not shown). Since the functional evidence suggests that GTF2IRD1, TFII-I and GTF2IRD2 act on a common mechanism in muscle tissue, we sought to examine whether all three family members were capable of colocalization in cultured cells. Firstly, the basal distribution of Myc and GFP-tagged GTF2IRD1 and TFII-I $\beta$  proteins were examined using confocal microscopy, which revealed a similar distribution. Fluorescence was detected throughout the nucleus in a broad patchy pattern, excluded from nucleoli, but otherwise evenly distributed with no evidence of a peripheral bias (Fig. 7E,F). By contrast, when GFP-tagged GTF2IRD1 and TFII-I were co-expressed with Myc–GTF2IRD2, it became obvious that these proteins had adopted the punctate nuclear pattern and both colocalized extensively with GTF2IRD2 (Fig. 7G–L). This colocalization required the presence of the N-terminal half of the protein as expression of the GFP–C-terminal half with Myc-tagged GTF2IRD1 resulted in separate distribution patterns (data not shown).



**Fig. 6. GTF2IRD2 nuclear bodies do not associate with recognized nuclear foci or chromatin domains and the large spherical inclusions are caused by the CHARLIE8 region.** Immunofluorescence detection of GFP-tagged GTF2IRD2 in transfected HeLa and COS-7 cells with various endogenous nuclear foci (red). (A–D) GTF2IRD2 bodies do not show strong associations with SC35 nuclear speckles in COS-7 (A), coilin-containing Cajal bodies in HeLa (B), Pml bodies in HeLa (C) or Bmi1-identified polycomb bodies in HeLa (D). (E) Confocal image of a DAPI-stained nucleus and (F) the same image overlaid with the GFP–GTF2IRD2 distribution. (G) COS-7 nucleus showing the distribution of GFP–GTF2IRD2 relative to trimethylated H3K4 active chromatin marks. (H) COS-7 nucleus showing GFP–GTF2IRD2 and trimethylated H3K9 indicating GFP–GTF2IRD2 is largely excluded from inactive heterochromatic regions. (I, J) Schematics of GFP-tagged protein constructs containing the N-terminal or C-terminal halves of GTF2IRD2 with the corresponding cellular distribution shown beneath. (I) GFP–N-terminal GTF2IRD2 localizes to the cytoplasm in a diffuse perinuclear pattern. (J) GFP–C-terminal GTF2IRD2 localizes exclusively to a small number of large round nuclear inclusions.

#### GTF2IRD2 interacts directly with GTF2IRD1 and TFII-I

Colocalization of these three protein family members could be due to binding of the proteins to common protein complexes in the cell or it could indicate direct interaction between these proteins. Reciprocal co-immunoprecipitation experiments revealed that recombinant GTF2IRD2, GTF2IRD1 and TFII-I



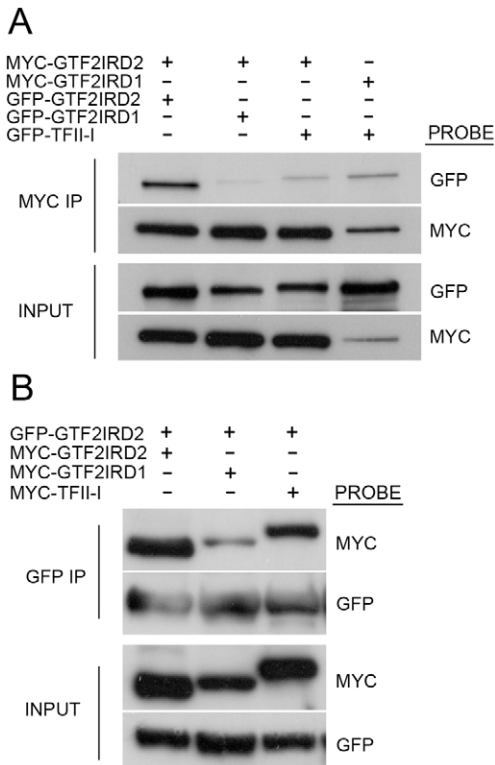
**Fig. 7. GTF2IRD2 localizes peripherally in the nucleus and influences the subnuclear distribution of GTF2IRD1 and TFII-I $\beta$ .** (A) A series of confocal z-sections through the same nucleus showing the peripheral distribution of GFP–GTF2IRD2. (B) Nucleus showing detection of endogenous nuclear pore complex (NPC). (C) Detection of NPC (red) and GFP–GTF2IRD2 (green) showing a lack of colocalization. (D) Higher power image from C showing how the GFP–GTF2IRD2 sits inside the plane of the NPCs. (E, F) Series of confocal z-sections through two nuclei shows the distribution of (E) GFP–GTF2IRD1 and (F) GFP–TFII-I. (G–I) Detection of Myc–GTF2IRD2 (G) by immunofluorescence and fluorescent GFP–GTF2IRD1 (H). The overlaid image (I) shows extensive colocalization of the two proteins in the punctate pattern typical of GTF2IRD2. (J–L) Similarly, detection of Myc–GTF2IRD2 (J) and fluorescent GFP–TFII-I (K) also shows extensive colocalization in the typical punctate pattern (L).

are all capable of cross interaction. The highest affinity found was between GTF2IRD2 and itself followed by TFII-I and GTF2IRD1 in decreasing order of affinity (Fig. 8).

#### Discussion

This study provides the first functional analysis of GTF2IRD2, which is of interest from two different perspectives. Firstly, this protein is the third member of a small family that are intimately concerned with the genetic cause of WBS (Antonell et al., 2010). Understanding the function of these genes will provide an essential insight into the cause of the disease and has the potential for insights into genetic mechanisms that underpin features of WBS patients; inroads into the function in one tissue informs on





**Fig. 8. GTF2IRD2 interacts with itself, TFII-I and GTF2IRD1 in decreasing order of affinity.** (A) Western blots of protein extracts from HEK293 cells transfected with combinations of constructs encoding GFP-tagged and Myc-tagged recombinant proteins before (INPUT) and after (MYC IP) immunoprecipitation using anti-Myc antibody to enrich for Myc-tagged protein complexes. The nature of the probe antibody is shown on the right. Band intensity indicates that GTF2IRD2 has a strong affinity for itself, weaker for TFII-I and weaker still for GTF2IRD1; GTF2IRD1 and TFII-I $\beta$  also co-immunoprecipitate together. (B) Western blot showing a reciprocal immunoprecipitation experiment to that in A, using GFP-GTF2IRD2 to precipitate the protein complexes. This analysis also shows a decreasing affinity of GTF2IRD2 for itself, TFII-I $\beta$  and GTF2IRD1 as above. Negative control immunoprecipitations using GFP only or Myc-tagged SETD7 showed no co-immunoprecipitation of target proteins (data not shown).

mechanism in other organs. Secondly, this protein is encoded by a gene that belongs to a group of functional human genes that have been derived from the random insertion of mobile element sequences into the genome and the exploitation of those sequences for functional purposes (Britten, 2004). It is interesting to consider how this exploitation has evolved and what properties are bestowed by the sequence derived from these mobile elements.

### Structure and evolution of GTF2IRD2

Previous analysis of GTF2IRD2 sequence indicated that it is an evolutionary derivative of the transcriptional regulator TFII-I (Hinsley et al., 2004; Makeyev et al., 2004; Tipney et al., 2004). However, during divergence from the common ancestor, GTF2IRD2 has undergone some significant rearrangements that preserve some of the features of TFII-I in the N-terminal half, while the C-terminal half has been lost and replaced by the in-frame insertion of a CHARLIE8 transposon. Sequence evidence suggests that subsequent to the insertion, both halves of the

protein have continued to diverge from their original forms. At some point the sequence stabilized, presumably as a result of evolved functional constraints, and the orthologs of GTF2IRD2 are now highly conserved in all of the eutherian mammal lineages for which good sequence is available.

It has been pointed out that the two RDs that remain in GTF2IRD2 compared to the six originally present in TFII-I, are most closely related to RD1 and RD6 of TFII-I (Makeyev et al., 2004). This means that the RD of TFII-I and its adjacent basic region, which is thought to represent the principal site of TFII-I DNA binding (Cheriyath and Roy, 2001; Roy, 2001), is missing in GTF2IRD2. In addition, a PredictNLS survey of the entire sequence indicates no evidence of a recognizable nuclear localization signal (Hinsley et al., 2004). However, the N-terminal leucine zipper and the sequence surrounding it in TFII-I is highly conserved in GTF2IRD2, as is the sequence shared in RD1 of both proteins. This would suggest that the dimerization properties, which have been shown to be mediated via the leucine zipper region and RD1 (also TFII-I RD2) working together to form an N-terminal dimerization domain (Cheriyath and Roy, 2001) have been highly conserved in GTF2IRD2. Other similarities include the presence of well-conserved SUMOylation motifs in both proteins and analysis of mass-purified, SUMO-ligated proteins has shown that TFII-I and GTF2IRD2 are indeed SUMOylation targets (Zhao et al., 2004; Ganesan et al., 2007). However, it is completely unknown what functions are brought by the addition of the CHARLIE8 sequences. Although sequence has diverged from the original transposon from which this region is derived, several recognizable transposase functional domains are identifiable within the sequence (Tipney et al., 2004) and it is likely that these have bestowed novel characteristics on this protein.

### Subcellular localization of GTF2IRD2, GTF2IRD1 and TFII-I

The localization of recombinant GTF2IRD2 was found to differ significantly from the localization of TFII-I $\beta$  and GTF2IRD1 which are broadly distributed throughout the nuclear compartment. GTF2IRD2 was found to be located on microtubules in the cytoplasm and in two types of distinct nuclear foci. The large spherical nuclear foci were shown to be entirely attributable to the CHARLIE8 region as a GFP-tagged CHARLIE8 fusion protein located exclusively to these domains. Analysis of the corresponding GFP-tagged N-terminal half revealed a loose perinuclear cytoplasmic location, which indicates that the N-terminal half lacks a nuclear localization signal (NLS) and it must be concluded that the NLS functions are now provided by the CHARLIE8 domain. Subcellular localization of the hAT transposases is poorly understood except for work showing that basic amino acid regions interspersed with the BED DNA-binding domain are responsible for nuclear localization (Michel and Atkinson, 2003). There is no current evidence that indicates hAT transposases adopt a punctate nuclear pattern or microtubule associations. Together these findings indicate that the unique pattern of GTF2IRD2 localization requires both the CHARLIE8 domain and the TFII-I N-terminal region working in unison and has evolved separately as part of a newly acquired evolutionary function. However, confirmation of these localization patterns will have to await the development of antibodies that are specific and sensitive enough to detect endogenous GTF2IRD2.



Confocal analysis of GTF2IRD2 localization revealed a predominance of peripheral localization, but the position does not coincide with the NPC and is slightly internal to the level of the nuclear lamina when compared to the localization of LAP2B (data not shown). No overlap with any other nuclear foci could be identified and no strong association was found with domains of inactive or active chromatin, although GTF2IRD2 was generally excluded from dense regions of heterochromatin. In contrast, when GTF2IRD1 and TFII- $\beta$  were expressed individually in COS-7 cells, the proteins were distributed throughout the interior of the nucleus except for exclusion from nucleoli. When GTF2IRD1 and TFII- $\beta$  were co-expressed with GTF2IRD2, they adopted the punctate pattern typical of GTF2IRD2. Based on the ability of these proteins to interact directly in co-immunoprecipitation experiments and the high conservation of the N-terminal region, which is involved in dimerization, we propose that the evolved function of GTF2IRD2 is to sequester TFII-I and GTF2IRD1 proteins to regions of the nucleus that prevent their normal functions as transcriptional regulators. It is possible that some sequestration to microtubules also occurs, but no clear evidence of this in co-expression studies could be seen. Under this hypothesis, the ability to interact with the GTF2I family of proteins has been strongly retained in the leucine zipper and RD1 surfaces, whereas the combined effect of the CHARLIE8 and the N-terminal TFII-I region has led to an entirely new pattern of protein localization in the cell. The resulting consequence is an ability to modulate GTF2IRD1 and TFII-I function by protein sequestration.

For this sequestration model to be effective, it requires that the levels of GTF2IRD2 in the nucleus are sufficient to have an impact on the levels of TFII-I or GTF2IRD1 proteins. *Gtf2ird2* transcript levels are, generally speaking, low (Tipney et al., 2004), as are *Gtf2ird1* transcript levels (O'Mahoney et al., 1998; Calvo et al., 2001), whereas *Gtf2i* transcripts are more readily detectable (Roy et al., 1997; Danoff et al., 2004). However, the *Gtf2i* gene generates four different TFII-I protein isoforms that vary in their properties for basal nuclear occupancy or nuclear/cytoplasmic shuttling (Hakre et al., 2006). Antibody detection of total TFII-I protein in the mouse brain suggests that most TFII-I exists in the cytoplasm (Danoff et al., 2004) where it is presumably tethered by p190RhoGAP (Jiang et al., 2005), Bruton's tyrosine kinase (Sacristán et al., 2004) or by interactions with PLC- $\gamma$  (Caraveo et al., 2006). We suggest that GTF2IRD2 sequestration would only be active against TFII-I or GTF2IRD1 proteins that enter the nucleus, which, in the case of the TFII-I isoforms, is largely dependent on the status of cell signaling (Roy, 2012). Therefore the relative stoichiometry of GTF2IRD2 and TFII-I isoforms in the nucleus is unstable and would be difficult to determine.

### Functional analysis of GTF2IRD2 and TFII- $\beta$ in the muscle transgenic model system

Analysis of TFII-I in the muscle transgenic system was conducted at the same time as GTF2IRD2 in order to be able to compare GTF2IRD2 with its closest homolog. The  $\beta$ -isoform was selected because it is the predominant isoform expressed in mouse tissues (Cheriyath and Roy, 2000) and it is constitutively nuclear under basal conditions where it is thought to act as a mild repressor of target genes via recruitment of histone deacetylases (HDACs) to sites of DNA binding (Hakre et al., 2006).

Analysis of muscle fiber type in TFII- $\beta$  transgenic mice supports previous reports of TFII- $\beta$  as a mild repressor of gene

targets and the notion that TFII-I and GTF2IRD1 are capable of regulating the same set of target genes (Tantin et al., 2004) because an almost identical slow to fast post-natal conversion of fiber type was observed in the TFII- $\beta$  mice compared with the original GTF2IRD1 transgenic mice (Issa et al., 2006). However, the magnitude of the shift was smaller despite estimated higher levels of transgenic *Gtf2i mRNA abundance (data not shown).*

In contrast, the GTF2IRD2 transgenic mice showed a consistent and profound post-natal conversion in the opposite direction from fast to slow. This is very apparent in the soleus muscle and to a lesser degree in the surrounding core limb muscles. However, unlike the soleus muscle and unlike the TFII- $\beta$  and GTF2IRD1 models, the surrounding muscles of adult GTF2IRD2 transgenic mice also have much thinner muscle fibers. Since patterning and muscle fiber appearance was normal at post-natal day 2, the thinness of the fibers could be caused by an attenuation of post-natal maturation hypertrophy. There is also a significant increase in the number of muscle fibers, which could represent a compensatory mechanism to combat the significant reduction in fiber size. These data indicate that the effects of GTF2IRD2 differ according to cell context and may reflect underlying molecular differences in epigenetic programming or the profile of transcriptional regulators present.

Based on the analysis of GTF2IRD2 behavior in cultured cells, the simplest mechanism to explain the fiber type conversion is to propose that sequestration of GTF2IRD1 and TFII- $\beta$  proteins leads to inactivation of their usual functions resulting in an antagonistic release from their repressive effects. This model predicts that co-expression of GTF2IRD2 and TFII- $\beta$  or GTF2IRD1 transgenes would set up an antagonistic balance in which a more even fiber type distribution has been restored. By interbreeding of GTF2IRD1 and GTF2IRD2 transgenic mice it was possible to show that this conclusion is correct. However, fiber size reduction in the surrounding muscle fibers was even more pronounced than in GTF2IRD2 transgenics alone, suggesting that the fiber size reduction mechanism does not work in this fashion. This model also predicts that endogenous GTF2IRD1 and/or TFII-I play a role in the maintenance of muscle fiber type as it is proposed that GTF2IRD2 only exerts its effects indirectly via the attenuation of these proteins. *Gtf2ird1* is expressed strongly during skeletal muscle development, but is downregulated in mature muscle (Palmer et al., 2007). Studies on *Gtf2ird1* knockout mice (Palmer et al., 2010; Howard et al., 2012) have so far failed to reveal fiber type changes, which may reflect very low levels of adult expression. By contrast, *Gtf2i* mRNA is relatively abundant in adult skeletal muscle tissue (Roy et al., 1997) and therefore TFII-I is the best candidate for such a role. Homozygous *Gtf2i* mutations cause embryonic lethality (Enkhtandakh et al., 2009) so proper analysis of TFII-I function in muscle tissue will have to await the development of tissue-specific conditional knockout mice.

### Materials and Methods

#### Bioinformatics

Data on conservation of synteny surrounding the GTF2I family of genes across vertebrate species, the relative chromosomal positions and the amino acid sequence of the three mammalian species of GTF2IRD2 was accessed via the Ensembl database (<http://www.ensembl.org/index.html>). Amino acid sequence was aligned using ClustalW (Thompson et al., 1994).

#### Plasmid constructs

Based on sequence information accessed via the Ensembl database, cDNA sequences corresponding to the open reading frame (ORF) of human *GTF2IRD1* ( $1\alpha 1$  isoform), mouse *Gtf2ird2* and mouse *Gtf2i* ( $\beta$ -isoform) were amplified from

human and mouse cell lines by RT-PCR and cloned into the pEGFP-C1 (Clontech) mammalian GFP vector. The pEGFP-GTF2IRD2 and pEGFP-TFII-I $\beta$  vectors were also modified by removal of the EGFP cDNA and re-ligation of a linker containing the sequence encoding the c-Myc epitope tag. C-terminal and N-terminal portions of the mouse *Gtf2ird2* gene were amplified and cloned into pEGFP-C1 using the same strategy.

*Gtf2ird2* and *Gtf2i* ORFs were ligated into the same plasmid construct used to generate the GTF2IRD1 transgenic line (Issa et al., 2006), which ensures skeletal-muscle-specific expression driven by the human skeletal actin gene (*ACTA1*).

#### Cell lines and immunofluorescence

For examination of subcellular localization, NIH3T3, COS-7 and HeLa cells were grown on glass coverslips using standard culture conditions and expression plasmids were transfected using Lipofectamine LTX and PLUS reagent (Invitrogen) according to manufacturer's instructions. Cells were fixed and permeabilized in 4% paraformaldehyde in PBS followed by ice-cold methanol, or by a combined 4% paraformaldehyde/0.125% Triton X-100 in PBS incubation. Disruption of microtubules was achieved by addition of nocodazole to the culture medium (dissolved in DMSO) to a final concentration of 1  $\mu$ M for 2 hours. Recovery following nocodazole treatment involved replacement with fresh culture medium for 1 hour. Antibodies used in immunofluorescence to detect the Myc-tag and endogenous proteins include: anti-Myc (monoclonal 9E10, Sigma), anti- $\alpha$ -tubulin (monoclonal DM1A, Sigma), anti-SC35 (monoclonal Abcam ab11826), anti-coilin (monoclonal Abcam ab11822), anti-PML (rabbit polyclonal Abcam ab53773), anti-SAM68 (rabbit polyclonal Abcam ab26803), anti-BMI1 (rabbit polyclonal Abcam ab38295), anti-nuclear pore complex proteins (Mab414 Abcam ab24609), anti-trimethyl-histone H3K4 (monoclonal Abcam ab12209) and anti-trimethyl-histone H3K9 (polyclonal Abcam ab8898). The fluorescent secondary antibody was Alexa Fluor 555 (Invitrogen). Fixed cells were blocked in 10% normal goat serum (Vector) in PBS and incubated with antibodies and washed according to standard methods. Nuclei were stained with DAPI in PBS (0.1 mg/ml) before mounting in immu-mount (Thermo-Shandon). Cells were imaged with conventional epifluorescence using a Zeiss Axioskop 40 with AxioCam MRC and Axiovision software or with an Olympus Fluoview FV1000 confocal microscope with Fluoview software (V2.0).

#### Immunohistochemistry and histology

Muscle tissue was cryostat sectioned and stained for the myosin heavy chain (MyHC) subtypes I and IIA as described previously (Issa et al., 2006). Detection of transgenic GTF2IRD2 expressed in muscle tissue was achieved using the anti-GTF2IRD2 antibody (Abnova B01P polyclonal) followed by Alexa Fluor 488 (Invitrogen). Fiber diameters and fiber counts were quantified from images of muscle sections using ImageJ software (<http://rsbweb.nih.gov/ij/>). All statistical comparisons were made using a one-tailed Student's *t*-test.

#### Western blotting and co-immunoprecipitation

Myosin heavy chain type I ( $\beta$  slow) was extracted for quantification using the procedure described by Hoh et al. (1976). Proteins were separated by electrophoresis on an 8% SDS-PAGE gel, blotted onto Immobilon-P PVDF membrane (Millipore) and MyHCI was detected using a BA-F8 monoclonal antibody (Borrione et al., 1988), followed by an anti-mouse Ig-HRP-conjugated secondary antibody (Santa Cruz Biotechnology) and detection with Western Lightning ECL reagents (PerkinElmer).

Co-immunoprecipitation of the GTF2I family of proteins was achieved by co-transfection of HEK293 cells with 5  $\mu$ g of each plasmid (except in the case of EGFP-transfected controls where only 100 ng was used) using a standard calcium phosphate precipitation protocol, followed by collection in lysis buffer [phosphate-buffered saline, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 1 mM PMSF and 1 $\times$  Complete Protease Inhibitor Cocktail (Roche)]. Cells were frozen on dry ice, rapidly thawed and forced through a 0.6 mm syringe ten times to lyse the cells.

Immunoprecipitation was performed using Protein-G-Sepharose-4 Fast Flow beads (GE Healthcare) and anti-Myc or anti-GFP (rabbit polyclonal, Abcam ab290) antibodies.

Immunoprecipitated proteins were separated by electrophoresis on a 10% SDS-PAGE gel, blotted onto Immobilon-P PVDF membrane and probed with anti-Myc or anti-GFP antibodies followed by either anti-mouse or anti-rabbit Ig-HRP-conjugated secondary antibody (Santa Cruz Biotechnology) and detection with ECL reagents.

#### Transgenic mice

Transgene constructs were isolated from the vector and microinjected into fertilized oocytes of C57BL/6  $\times$  DBA/2J F1 mice to establish the transgenic lines expressing GTF2IRD2 [B6D2-Tg(*ACTA1-Gtf2ird2*)30Hrd] and TFII-I $\beta$  [B6D2-Tg(*ACTA1-Gtf2i*)29Hrd] on an isogenic background to the GTF2IRD1 line (Issa et al., 2006) [B6D2-Tg(*ACTA1-GTF2IRD1.1x1*)22Hrd] for comparative purposes. Transgene copy number was estimated via Southern blot analysis and transgene

expression was measured by northern blot as described elsewhere (Issa et al., 2006).

#### Acknowledgements

The authors would like to thank Kim Guven for making the transgenic constructs, Josephine Joya for generating the transgenic founder mice, Dr Geraldine O'Neil for assistance with the tubulin experiments, Axel Neumann for the Pml antibody and nuclear microscopy work and Archa Fox for advice on nuclear speckle patterns. The authors also thank Paulina Carmona-Mora and Anthony Kee for critical review of the manuscript.

#### Funding

This work was supported by the Australian Research Council [grant number DP0984430 to S.J.P. and E.C.H.]; the Australian National Health and Medical Research Council [grant number 423401 to S.J.P. and E.C.H.]; and J.W. held a University of New South Wales Postgraduate Award.

#### References

- Antonell, A., Del Campo, M., Magano, L. F., Kaufmann, L., de la Iglesia, J. M., Gallastegui, F., Flores, R., Schweigmann, U., Fauth, C., Kotzot, D. et al. (2010). Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. *J. Med. Genet.* **47**, 312-320.
- 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.
- Borrione, A. C., Zanellato, A. M., Saggini, L., Mazzoli, M., Azzarello, G. and Sartore, S. (1988). Neonatal myosin heavy chains are not expressed in Ni-induced rat rhabdomyosarcoma. *Differentiation* **38**, 49-59.
- Brennan, K. J. and Hardeman, E. C. (1993). Quantitative analysis of the human alpha-skeletal actin gene in transgenic mice. *J. Biol. Chem.* **268**, 719-725.
- Britten, R. J. (2004). Coding sequences of functioning human genes derived entirely from mobile element sequences. *Proc. Natl. Acad. Sci. USA* **101**, 16825-16830.
- 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.
- Caraveo, G., van Rossum, D. B., Patterson, R. L., Snyder, S. H. and Desiderio, S. (2006). Action of TFII-I outside the nucleus as an inhibitor of agonist-induced calcium entry. *Science* **314**, 122-125.
- Castel, D. E., Zhuang, S., Gudi, T., Tang, J., Vuica, M., Desiderio, S. and Pilz, R. B. (2002). cGMP-dependent protein kinase I beta physically and functionally interacts with the transcriptional regulator TFII-I. *J. Biol. Chem.* **277**, 32003-32014.
- Cheriyath, V. and Roy, A. L. (2000). Alternatively spliced isoforms of TFII-I. Complex formation, nuclear translocation, and differential gene regulation. *J. Biol. Chem.* **275**, 26300-26308.
- 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.
- 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.
- Doi-Katayama, Y., Hayashi, F., Inoue, M., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Kigawa, T., Yoshida, M., Shirouzu, M. et al. (2007). Solution structure of the general transcription factor 2I domain in mouse TFII-I protein. *Protein Sci.* **16**, 1788-1792.
- Enkhtandakh, B., Makeyev, A. V., Erdenechimeg, L., Ruddle, F. H., Ching, N. O., Tussie-Luna, M. I., Roy, A. L. and Bayarsaihan, D. (2009). Essential functions of the Williams-Beuren syndrome-associated TFII-I genes in embryonic development. *Proc. Natl. Acad. Sci. USA* **106**, 181-186.
- Ganesan, A. K., Kho, Y., Kim, S. C., Chen, Y., Zhao, Y. and White, M. A. (2007). Broad spectrum identification of SUMO substrates in melanoma cells. *Proteomics* **7**, 2216-2221.
- Hakre, S., Tussie-Luna, M. I., Ashworth, T., Novina, C. D., Settleman, J., Sharp, P. A. and Roy, A. L. (2006). Opposing functions of TFII-I spliced isoforms in growth factor-induced gene expression. *Mol. Cell* **24**, 301-308.
- 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.
- Hoh, J. Y., McGrath, P. A. and White, R. I. (1976). Electrophoretic analysis of multiple forms of myosin in fast-twitch and slow-twitch muscles of the chick. *Biochem. J.* **157**, 87-95.
- Howard, M. L., Palmer, S. J., Taylor, K. M., Arthurson, G. J., Spitzer, M. W., Du, X., Pang, T. Y., Renoir, T., Hardeman, E. C. and Hannan, A. J. (2012). Mutation of *Gtf2ird1* from the Williams-Beuren syndrome critical region results in facial dysplasia, motor dysfunction, and altered vocalisations. *Neurobiol. Dis.* **45**, 913-922.

- Issa, L. L., Palmer, S. J., Guven, K. L., Santucci, N., Hodgson, V. R., Popovic, K., Joya, J. E. and Hardeman, E. C. (2006). MusTRD can regulate postnatal fiber-specific expression. *Dev. Biol.* **293**, 104-115.
- Jiang, W., Sordella, R., Chen, G. C., Hakre, S., Roy, A. L. and Settleman, J. (2005). An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression. *Mol. Cell* **17**, 23-35.
- Kim, D. W., Cheriya, V., Roy, A. L. and Cochran, B. H. (1998). TFII-I enhances activation of the c-fos promoter through interactions with upstream elements. *Mol. Cell Biol.* **18**, 3310-3320.
- Lucena, J., Pezzi, S., Aso, E., Valero, M. C., Carreiro, C., Dubus, P., Sampaio, A., Segura, M., Barthelemy, I., Zindel, M. Y. et al. (2010). Essential role of the N-terminal region of TFII-I in viability and behavior. *BMC Med. Genet.* **11**, doi: 10.1186/1471-2350-11-61.
- Makeyev, A. V., Erdenechimeg, L., Mungunsukh, O., Roth, J. J., Enkhmandakh, B., Ruddle, F. H. and Bayarsaihan, D. (2004). GTF2IRD2 is located in the Williams-Beuren syndrome critical region 7q11.23 and encodes a protein with two TFII-I-like helix-loop-helix repeats. *Proc. Natl. Acad. Sci. USA* **101**, 11052-11057.
- Manzano-Winkler, B., Novina, C. D. and Roy, A. L. (1996). Additions and Corrections to TFII-I is required for transcription of the naturally TATA-less but initiator-containing Vbeta promoter. *J. Biol. Chem.* **271**, 16982.
- Michel, K. and Atkinson, P. W. (2003). Nuclear localization of the Hermes transposase depends on basic amino acid residues at the N-terminus of the protein. *J. Cell. Biochem.* **89**, 778-790.
- Morris, C. A. and Mervis, C. B. (2000). Williams syndrome and related disorders. *Annu. Rev. Genomics Hum. Genet.* **1**, 461-484.
- Novina, C. D., Kumar, S., Bajpai, U., Cheriya, V., Zhang, K., Pillai, S., Wortis, H. H. and Roy, A. L. (1999). Regulation of nuclear localization and transcriptional activity of TFII-I by Bruton's tyrosine kinase. *Mol. Cell Biol.* **19**, 5014-5024.
- 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, MusTRD1. *Mol. Cell Biol.* **18**, 6641-6652.
- Osborne, L. R. and Mervis, C. B. (2007). Rearrangements of the Williams-Beuren syndrome locus: molecular basis and implications for speech and language development. *Expert Rev. Mol. Med.* **9**, 1-16.
- Palmer, S. J., Tay, E. S., Santucci, N., Cuc Bach, T. T., Hook, J., Lemckert, F. A., Jamieson, R. V., Gunning, P. W. and Hardeman, E. C. (2007). Expression of Gtf2ird1, the Williams syndrome-associated gene, during mouse development. *Gene Expr. Patterns* **7**, 396-404.
- Palmer, S. J., Santucci, N., Widagdo, J., Bontempo, S. J., Taylor, K. M., Tay, E. S., Hook, J., Lemckert, F., Gunning, P. W. and Hardeman, E. C. (2010). Negative autoregulation of GTF2IRD1 in Williams-Beuren syndrome via a novel DNA binding mechanism. *J. Biol. Chem.* **285**, 4715-4724.
- Polly, P., Haddadi, L. M., Issa, L. L., Subramaniam, N., Palmer, S. J., Tay, E. S. and Hardeman, E. C. (2003). hMusTRD1alpha represses MEF2 activation of the troponin I slow enhancer. *J. Biol. Chem.* **278**, 36603-36610.
- Ren, X., Siegel, R., Kim, U. and Roeder, R. G. (2011). Direct interactions of OCA-B and TFII-I regulate immunoglobulin heavy-chain gene transcription by facilitating enhancer-promoter communication. *Mol. Cell* **42**, 342-355.
- 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.
- Roy, A. L. (2001). Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene* **274**, 1-13.
- Roy, A. L. (2006). Transcription factor TFII-I conducts a cytoplasmic orchestra. *ACS Chem. Biol.* **1**, 619-622.
- Roy, A. L. (2007). Signal-induced functions of the transcription factor TFII-I. *Biochim. Biophys. Acta* **1769**, 613-621.
- Roy, A. L. (2012). Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later. *Gene* **492**, 32-41.
- Roy, A. L., Carruthers, C., Gutjahr, T. and Roeder, R. G. (1993). Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature* **365**, 359-361.
- Roy, A. L., Du, H., Gregor, P. D., Novina, C. D., Martinez, E. and Roeder, R. G. (1997). Cloning of an in- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. *EMBO J.* **16**, 7091-7104.
- Rubin, E., Lithwick, G. and Levy, A. A. (2001). Structure and evolution of the hAT transposon superfamily. *Genetics* **158**, 949-957.
- Sacristán, C., Tussie-Luna, M. I., Logan, S. M. and Roy, A. L. (2004). Mechanism of Bruton's tyrosine kinase-mediated recruitment and regulation of TFII-I. *J. Biol. Chem.* **279**, 7147-7158.
- Sakurai, T., Dorr, N. P., Takahashi, N., McInnes, L. A., Elder, G. A. and Buxbaum, J. D. (2011). Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. *Autism Res.* **4**, 28-39.
- 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.
- Tapia-Páez, I., Tammimies, K., Massinen, S., Roy, A. L. and Kere, J. (2008). The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. *FASEB J.* **22**, 3001-3009.
- Tassabehji, M., Hammond, P., Karmiloff-Smith, A., Thompson, P., Thorgerisson, S. S., Durkin, M. E., Popescu, N. C., Hutton, T., Metcalfe, K., Rucka, A. et al. (2005). GTF2IRD1 in craniofacial development of humans and mice. *Science* **310**, 1184-1187.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
- 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.
- 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.
- 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.
- Young, E. J., Lipina, T., Tam, E., Mandel, A., Clapcote, S. J., Bechard, A. R., Chambers, J., Mount, H. T., Fletcher, P. J., Roder, J. C. et al. (2008). Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice. *Genes Brain Behav.* **7**, 224-234.
- Zhao, Y., Kwon, S. W., Anselmo, A., Kaur, K. and White, M. A. (2004). Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J. Biol. Chem.* **279**, 20999-21002.