Hum Genet
DOI 10.1007/s00439-015-1591-0

ORIGINAL INVESTIGATION

The nuclear localization pattern and interaction partners of GTF2IRD1 demonstrate a role in chromatin regulation

Paulina Carmona-Mora[1] · Jocelyn Widagdo[2] · Florence Tomasetig[1] · Cesar P. Canales[1] · Yeojoon Cha[1] · Wei Lee[1] · Abdullah Alshawaf[3] · Mirella Dottori[3] · Renee M. Whan[4] · Edna C. Hardeman[1] · Stephen J. Palmer[1]

Received: 11 February 2015 / Accepted: 4 August 2015

Abstract *GTF2IRD1* is one of the three members of the *GTF2I* gene family, clustered on chromosome 7 within a 1.8 Mb region that is prone to duplications and deletions in humans. Hemizygous deletions cause Williams–Beuren syndrome (WBS) and duplications cause WBS duplication syndrome. These copy number variations disturb a variety of developmental systems and neurological functions. Human mapping data and analyses of knockout mice show that *GTF2IRD1* and *GTF2I* underpin the craniofacial abnormalities, mental retardation, visuospatial deficits and hypersociability of WBS. However, the cellular role of the GTF2IRD1 protein is poorly understood due to its very low abundance and a paucity of reagents. Here, for the first time, we show that endogenous GTF2IRD1 has a punctate pattern in the nuclei of cultured human cell lines and neurons. To probe the functional relationships of GTF2IRD1 in an unbiased manner, yeast two-hybrid libraries were screened, isolating 38 novel interaction partners, which were validated in mammalian cell lines. These relationships illustrate GTF2IRD1 function, as the isolated partners are mostly involved in chromatin modification and transcriptional regulation, whilst others indicate an unexpected role in connection with the primary cilium. Mapping of the sites of protein interaction also indicates key features regarding the evolution of the GTF2IRD1 protein. These data provide a visual and molecular basis for GTF2IRD1 nuclear function that will lead to an understanding of its role in brain, behaviour and human disease.

Abbreviations

hESC	Human embryonic stem cells
PLA	Proximity ligation assay
STED	Stimulated emission depletion
WBS	Williams–Beuren syndrome
Y2H	Yeast two-hybrid

Introduction

GTF2IRD1 (GTF2I repeat domain containing protein I) was initially identified in three independent yeast one-hybrid screens as a protein that interacted with DNA baits composed of triplicated versions of the upstream regions of *TNNI1* (O'Mahoney et al. 1998), *Hoxc8* (Bayarsaihan and Ruddle 2000), and *goosecoid* (Ring et al. 2002). In the human, the gene encoding GTF2IRD1 is located in a cluster within chromosome 7q11.23 with two closely related genes; *GTF2I* encoding TFII-I and *GTF2IRD2* encoding GTF2IRD2. The 7q11.23 region contains three blocks of low-copy repeats (LCRs) that cause susceptibility to non-allelic homologous recombination during meiosis that results in offspring carrying hemizygous deletions resulting in Williams–Beuren syndrome (WBS OMIM#194050) (Franke et al. 1999; Osborne et al. 1999; Pérez Jurado et al. 1998; Tassabehji et al. 1999; Tipney et al. 2004)

Electronic supplementary material The online version of this article (doi:10.1007/s00439-015-1591-0) contains supplementary material, which is available to authorized users.

✉ Stephen J. Palmer
s.palmer@unsw.edu.au

[1] Cellular and Genetic Medicine Unit, School of Medical Sciences, UNSW Australia, Sydney, NSW 2052, Australia

[2] Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia

[3] Centre for Neural Engineering, The University of Melbourne, Melbourne, VIC 3010, Australia

[4] Biomedical Imaging Facility, Mark Wainwright Analytical Centre, UNSW Australia, Sydney, NSW 2052, Australia

Published online: 15 August 2015

or duplications that cause WBS duplication syndrome (OMIM#609757) (Depienne et al. 2007; Merla et al. 2010; Sanders et al. 2011; Somerville et al. 2005; Torniero et al. 2007; Van der Aa et al. 2009).

Patients with WBS duplication syndrome have only recently been described and little mapping data exists that can discriminate the individual genetic contribution to the phenotypes but in WBS, a series of atypical deletion patients indicate that loss of *GTF2IRD1* and/or *GTF2I* is necessary to manifest the craniofacial abnormalities, mental retardation, visuospatial construction deficits and hypersociability of WBS (Antonell et al. 2010). Given the profound importance of copy number variations (CNVs) of these genes to the neurological abnormalities of WBS, it is not unreasonable to suppose that CNVs which increase *GTF2IRD1* and *GTF2I* gene dosage may also play an important role in the consistent speech delay and increased rates of autism and schizophrenia found in WBS duplication syndrome and evidence from mouse models supports this idea (Osborne 2010). Therefore, it is very important to understand the function of these related proteins to comprehend the consequences of their altered dosage.

Cultured cell transfection studies and transgenic experiments both indicate that GTF2IRD1 has strong gene repression capabilities (Issa et al. 2006; Tay et al. 2003) and binding studies demonstrated that GTF2IRD1 has sequence-specific DNA recognition properties for GGATTA-containing sequences that are conferred by a subset of the five I-repeat domains (RDs) that it contains (Polly et al. 2003; Thompson et al. 2007; Vullhorst and Buonanno 2003, 2005). The *GTF2IRD1* upstream region (GUR) contains three GGATTA binding sites and EMSA studies have shown that all three are required to achieve high-affinity GTF2IRD1 binding (Palmer et al. 2010). This may explain why GTF2IRD1 was readily isolated from the artificial triplicated bait constructs of the original yeast one-hybrid assays. However, it is unclear what evolutionary advantage was bestowed by the multiple duplication of this DNA binding domain and how the RDs work in DNA binding site selection of target genes.

Apart from the RDs, the human GTF2IRD1 protein contains a short leucine zipper near the N-terminus implicated in dimerization (Vullhorst and Buonanno 2003), a nuclear localization signal (NLS) near the C-terminus, two SUMOylation motifs of which one is clearly highly conserved and functional (Widagdo et al. 2012), a highly conserved C-terminal domain that may be important for SUMOylation due to the binding of the E3 SUMO-ligase, PIASxβ (Widagdo et al. 2012) and a polyserine tract near the C-terminus that is missing in all fish species but present in amphibians and may, therefore, be a more recent evolutionary refinement.

Much of the thinking regarding GTF2IRD1 function is based on homology with TFII-I, which has been studied more intensively. It is undisputed that *GTF2I* and *GTF2IRD1* evolved from a common ancestor but the level of functional overlap and redundancy between TFII-I and GTF2IRD1 is currently unclear. Direct protein interaction between them is possible (Palmer et al. 2012) and some of the data indicates similar target gene sets and mechanisms of regulation (Jackson et al. 2005; Palmer et al. 2012; Tantin et al. 2004). However, unlike GTF2IRD1, much of the data on TFII-I indicate a very broad role in both the cytoplasm and the nucleus and individual isoforms show different subcellular localization patterns and functional properties (Roy 2012). Some isoforms of TFII-I are thought to reside in the cytoplasm, where they are tethered by interactions with Bruton's tyrosine kinase (Yang and Desiderio 1997) or p190RhoGAP (Jiang et al. 2005) and shuttle into the nucleus in response to signalling events and can also interact with PLC-γ in a way that competitively inhibits its binding to TRPC3, thus altering agonist-induced calcium entry into the cell (Caraveo et al. 2006). At the same time, TFII-I isoforms have a series of nuclear roles that include direct DNA binding to the regulatory regions of various genes, including *c-fos* (Roy 2012).

The majority of studies indicate that *Gtf2ird1*$^{-/-}$ null mice survive but have craniofacial and neurological abnormalities (Howard et al. 2012; Schneider et al. 2012; Tassabehji et al. 2005; Young et al. 2008), whereas *Gtf2i*$^{-/-}$ mice are embryonic lethal (Enkhmandakh et al. 2009). These data, in combination with the data on the cytoplasmic roles of TFII-I, might be taken to suggest that TFII-I plays a broader and more critical cellular role than GTF2IRD1. However, evolutionary conservation studies indicate that the common ancestor of these two genes bore a stronger sequence similarity to the current *GTF2IRD1* (Gunbin and Ruvinsky 2013). Thus, during the initial period following duplication, proto-*GTF2I* was presumably liberated from functional constraints, whereas the proto-*GTF2IRD1* retained most of the ancestral gene's role. The *GTF2I/GTF2IRD1* duplication pre-dates the formation of cartilaginous fish but the *GTF2I* gene has been lost in the two bony fish infraclasses (Teleostei and Holostei), whereas *GTF2IRD1* has been retained in all vertebrates since its formation (Gunbin and Ruvinsky 2013). These data support the likelihood of functional overlaps between these two proteins and suggest that, at least in bony fish, GTF2IRD1 is sufficient to support all of the functions they provide in other species.

Based on the well-established DNA binding properties of GTF2IRD1 and the clear impact on transcriptional regulation when over-expressed in vitro (Polly et al. 2003; Vullhorst and Buonanno 2003) and in vivo (Issa et al. 2006), it has been assumed that GTF2IRD1 is a conventional

transcription factor that has a consistent set of gene targets that will be dysregulated in its absence. But, despite demonstrated alterations in behaviour and motor function in *Gtf2ird1* knockout mice (Howard et al. 2012; Young et al. 2008) and electrophysiological changes in CNS neurons (Proulx et al. 2010), evidence for such a gene set from transcriptional analysis of knockout brain tissue has so far proved elusive (O'Leary and Osborne 2011). This has led to speculation on the possibility of missing roles for GTF2IRD1, such as a cytoplasmic function similar to TFII-I (O'Leary and Osborne 2011).

Therefore, there is a strong need to supply some fundamental information regarding GTF2IRD1, which has been hampered by the lack of good quality antibodies and a lack of understanding about its protein–protein interactions. In this paper, we examine the subcellular localization of endogenous GTF2IRD1 and demonstrate that it is distributed in a speckled pattern in the nucleus that brings it into close proximity with several markers of chromatin/transcriptional regulation. To identify the range of protein interactions engaged in by GTF2IRD1, we used yeast two-hybrid library screenings to generate an unbiased comprehensive list of protein partners. Most of the proteins isolated support a role in the regulation of chromatin. Interactions with other DNA binding proteins and transcriptional co-factors suggest that GTF2IRD1 binds to chromatin targets using cooperative mechanisms. In addition, interactions with several components of the primary cilium and ARM repeat proteins offer an intriguing new direction in GTF2IRD1 research.

Results

Endogenous GTF2IRD1 exists in a punctate pattern in the nucleus

Previous analyses of endogenous GTF2IRD1 have been hampered by a lack of good quality antibodies and the low abundance of the protein. However, two commercial antibodies were identified, the first of which (333A) is usable in immunofluorescence and immunoprecipitation but is specific for the human protein and is a very poor detection reagent on western blots. The second (M19) is highly sensitive as a detection reagent on western blots but has relatively poor specificity and does not work for immunofluorescence (Fig. 1).

In whole cell extracts of HeLa cells, M19 detects two major bands in the 110–130 kDa range (Fig. 1a). After immunoprecipitation using the 333A antibody, only the upper band (running at approximately 130 kDa) is detected by M19, indicating that this corresponds to endogenous GTF2IRD1 (predicted molecular weight: 106 kDa). This result was confirmed using a pool of 4 anti-GTF2IRD1 siRNAs. While no change in the band pattern from whole cell extracts was observed when HeLa cells were treated with control siRNA, the upper band at 130 kDa was lost when cells were treated with anti-GTF2IRD1 siRNA (Fig. 1a).

The lower molecular weight band detected by M19 could not be identified. We considered the possibility that it constituted TFII-I or GTF2IRD2 and was detected as a result of cross-reactivity with the M19 antibody. However, this possibility was dismissed by immunoblotting using anti-TFII-I and anti-GTF2IRD2 antibodies. These antibodies both identified bands that run at a very similar molecular weight to GTF2IRD1 and not at the lower level of approximately 110 kDa. The possibility that the anti-GTF2IRD1 siRNAs affected the levels of TFII-I and GTF2IRD2 or that M19 was cross-reacting with these related proteins was also excluded by probing extracts treated using the siRNA knockdown oligonucleotides (Fig. 1a).

Endogenous GTF2IRD1 has never been convincingly localized within the cell by immunofluorescence. Using the 333A antibody we were able to detect a punctate signal in the nuclei of HeLa cells and this signal was completely abrogated in the majority of cells treated with the anti-GTF2IRD1 siRNAs (Fig. 1b). Cell counting indicated that approximately 98 % of cells had no signal while the remaining 2 % showed the normal pattern, indicating that they had failed to be transfected by the siRNA oligonucleotides. Use of super-resolution confocal STED (stimulated emission depletion) microscopy demonstrated that the speckles were distributed evenly throughout the nuclei of HeLa, HEK-293 (Fig. 1c) and SH-SY5Y cells (data not shown) in large numbers.

We considered the possibility that the pattern of GTF2IRD1 localization was specific to immortalized cell lines. Since GTF2IRD1 function has been associated with neurobehavioural abnormalities in mouse studies (Howard et al. 2012; Young et al. 2008) and we had previously shown expression in the mouse brain (Palmer et al. 2007), an understanding of localization in neurons was sought. Human cerebellum samples were obtained for immunofluorescence analysis but no convincing evidence of GTF2IRD1 localization could be found in frozen or fixed tissue (data not shown). Therefore, neurons differentiated from human embryonic stem (ES) cells were analysed using the 333A antibody. Co-immunofluorescence with anti-β-tubulin III and anti-MAP2ab antibodies showed punctate nuclear expression of GTF2IRD1 in neuronal cells (Fig. 2), consistent with its expression pattern in the immortalized cell lines (Fig. 1). Of note, GTF2IRD1 nuclear expression was also observed in β-tubulin III negative and MAP2ab negative cells, which correspond to subpopulations of early neural progenitors (Dottori, personal communication).

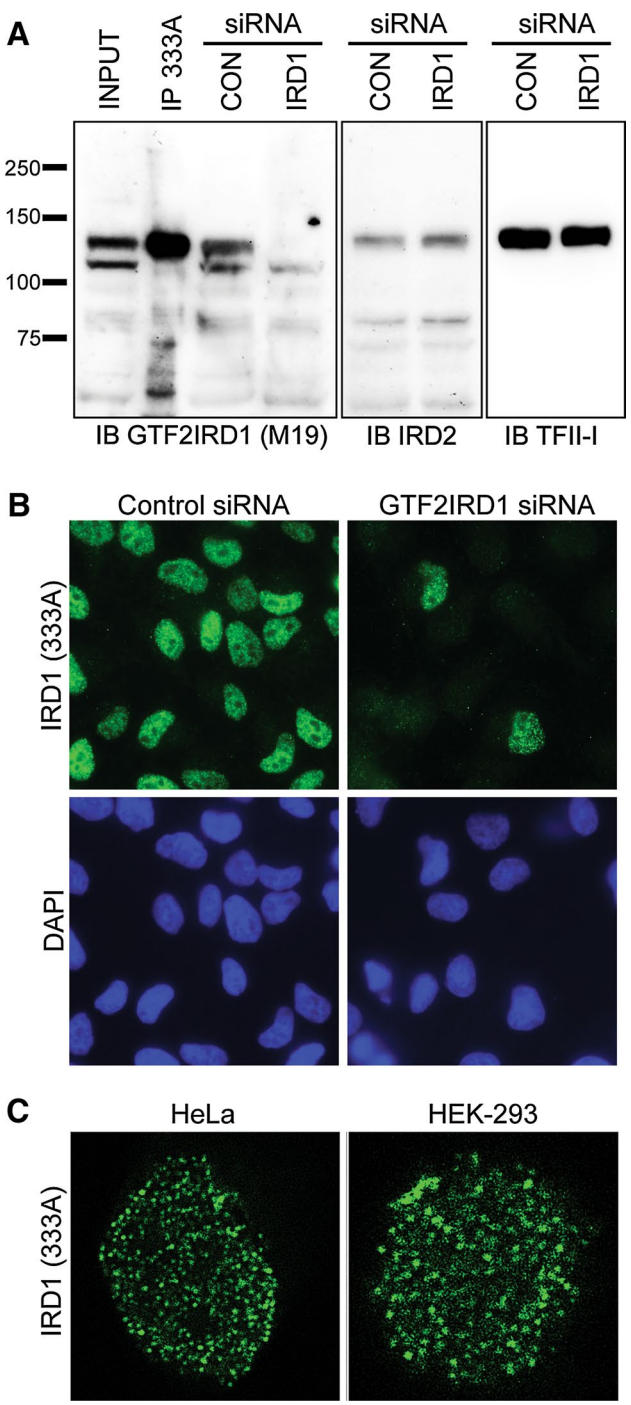


Fig. 1 Detection of endogenous human GTF2IRD1. **a** Western blot analysis of endogenous GTF2IRD1, GTF2IRD2 and TFII-I in HeLa cell extracts. A single western blot was cut into strips (indicated by the *boxes*), probed using anti-GTF2IRD1 (M19), anti-GTF2IRD2 (IRD2) and anti-TFII-I antibodies, and the resulting film exposures realigned. The anti-GTF2IRD1 M19 antibody detects two bands above 100 kDa in the whole cell extract (No siRNA) but after immunoprecipitation using the anti-GTF2IRD1 333A antibody (IP 333A), only the upper band is detected. In extracts from cells transfected with a negative control siRNA (CON siRNA), both bands are detected but in cells transfected with the anti-GTF2IRD1 siRNA pool (IRD1 siRNA), the upper band disappears. Immunoblotting (IB) for GTF2IRD2 and TFII-I shows that the lower band does not correspond to these proteins and the anti-GTF2IRD1 siRNA lane is unaffected in both blots, showing that there is no evidence of compensatory protein level change or antibody cross-reactivity. **b** Immunofluorescence analysis of endogenous GTF2IRD1 protein using the 333A antibody on HeLa cells treated with control or targeting siRNA. **c** Immunofluorescence analysis of endogenous GTF2IRD1 distribution in the nucleus of HeLa and HEK-293 cells using stimulated emission depletion (STED) super resolution confocal microscopy

Endogenous GTF2IRD1 is found in close proximity to elements of chromatin regulation

These results prompted an exploration of co-localization with other nuclear speckling bodies as a potential means to understand GTF2IRD1 function. Comparison of the distribution patterns using co-immunofluorescence revealed that none of the markers tested showed consistent one-to-one overlap with GTF2IRD1. However, qualitative assessment indicated that some GTF2IRD1 protein showed fractional overlap with markers of chromatin and transcriptional regulation, including histone H3 methylation marks, members of the heterochromatin protein 1 family and SP1 (Fig. 3). Co-localization of GTF2IRD1 and the nuclear bodies detected by antibodies against, coilin, LAP2, the nuclear pore complex (NPC) and PML was very limited, although the sparse PML bodies typically showed at least one overlapping signal per nucleus (Supplementary Material, Fig. S1).

Proximity ligation assays (PLA) were used as a quantitative means to assess the incidence of close proximity between endogenous GTF2IRD1 protein and the markers of chromatin/transcriptional regulation in HeLa cells (Fig. 3). These analyses indicated that GTF2IRD1 is found in close proximity with the heterochromatin proteins HP1β and HP1γ at the highest frequency, followed by SP1, HP1α, H3K27Me2/3 and H3K4Me3 in descending order (Fig. 4). PLA signal detected for GTF2IRD1 and H3K9Me3 was equivalent to the combined background control levels, suggesting that the incidence of close proximity between these proteins is equal to or approaching zero.

These observations suggested an association with elements of the chromatin regulation machinery, but stronger evidence of such interactions prevented firmer functional conclusions. While some GTF2IRD1 protein binding partners have been reported (Tussie-Luna et al. 2002; Widagdo et al. 2012) this area is unexplored and we therefore set out to address this need using a comprehensive screening approach.

Yeast two-hybrid library screening for novel GTF2IRD1 interacting partners

Yeast two-hybrid (Y2H) screening was chosen because it is unbiased and can identify both transient and stable

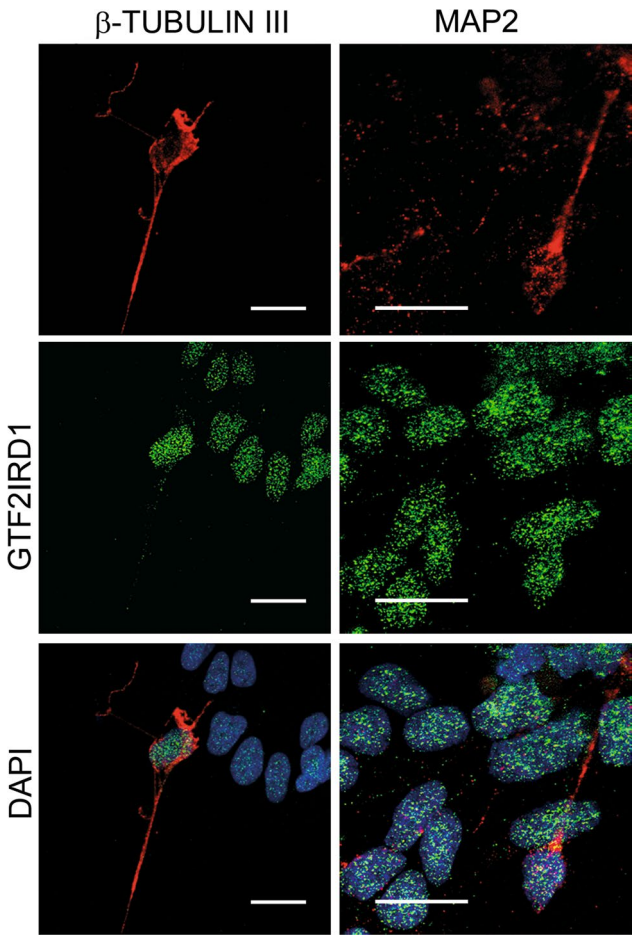


Fig. 2 Endogenous GTF2IRD1 adopts a speckled nuclear pattern in hESC-derived neuronal cell cultures. Immunofluorescence analysis of hESC-derived cells, driven into the neuronal pathway of development, shows that GTF2IRD1 has the same nuclear pattern found in immortalized cell lines. GTF2IRD1 is expressed in all cells including differentiating neurons, as marked by β-tubulin III and MAP2ab antibodies. *Scale bars* represent 20 μm

interactions. To provide as complete a list of binding partners as possible, two Y2H screens were performed using a universal normalized mouse cDNA library (derived from a collection of different mouse tissues) and a human brain normalized cDNA library. These screens led to the initial isolation of 191 positive yeast colonies.

Clones with prey sequences duplicating other clones or out of frame with the GAL4 DNA binding domain were discarded. Most of the remaining clones were retransformed into haploid AH109 yeast using the original rescued prey plasmid or a reconstructed prey plasmid containing the full-length prey cDNA (Supplementary Material, Table 1), together with the bait construct or the empty pGBKT7 control plasmid (Supplementary Material, Fig. S2). Prey clones that were resistant to the quadruple dropout (QDO) media in the presence of the empty control plasmid were presumed to encode proteins that bind directly to the GAL4 DNA binding domain and were also discarded.

This refinement process led to the identification of 40 individual GTF2IRD1 binding protein candidates (Table 1). Five of these clones were not pursued beyond sequence identification as they were known to be solely cytoplasmic, extracellular or cell membrane localized and were less likely to be of biological relevance. Two of the proteins have been described previously as interacting nuclear partners (Tussie-Luna et al. 2002; Widagdo et al. 2012). Of the remaining 33 proteins, 26 either shuttle into the nucleus or are primarily located in the nucleus according to known functions or predictions summarized in the subcellular localization database, COMPARTMENTS (Binder et al. 2014). The KPNA proteins were predicted to have been isolated due to their binding to the GTF2IRD1 nuclear localization signal (NLS). Preliminary Y2H studies mapped this interaction to the C-terminal domain of GTF2IRD1, where the NLS is located, suggesting that this was indeed the case and these proteins were not pursued beyond this point (data not shown). It was striking that six of the remaining non-nuclear proteins are associated with or have links with centrosome and primary cilia function (Table 1), an association that has never been previously noted in connection with GTF2IRD1.

Domain characterization for the novel protein interactions of GTF2IRD1

To map the binding domains of these proteins in GTF2IRD1, a range of Y2H bait plasmids was constructed containing 8 separate 88 amino acid regions of the GTF2IRD1 protein (Fig. 5a) containing known functional units or sequences that are strongly conserved between species, as described previously (Widagdo et al. 2012). A selected set of prey proteins was co-transformed with each of the 8 domain-specific plasmids and plated on QDO/x-α-GAL media (Fig. 5b). Protein interactions were mapped to several of the domains, sometimes multiple domains. The highest number of interactions mapped to the regions containing the SUMOylation motifs (Fig. 5c).

Subcellular localization of GTF2IRD1 and its novel protein partners

To gather further evidence for interactions in a mammalian cell context and to check the subcellular localization characteristics of the putative protein partners, plasmids encoding epitope-tagged versions or EGFP fusion proteins were either obtained or constructed (see Supplementary Material, Table S1). These plasmids were co-transfected into HeLa cells with plasmids encoding either GTF2IRD1-EGFP or Myc-tagged GTF2IRD1 and co-localization was analysed by fluorescence microscopy (Fig. 6).

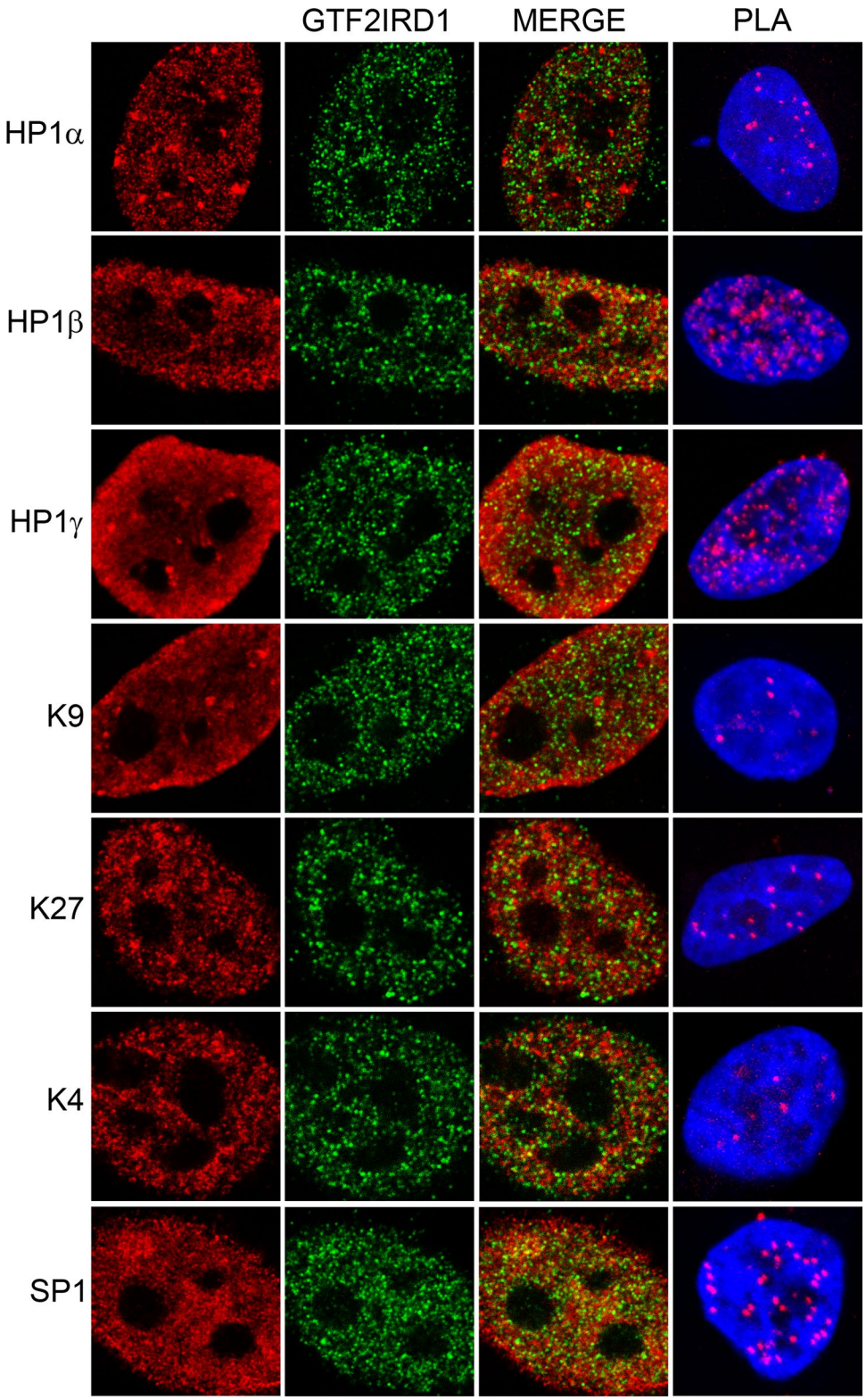
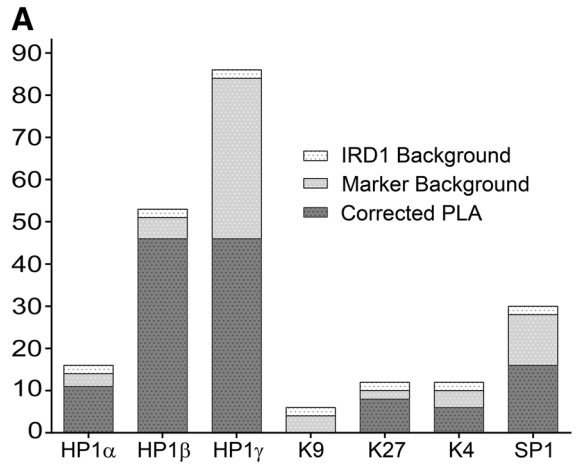


Fig. 3 Co-localization of endogenous GTF2IRD1 with markers of chromatin/transcriptional regulation using confocal immunofluorescence analysis and PLA. The markers include HP1 α, β, γ, H3K9Me3 (K9), H3K27Me2/3 (K27), H3K4Me3 (K4) and SP1. The PLA images show representative Z-stack confocal reconstructions of DAPI-stained nuclei overlaid with the PLA dots generated by the same antibody pairings as the adjacent immunofluorescence images

The majority of the candidate proteins localized to the nucleus, as expected, with some degree of nuclear speckling in many cases but most showed little direct overlap with the tagged GTF2IRD1 protein (Fig. 6a). Candidate proteins that localized predominantly outside of the nucleus in these assays (Fig. 6b) may shuttle into the nucleus under normal circumstances to interact with GTF2IRD1 and thus, these findings should not prejudice the likelihood of



B

	Combined		Background controls				
	Marker/GTF2IRD1		GTF2IRD1		Marker		Corrected
PLA	Dots	SD	Dots	SD	Dots	SD	Mean
HP1α + GTF2IRD1	16	10	2	2	3	2	**11**
HP1β + GTF2IRD1	54	22	2	2	5	5	**47**
HP1γ + GTF2IRD1	86	26	2	2	38	10	**46**
H3K9 + GTF2IRD1	7	7	2	2	4	2	**1**
H3K27 + GTF2IRD1	12	9	2	2	2	2	**8**
H3K4 + GTF2IRD1	12	6	2	2	4	3	**6**
SP1+ GTF2IRD1	30	15	2	2	12	9	**16**

Fig. 4 PLA quantification of the incidence of mean close proximity per nucleus between GTF2IRD1 and markers of chromatin/transcriptional regulation. The markers include HP1 α, β, γ, H3K9Me3 (K9), H3K27Me2/3 (K27), H3K4Me3 (K4) and SP1. **a** Histogram representing the total mean PLA dots per nucleus, the contribution of the total caused by GTF2IRD1 (IRD1) background signal, the marker background signal and the resulting estimate of the corrected PLA mean. **b** Table of the same data shown in A (rounded to integers) with associated estimates of standard deviation (SD) for each mean

a genuine biological interaction. However, these proteins were not analysed in the next stage of verification, involving co-immunoprecipitation (co-IP), since they were not occupying the same cellular compartment as GTF2IRD1.

GTF2IRD1 interacts with chromatin modifiers and transcriptional regulators in mammalian cells

The majority of the candidate proteins that showed significant distribution in the nuclear compartment were selected for a further level of validation using co-IP analysis of the recombinant proteins. Plasmids encoding tagged fusion proteins of GTF2IRD1 and each candidate protein were co-transfected into HeLa cells and protein complexes were immunoprecipitated using the anti-GFP antibody. Co-IP proteins were analysed on western blots using the anti-Myc antibody. Negative controls were performed by co-transfection with the empty pEGFP vector.

PKP2 is a member of the plakophilin family that plays dual roles in the nucleus and in desmosomes (Chen et al. 2002). In HeLa cells, the protein predominantly localized to the cell surface and, therefore, interaction with GTF2IRD1 was not tested by co-IP. However, the family member PKP1 is known to localize more readily to the nucleus (Hatzfeld et al. 2000) and we considered the possibility that the interaction of GTF2IRD1 with PKP family members may be conserved. To address this question, the prey vector pGADT7 containing the PKP1 open reading frame was co-transformed with GTF2IRD1 and an interaction in yeast was verified (Supplementary Material, Fig. S2). Second, the localization of PKP1-EGFP to the nuclei of HeLa cells was confirmed (Fig. 6b) and PKP1 was included in the co-IP experiments.

All of the candidate proteins tested were found to co-immunoprecipitate with GTF2IRD1 from the HeLa cell extracts with varying levels of recovery and all the control interactions with GFP were negative as expected (Fig. 7). The majority of experiments were performed using EGFP-tagged versions of the candidate proteins and GTF2IRD1-Myc (Fig. 7a) but SETD6 and ZMYM2 co-IPs were performed using the reverse configuration (Fig. 7b) because only Myc-tagged versions of these constructs were available.

Co-immunofluorescence analysis of endogenous ZMYMs in mammalian cells

Of the previously unreported candidate proteins isolated in these screens, the proteins ZMYM2 and ZMYM3 were the most prominent to us because they have been previously isolated using immunoaffinity purification from endogenous HeLa cell extracts as part of a complex that contained TFII-I, BHC110, BHC80, CoREST, HDAC1 and HDAC2 (Hakimi et al. 2003). Therefore, while direct binding of TFII-I to ZMYM2 and ZMYM3 had not been demonstrated, it seemed plausible that direct interactions between ZMYM proteins and members of the TFII-I family, including GTF2IRD1, are a conserved feature that confers the ability to integrate into HDAC-containing silencing complexes.

To examine this hypothesis, anti-ZMYM2 and anti-ZMYM3 antibodies were obtained and co-immunofluorescence analysis of their endogenous co-localization with endogenous GTF2IRD1 and TFII-I was conducted in HeLa cells. All 4 of these proteins were distributed in punctate patterns, and some co-localization was apparent, but the overlap was not one-to-one as a proportion of the red and green signal was still very obvious (Fig. 8a). However, it was noteworthy that the co-localization of GTF2IRD1 and the ZMYMs was not dissimilar from the co-localization of TFII-I and the ZMYMs; previously associated by co-immunoprecipitation in HeLa cells (Hakimi et al. 2003).

Table 1 Summary of Y2H screens

Gene symbol	Name	Location	Validation
AKIRIN2	Akirin 2	Nuclear	Yeast
ALMS1	Alstrom syndrome 1	Primary cilia/centrosome	Yeast
ARMCX5	Armadillo repeat containing, X-linked 5	Nuclear/cytoplasmic	Yeast
ATF7IP	Activating transcription factor 7 interacting protein	Nuclear	co-IP
ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	Cytoplasmic	n.d.
BBS4	Bardet–Biedl syndrome 4	Primary cilia/centrosome	Yeast
CSPP1	Centrosome and spindle pole associated protein 1	Cilia/centrosome	Yeast
DCAF6[a]	DDB1 and CUL4 associated factor 6	Nucleus	co-IP
ELF2	E74-like factor 2 (ets domain transcription factor)	Nucleus	Yeast
FAM47E	Family with sequence similarity 47, member E	Nucleus	Yeast
FBXW10	F-box and WD repeat domain containing 10	Nuclear/cytoplasmic	Yeast
FHAD1	Forkhead-associated (FHA) phosphopeptide binding domain 1	Unknown	Yeast
HOMEZ	Homeobox and leucine zipper encoding	Nucleus	co-IP
HSF2BP	Heat shock transcription factor 2 binding protein	Cytoplasm	Yeast
HTRA4	HtrA serine peptidase 4	Extracellular	n.d.
INTS12	Integrator complex subunit 12	Nucleus	co-IP
KPNA1	Karyopherin alpha 1 (importin alpha 5)	Nucleus	Yeast
KPNA2[a]	Karyopherin alpha 2 (RAG cohort 1, importin alpha 1)	Nucleus	Yeast
KPNA3	Karyopherin alpha 3 (importin alpha 4)	Nucleus	Yeast
KPNA4	Karyopherin alpha 4 (importin alpha 3)	Nucleus	Yeast
MBD3L1	Methyl-CpG-binding domain protein 3-like 1	Nucleus	co-IP
NAP1L2	Nucleosome assembly protein 1-like 2	Nucleus	co-IP
OPHN1	Oligophrenin 1	Cytoplasm	Yeast
PARPBP	PARP1 binding protein	Nucleus	Yeast
PIAS1[a]	Protein inhibitor of activated STAT-1	Nucleus	Yeast
PIAS2[b]	Protein inhibitor of activated STAT, 2	Nucleus	Reported
PKP2	Plakophilin 2	Desmosome/nucleus	Yeast
SCNM1	Sodium channel modifier 1	Nucleus	Yeast
SETD6	SET domain containing 6	Nucleus	co-IP
SPTLC1	Serine palmitoyltransferase, long chain base subunit 1	Endoplasmic reticulum	n.d.
TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kD	Nucleus	Yeast
TMEM55A	Transmembrane protein 55A	Membrane	n.d.
TRIP11	Thyroid hormone receptor interacting protein 11	Golgi/primary cilia	Yeast
USP20	Ubiquitin specific peptidase 20	Cytoplasm/centrosome	Yeast
USP33	Ubiquitin specific peptidase 33	Cytoplasm/centrosome	Yeast
VIMP	VCP-interacting membrane protein	Endoplasmic reticulum	n.d.
ZC4H2[a]	Zinc finger, C4H2 domain containing	Nucleus	co-IP
ZMYM2	Zinc finger, MYM-type 2	Nucleus	co-IP
ZMYM3	Zinc finger, MYM-type 3	Nucleus	co-IP
ZMYM5[a,b]	Zinc finger, MYM-type 5	Nucleus	Reported

Combined results of GTF2IRD1 interacting partners identified in two independent screens (mouse universal and human brain): gene symbols are sorted alphabetically. Location data are based on reported subcellular localizations or predicted/inferred information using COMPARTMENTS (Binder et al. 2014). Several clones occurred in both screens (a) and 2 genes (b) have been described previously (Tussie-Luna et al. 2002; Widagdo et al. 2012). Some interactions were not pursued beyond sequence analysis (not done—n.d.). Other clones were either validated solely by retransformation in yeast (yeast), or by yeast retransformation and subsequent transfection into HeLa cell lines and co-immunoprecipitation (co-IP)

To explore these associations quantitatively, the frequency of close proximity per nucleus between endogenous GTF2IRD1 or TFII-I and ZMYM2 and ZMYM3 was estimated using PLA (Fig. 8b). These data indicated that the incidence of close proximity per nucleus was similar and at relatively high levels for both GTF2IRD1 and TFII-I with the ZMYM proteins.

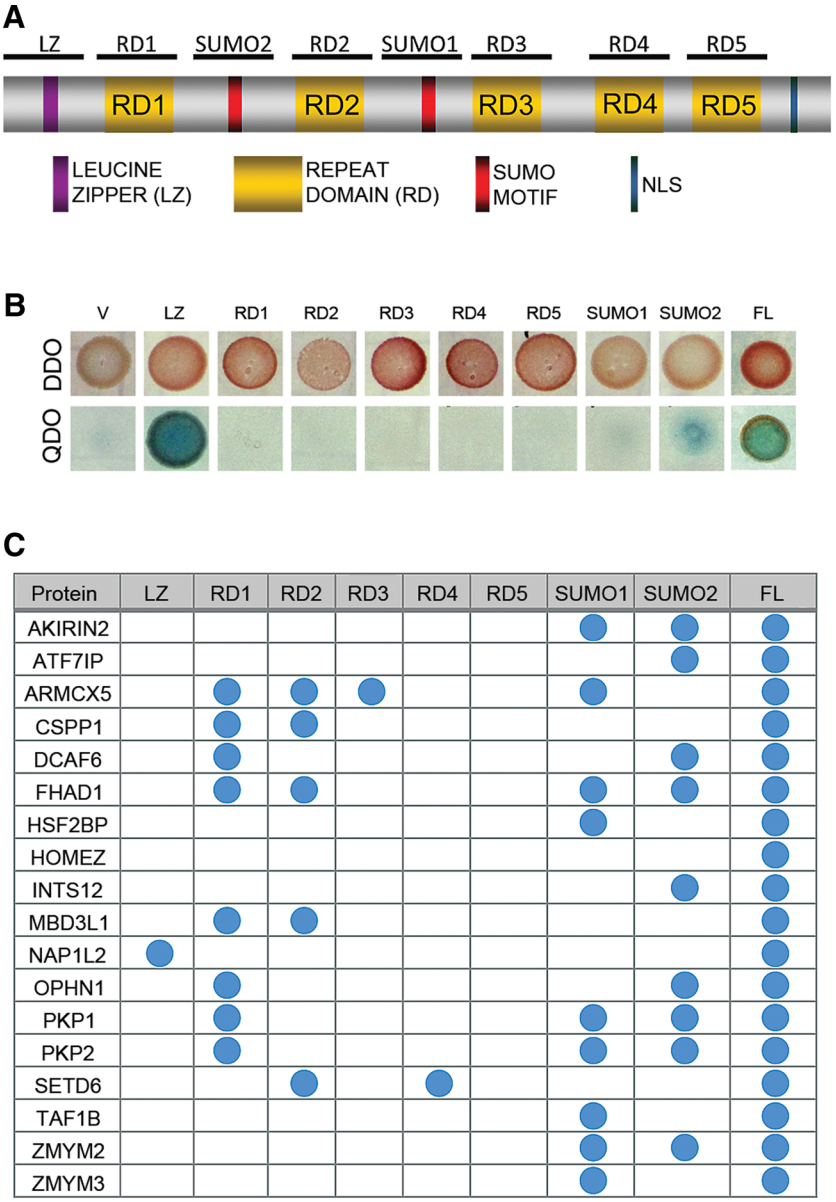


Fig. 5 Mapping of interaction domains in GTF2IRD1 with the proteins identified in the Y2H screens. **a** Diagram of human GTF2IRD1 and the corresponding subdomains used for mapping (*black bars* above). The domains include the leucine zipper region (LZ), five repeat domains (RDs), two regions containing SUMO motifs and a nuclear localization signal (NLS). **b** Representative example images of yeast colonies plated on double dropout agar (DDO) as a control, or quadruple dropout (QDO) agar containing x-α-gal. Each colony represents yeast co-transformed with the empty vector control (V), domain-specific or full-length (FL) bait plasmids together with the prey plasmid identified in the Y2H screen. The example shown is the NAP1L2 interaction. Slight background activity in some surviving yeast is typical of Y2H assays and is ignored. **c** Summary of the domain mapping results using yeast co-transformation

Discussion

In this paper, we have shown for the first time that the endogenous human GTF2IRD1 protein is localized primarily to the nucleus in immortalized human cell lines and in neurons and progenitors differentiated from human embryonic stem cells. The expression of GTF2IRD1 in progenitor and neuronal populations derived from human ES cells suggests a function of this protein from early stages of human neuronal development.

The localization of GTF2IRD1 within the nucleus assumes a speckled pattern that is similar to the TFII-I pattern in HeLa cells (Tanikawa et al. 2011). GTF2IRD1 is clearly excluded from the nucleoli but the distribution does not directly match any of the standard markers of nuclear subcompartments (Sleeman and Trinkle-Mulcahy 2014). However, PLA quantification demonstrated that the strongest potential association of the markers chosen was the HP1 proteins. HP1 proteins have a chromodomain that recognizes the H3K9me2/3 mark and were originally associated with heterochromatin but are now recognized as having multiple roles in transcriptional activation, sister chromatid cohesion, chromosome segregation, telomere maintenance, DNA repair and RNA splicing (Canzio et al. 2014). While HP1α and HP1β are generally localized to heterochromatin, HP1γ is often found in euchromatin at transcription start sites (Sridharan et al. 2013). GTF2IRD1 was also found in close proximity with the transcription factor SP1, which is positively regulated by direct binding of ATF7IP (Fujita et al. 2003); identified as a novel interaction

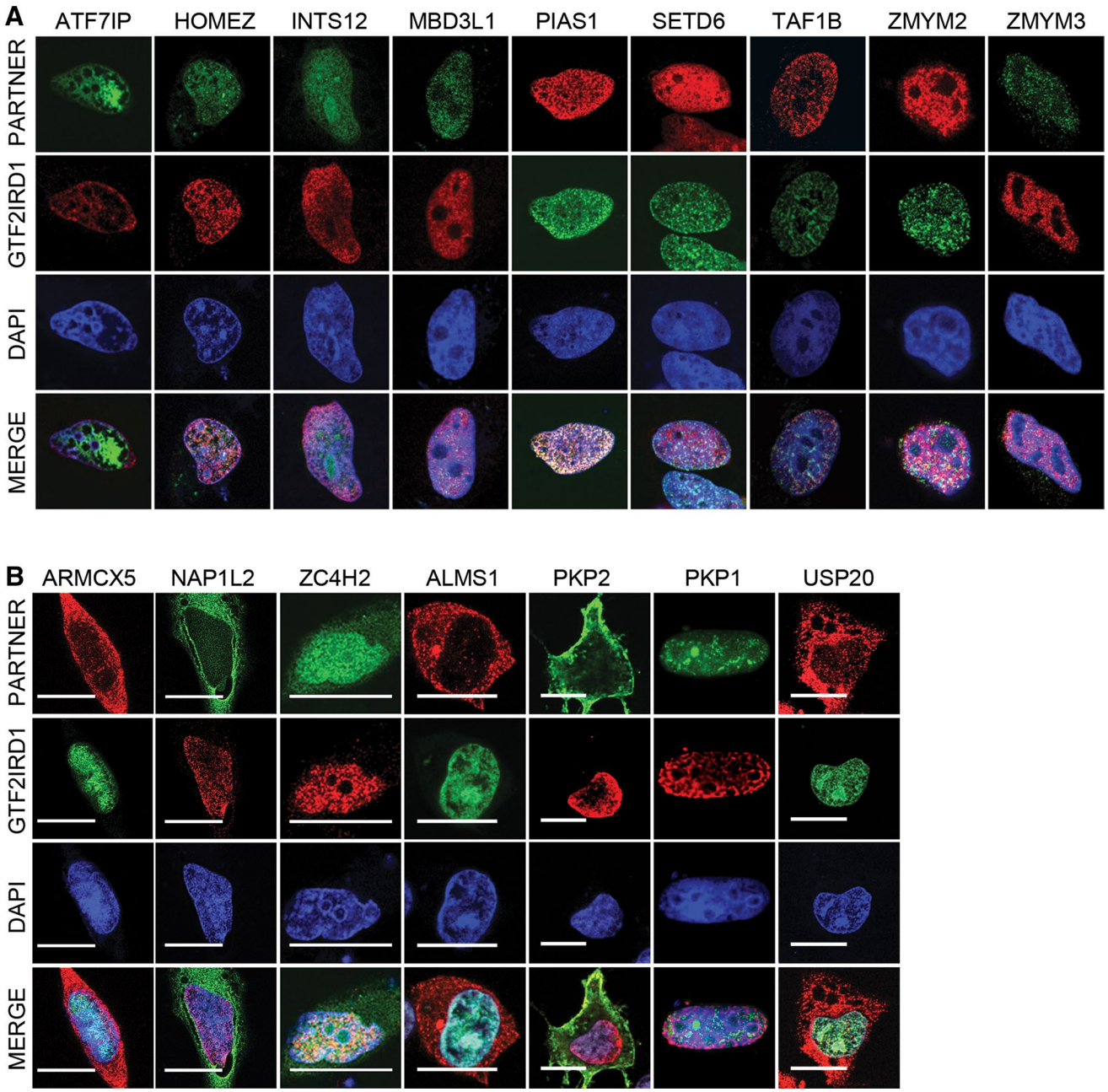


Fig. 6 Subcellular localization of constitutively expressed GTF2IRD1 and the novel putative protein partners. **a** Confocal immunofluorescence analysis of HeLa cells transfected with plasmids encoding human GTF2IRD1-Myc (*red*) or GTF2IRD1-EGFP (*green*) together with plasmids encoding the partner, also tagged with Myc, EGFP or FLAG (PIAS1 only). All of the proteins in **a** were found to localize to the nucleus. Over-expression of some proteins typically causes abnormal appearance of the nucleus, which is apparent in the DAPI images. This is assumed to be a consequence of the impact of the partner on nuclear behaviour. **b** An identical analysis with partner proteins that were found to show cytoplasmic and nuclear localization. PKP1 was not identified in the Y2H screens but was selected due to homology to PKP2, which localizes to the cell periphery. *Scale bars* represent 20 µm (color figure online)

partner in this study. Close proximities were observed at a lower frequency with the chromatin mark H3K4Me3, found at the transcription start site of active genes and with H3K27Me2/3, which is a mark mediated via PRC2, a key repressive factor for the regulation of developmental genes (Golbabapour et al. 2013).

Based on these associations alone, one might speculate that GTF2IRD1 plays a role in transcriptional regulation and developmental gene silencing. These ideas fit very well with previous observations regarding the repression of multiple tissue-specific genes in a transgenic system (Issa et al. 2006), the direct negative

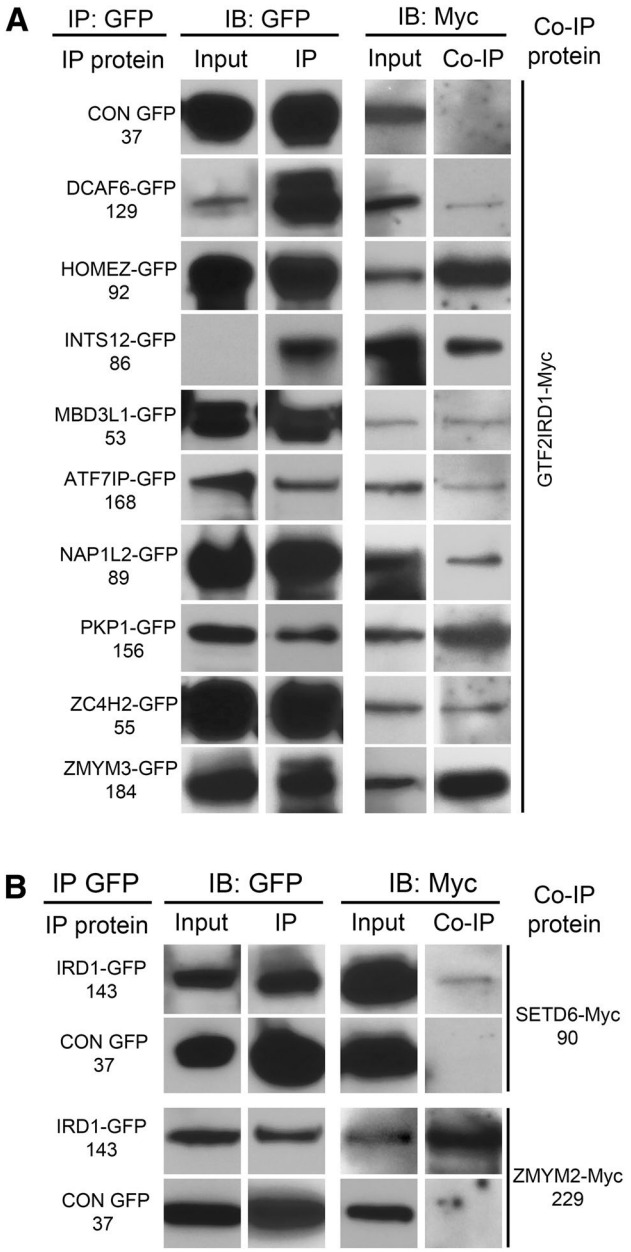
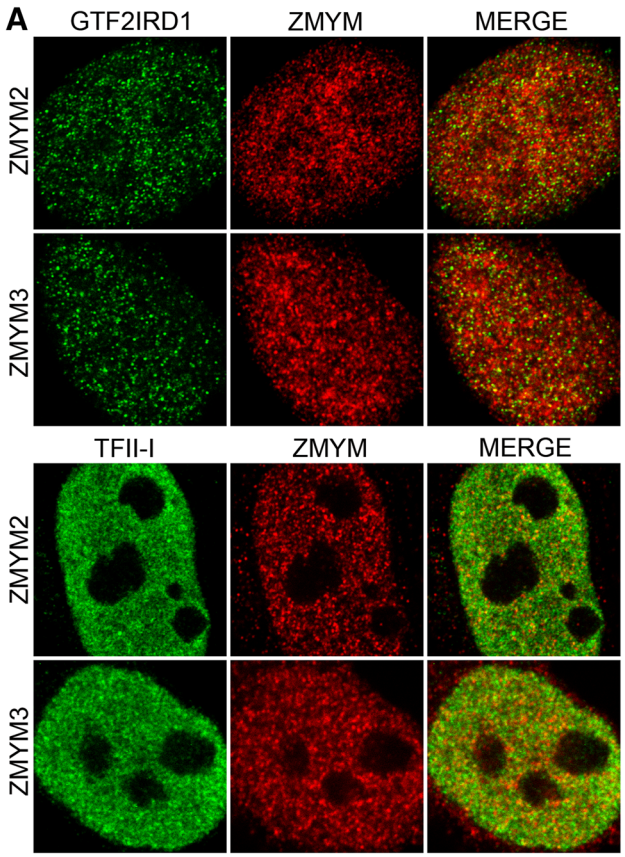


Fig. 7 Novel GTF2IRD1 interactions with nuclear proteins revealed by co-immunoprecipitation in mammalian cells. *Panels* show western blot analyses of HeLa cells transiently transfected with the indicated constructs. **a** Protein partners were immunoprecipitated (IP) with anti-GFP antibody and immunoblotted (IB) with anti-GFP to show successful immunoprecipitation. In one case (INTS12-GFP), the loading of the input was too low to be detected but sufficient protein was recovered in the IP. Immunoblotting with anti-Myc to reveal co-immunoprecipitation (Co-IP) of GTF2IRD1-Myc showed that GTF2IRD1 was recovered in all experiments, except for the pEGFP vector control (CON GFP). **b** Due to limited plasmid clone availability, some partners were assayed in the reverse configuration. HeLa cells were transfected with plasmids encoding GTF2IRD1-EGFP or EGFP alone and SETD6-Myc or ZMYM2-Myc. Proteins were immunoprecipitated using anti-GFP antibody and the interactions detected by immunoblotting with anti-Myc antibody. *Numbers* below the construct names represent the molecular weight in kDa, which was assessed as approximately correct in all cases against molecular weight markers



B

	Combined		Background controls				
	GTF2IRD1/ZMYM		GTF2IRD1		ZMYM		Corrected
PLA	Dots	SD	Dots	SD	Dots	SD	Mean
GTF2IRD1 + ZMYM2	118	45	35	15	8	5	**75**
GTF2IRD1 + ZMYM3	74	37	35	15	3	2	**36**
	TFII-I/ZMYM		TFII-I		ZMYM		
TFII-I + ZMYM2	87	33	13	7	8	5	**66**
TFII-I + ZMYM3	92	41	13	7	5	3	74

Fig. 8 Endogenous co-localization of GTF2IRD1 and TFII-I with ZMYM2 and ZMYM3 using confocal immunofluorescence analysis and PLA. **a** Confocal immunofluorescence analysis of HeLa cells using antibodies against the proteins indicated. Despite the previously reported association of TFII-I with ZMYM2 and ZMYM3 in immunoaffinity analysis of HeLa cell extracts, only partial co-localization is observed. **b** PLA quantification of the incidence of mean close proximity per nucleus between GTF2IRD1 and TFII-I with ZMYM2 and ZMYM3. The *table* indicates the total combined mean PLA dots per nucleus, the contribution of the total caused by GTF2IRD1 (IRD1) background signal, the ZMYM background signal and the resulting estimate of the corrected PLA mean. The adjacent columns show the associated estimates of standard deviation (SD) for each mean

autoregulation of the *GTF2IRD1* promoter/enhancer by its own protein product (Palmer et al. 2010) and the fact that mouse *Gtf2ird1* is widely and robustly expressed during development but restricted to specific cell types such as neurons and brown adipose tissue during adulthood (Palmer et al. 2007).

A large group of novel GTF2IRD1 protein–protein interactions was identified by Y2H screening. Many of these interactions were tested in mammalian cells via co-localization and co-IP; the latter approach forming the basis for a broader interactional network summary (Supplementary Material, Fig. S3). As anticipated, a large number of GTF2IRD1 novel partners are nuclear-localized or have the capability to shuttle into the nucleus, while some are generally cytoplasmic or extracellular and are therefore more likely to be artefacts of the screening system, although there is no additional evidence to support that conclusion. Putting the proteins of primary interest into functional groups, several broad categories emerge; such as nuclear import functions (KPNA1-4), post-translational modifications of ubiquitination (e.g. USP20, USP33 and FBXW10) and SUMOylation (PIAS1 and PIAS2), DNA binding proteins and transcriptional co-regulators (e.g. ELF2, HOMEZ, TRIP11, ZC4H2) and the largest group, which is primarily associated with chromatin regulation (e.g. SETD6, ATF7IP, DCAF6, ZMYM2, ZMYM3, ZMYM5, MBD3L1 and NAP1L2).

The association of GTF2IRD1 with gene silencing functions is consistent with the identification of multiple binding partners that play a role in transcriptional regulation through chromatin modification (Supplemental Material, Table S2). One might predict on this basis that a major functional role for GTF2IRD1 is to nucleate complexes of proteins that are capable of changing histone marks and direct them to specific locations in the genome, either through the direct DNA binding properties of GTF2IRD1 or by association with other transcription factors. The identification of 3 members of the ZMYM family in the screens is consistent with the isolation of ZMYM2 and ZMYM3 in association with TFII-I in the same HDAC-containing complex using immunoaffinity purification from endogenous HeLa cell extracts (Hakimi et al. 2003). This would suggest that the interaction between ZMYM proteins and members of the GTF2I family is an evolutionary conserved feature. Endogenous co-immunofluorescence analysis of the ZMYM proteins suggested some overlap with GTF2IRD1 and TFII-I and the PLA quantification demonstrated frequent close proximity of GTF2IRD1 and TFII-I with the ZMYMs. However, it was also clear that only a fraction of the GTF2IRD1 and TFII-I protein population was in association with the ZMYMs, suggesting that while endogenous TFII-I has been co-purified in complexes containing ZMYM2 and ZMYM3 (Hakimi et al. 2003), these data do not necessarily provide a picture of how TFII-I is distributed in various complexes and such interactions may be transient and fractional.

Several proteins fall into a category that could indicate a signalling role. The largest grouping of these includes 3 proteins that localize to the primary cilium/centrosome complex (ALMS1, BBS4 and CSPP1) as well as 3 other proteins that are linked with primary cilium function (TRIP11, USP20 and USP33). No previous reports have indicated a role for GTF2IRD1 in this structure and there is no evidence as yet to suggest that GTF2IRD1 shuttles to this site but the isolation of 6 proteins belonging to this grouping in an unbiased screen seems beyond the likelihood of coincidence and could initiate a valuable new line of future investigation. All 3 of the main proteins isolated are associated with ciliopathies: mutations in BBS4 cause Bardet–Biedl syndrome 4 (OMIM #615982); ALMS1, Alstrom syndrome (OMIM #203800) and CSPP1, Joubert syndrome 21 (OMIM #615636). Primary cilia in specialized sensory cells are well known but it is now clear that these structures are virtually universal in all cell types, playing critical roles in the sonic hedgehog and Wnt signalling pathways and are particularly important in the developing brain (Guemez-Gamboa et al. 2014; Han et al. 2009).

GTF2IRD1 was shown to interact with 3 members of the ARM repeat-containing family; PKP1, PKP2 and ARMCX5. The plakophilins localize to the cytoplasmic surface of desmosomes but also localize to the nucleoplasm in a wide range of cells. They are widely viewed as signalling proteins that shuttle between these locations playing roles of structural scaffold at the desmosome and transcriptional regulation in the nucleus (Bass-Zubek et al. 2009), being capable of potentiating β-catenin/TCF-mediated transcriptional regulation (Chen et al. 2002). ARMCX5 function is poorly understood but evidence suggests that the *Armcx* genes arose as a cluster on the X chromosome as a result of retrotransposition from *Armc10*. These genes encode proteins that are highly expressed in the developing and adult nervous system, localize both to the nucleus and to mitochondria and play a role in the distribution and dynamics of the mitochondria (Lopez-Domenech et al. 2012).

Eighteen of the novel partner proteins were mapped to interaction domains in GTF2IRD1, highlighting two important points. First, the majority of the interactions localized to the SUMO domains, indicating that some of these interactions may be regulated by post-translational modification of this domain, as was previously shown for ZMYM5 (Widagdo et al. 2012). The second point is that the other major site of interaction is RD1, although binding to the other repeat domains was also common. The repeat domains are known to be the site of DNA binding activity and RDs 2–5 all show varying levels of DNA binding and sequence specificity (Vullhorst and Buonanno 2005), whereas RD1 does not bind DNA. It is difficult to understand what evolutionary advantage was bestowed by the internal duplication of the repeat domains, leading to their expansion to 5 copies in humans and 6 copies in mice. It seems unlikely that this was driven by a need to

refine direct DNA binding properties, although this does restrict high affinity binding to sites that contain at least 2 GGATTA recognition sequences, as shown for the autoregulation of the GTF2IRD1 promoter/enhancer (Palmer et al. 2010). If, however, it is assumed that the RDs form an important protein interaction surface, an evolutionary expansion of this domain would initially amplify the number of proteins with which GTF2IRD1 could interact and subsequent divergence of the repeat domain sequence could diversify the range of simultaneous protein–protein interactions. Alternatively, adding repeat domains could provide a secondary interaction surface for the same partner protein, thus allowing greater control over the binding reaction. The multiple binding sites of several of the partner proteins in the Y2H mapping experiments indicate that the latter scenario is possible.

In conclusion, this paper provides visual and biochemical insights into the functions of GTF2IRD1 using unbiased systems. These data form the basis for a set of testable hypotheses that places GTF2IRD1 as a nuclear protein capable of engaging partner proteins in transcriptional regulation through chromatin modification. In addition, GTF2IRD1 may be located to sites on the genome through interactions with other DNA binding proteins identified in the Y2H screens and may integrate cytoplasmic signals via ARM repeat proteins or members of the centrosome/primary cilium complex.

Materials and methods

Plasmids

For Y2H experiments, bait (pGBKT7) and prey (pGADT7) plasmids were obtained from the Matchmaker Gold Yeast Two-Hybrid System (Clontech Laboratories). pGBKT7 constructs containing full-length human *GTF2IRD1* cDNA or the 88 amino acid domain regions (Fig. 5) have been described previously (Widagdo et al. 2012). Mammalian expression constructs for GTF2IRD1 (pMyc-GTF2IRD1 and pEGFP-GTF2IRD1) were described previously (Widagdo et al. 2012). A detailed list of all other constructs used in this study is shown in Supplementary Material, Table S1.

Yeast two-hybrid assays and library screening

Saccharomyces cerevisiae strain AH109 was transformed with both prey and bait plasmids using the standard lithium acetate/polyethylene glycol protocol (Gietz and Woods 2002). Diploids were grown on double dropout (DDO) selective medium which lacks tryptophan and leucine at 30 °C for 4–5 days. To identify protein interactions, cells were grown on a quadruple dropout (QDO) medium with x-α-gal (40 µg/mL), deficient in tryptophan, leucine, histidine and adenine. For library screening, pGBKT7-GTF2IRD1 was transformed into the AH109 yeast strain and mated with the Mate and Plate (Clontech Laboratories) Universal Mouse (Normalized) library (#630482) or the Human Brain (Normalized) cDNA library (#630486), according to the manufacturer's protocol.

Clones appearing on QDO medium after 4–7 days were further analysed by re-streaking onto QDO/x-α-gal plates. Plasmids were extracted from the yeast, grown in *E.coli* and retransformed into haploid AH109 yeast together with pGBKT7-GTF2IRD1 to confirm the interactions. Identification of inserts in the prey plasmids was performed by Sanger sequencing and BLASTn (NCBI) searches. Protein information was retrieved from the UniProt Consortium database.

Cell lines and transfections

HeLa and HEK-293 cells were grown in Dulbecco's modified Eagle's medium, supplemented with 10 % foetal bovine serum, and 1× penicillin (100 U/ml)/streptomycin (100 µg/ml) at 37 °C in 5 % CO_2. For siRNA transfection, the ON-TARGETplus GTF2IRD1 siRNA SMART pool (L-013262-00-0005, sequences: GUGUGCAGAUCCUGUUUAA, UCACGGGUCUGCCUGAUGA, AGUAUCCACUUCAUCAUUA, UCCCGGGACCUCUUAAUUA; Dharmacon) was transfected into HeLa cells (100 pmol/well of a 6-well plate) using Lipofectamine 2000 (Life Technologies) following the product protocol. Transfected cells were incubated for 48 h before analysis. Transient transfections of mammalian expression vectors were performed in HeLa cells using Lipofectamine LTX (Life Technologies) according to the manufacturer's protocol.

The H9 (WA-09, WiCell) human ES cell line was cultured feeder-free on vitronectin-coated plates using MTeSR-1 defined media according to the manufacturer's instructions (Stem Cell Technologies) and maintained at 37 °C with 5 % CO_2. Colonies were mechanically dissected every 7 days and transferred to freshly prepared coated plates. Cell culture media was changed every day. Neural inductions of human ES cells were set up as described (Denham et al. 2012) with some slight modifications. Briefly, human ES cells were mechanically dissected into pieces approximately 0.5 mm in diameter and transferred to laminin-coated organ culture plates in N2B27 medium containing 1:1 mix of Neurobasal medium (NBM) with DMEM/F12 medium. Neurobasal media contained Neurobasal A medium supplemented with 1 % N_2, 2 % B27, 2 mM L-glutamine and 0.5 % Penicillin/Streptomycin (all sourced from Gibco). Cells were cultured in N2B27 media for 14 days without passaging. SB431542 (10 µM, Tocris) and noggin (500 ng/ml, Peprotech) were

supplemented in the N2B27 media for the first 7 days, followed by basic fibroblast growth factor (bFGF; 20 ng/ml, Peprotech) supplementation only for the remaining 7 days. Fresh supplemented media were replaced every second day. Following 14 days, colonies were dissected into pieces and cultured in suspension in NBM supplemented with epidermal growth factor (EGF) and bFGF at 20 ng/mL each (Peprotech) for 1 week to generate neurospheres. Neuronal differentiation was performed by mechanically disaggregating neurospheres and plating the cells onto poly-D-lysine/laminin dishes in unsupplemented NBM for 1–2 weeks, as previously described (Denham and Dottori 2011).

Antibodies

Anti-GTF2IRD1 antibodies included the 333A rabbit polyclonal (A301-333A-1 Bethyl Laboratories Inc.), the epitope for which maps to between amino acids 909–959 of the human protein, and the M19 rabbit polyclonal (sc-14714 Santa Cruz), for which the epitope maps to 'within an internal region of WBSCR11 of mouse origin', according to the manufacturer's description. For TFII-I, the antibodies used were #4562 rabbit polyclonal (Cell Signaling Technology, proprietary epitope information) for western blot or the #sc-9943 goat polyclonal for immunofluorescence (Santa Cruz) for which the epitope is described as 'mapping near the C-terminus of human TFII-I'. The antibody obtained for GTF2IRD2, #H00084163-B01P mouse polyclonal (Abnova) was produced using the full-length human protein as an immunogen. The monoclonal mouse antibodies against MAP2ab (#MA5-12823, Thermo Scientific) and β-tubulin III (#MAB1637 Merk Millipore) were used as neuronal markers. Antibodies against nuclear subcompartments included mouse monoclonal antibodies; (Abcam), anti-coilin (ab11822-50), anti-Histone H3 trimethyl K4 (ab12209), anti LAP2 (ab11823), anti-nuclear pore complex (ab24609) and anti-SC-35 (ab11826); (Active Motif), anti-histone H3 tri-methyl K9 (#39286), anti-histone H3 di/tri-methyl K27 (#39538), anti-HP1α (#39978), anti-HP1β (#39980) and anti-HP1γ (#39982); (Abnova), SP1 (H00006667-M02) and (Santa Cruz) anti-PML (N-19) sc-9862 goat polyclonal IgG.

Antibodies against epitope tags and GFP included; mouse monoclonal anti-Myc antibody clone 9E10 (Sigma), anti-FLAG rabbit polyclonal F7425 (Sigma), anti-HA mouse monoclonal #MMS-101R-500 (Covance) and the rabbit polyclonal anti-GFP #ab290 (Abcam). To detect the endogenous ZMYM proteins, we used the rabbit polyclonal anti-ZNF198 (ZMYM2) A301-710A and anti-ZNF261 (ZMYM3) A300-200A (Bethyl Laboratories).

Immunoprecipitation and immunoblotting

Cells were lysed 24 h after expression of vector transfection or 48 h after siRNA transfection in lysis buffer (20 mM Tris–HCl, pH7.4; 420 mM NaCl; 10 mM $MgCl_2$; 2 mM EDTA; 10 % Glycerol; 1 % Triton X-100; 2.5 mM β-Glycerophosphate; 1 mM NaF) supplemented with protease inhibitor cocktail (Sigma P8340) and incubated for 30 min on ice before being sonicated twice for 7 s on ice. Cell lysates were centrifuged at 20,000*g* for 10 min at 4 °C to remove cell debris and pre-cleared by incubation with Pure Proteome Protein A/G magnetic beads (Millipore, #LSKMAGAG02) for 30 min at 4 °C. The anti-GFP antibody (ab290, Abcam) was coupled to the protein A/G Magnetic beads for 30 min at room temperature, and then washed three times in PBS/Tween 20 (0.2 %). Pre-cleared lysates were incubated with the antibody-bound beads at 4 °C overnight. Beads were washed in PBS/Tween 20 four times and proteins were eluted by boiling in 1× Laemmli sample buffer containing 0.1 M DTT. Proteins were resolved by 8 or 10 % SDS-PAGE and transferred to a PVDF membrane for western blot analysis using standard methods. Briefly, membranes were blocked for 1 h in blocking solution (TBS/Tween 20 5 % non-fat milk powder), incubated with the primary antibody for 2 h in the same solution and washed for 3× 10 min in TBS/Tween 20. The secondary antibody incubation was conducted for 45 min in blocking solution, washed as before and signal was detected using the ECL substrates, Clarity (Bio-Rad) or Luminata Forte (Merck Millipore), and exposure to X-ray film.

Immunofluorescence

For immunofluorescence analysis of endogenous proteins, HeLa cells were washed with PBS and then fixed and permeabilized for 15 min in 4 % PFA/0.25 % Triton-X100. For analysis of transfected proteins, 24 h after transfection, HeLa cells were washed with PBS and fixed in ice-cold methanol for 10 min. After fixation, all cells were incubated with blocking buffer (10 % BSA in PBS) for 1 h at room temperature, followed by the primary antibody incubation in 1 % BSA in PBS. Detection was carried out using secondary antibodies conjugated to Alexa Fluor Dyes (Molecular Probes). ProLong Gold Antifade reagent with DAPI (Molecular Probes) was used as mounting media in all preparations, except for stimulated emission depletion (STED) imaging, where DAPI was excluded.

For immunofluorescence analysis where both antibodies were derived from rabbit serum (endogenous GTF2IRD1 and ZMYM2 or ZMYM3), incubations were performed in a sequential manner. HeLa cells were washed with PBS

followed by fixation/permeabilization for 15 min in 4 % PFA/0.25 % Triton-X100. After blocking with 10 % foetal calf serum (FCS) for 30 min, cells were incubated overnight at 4 °C with rabbit anti-GTF2IRD1 (333A) and then blocked for 1.5 h at room temperature with a Fab fragment goat anti-rabbit (1:45, Jackson Immunoresearch Laboratories, Inc), which converts the presentation of the rabbit IgG (H + L) of the primary antibody into a goat antigen. This was followed by incubation with a secondary anti-goat antibody conjugated to Alexa Fluor 488. The primary rabbit antibody of the protein partners was added and detected with a secondary anti-rabbit antibody conjugated to Alexa Fluor 594 in the second phase. Negative controls were performed to ensure the complete blocking of rabbit IgG from the first primary antibody.

Cells were visualized by confocal microscopy using a Leica TCS SP5 microscope under ×63 or ×100 magnification. For stimulated emission depletion (STED) experiments, the imaging system was connected to a 592 nm continuous wave depletion laser.

Proximity ligation assay (PLA)

PLA was performed using HeLa cells grown on 12 × 12 mm coverslips and fixed for endogenous protein detection as described earlier (Immunofluorescence). The assay was carried out using the Duolink kit (Olink AB), following the manufacturer's protocol, using red detection reagents and the secondary probes provided when the primary antibody pairs were from mouse, rabbit or goat origin. Background controls were produced by performing two parallel experiments in which the two primary antibodies are left out of the procedure.

In the cases where a PLA antibody pair consisted of two primary antibodies raised in rabbit, the protocol was modified for sequential primary antibody incubation. After fixation and blocking, samples were incubated overnight with rabbit anti-GTF2IRD1 333A, followed by blocking with goat anti-rabbit Fab fragment (1:45 for 1.5 h) followed by an incubation for 1 h at 37 °C with the PLA probe anti-Goat PLUS Duolink. Then the second rabbit primary antibody was added at room temperature for 2 h and blocked for 1 h at 37 °C with PLA probe anti-Rabbit MINUS. The procedure then continued with the standard PLA protocol. Background controls, lacking one or both primary antibodies with Fab fragment incubation were performed to ensure complete blocking of the first primary antibody rabbit IgG.

For each sample, three-dimensional acquisitions (z-stacks) of the nucleus were obtained and the maximum intensity projection pictures were created and analysed for the number of PLA positive puncta. Positive PLA signal was counted (Image J, manual cell counter tool) as nuclear dots per cell, in 30 cells per experiment ($n = 2$ experiments). Background control levels were estimated by counting the dots per nucleus per cell in the background control slides using the same procedure.

Online resources

The interactional network of GTF2IRD1, was generated using Cytoscape 3.1.1 (The Cytoscape Consortium) (Shannon et al. 2003), retrieving protein–protein interactions from IntAct database: Cytoscape: http://www.cytoscape.org/ and IntAct, EMBL-EBI: http://www.ebi.ac.uk/intact/.

Acknowledgments We thank Kylie M. Taylor for her technical assistance. We are grateful for the plasmid constructs provided by the researchers detailed in the Supplementary Table 1. We extend thanks to the Biomedical Imaging Facility, from the Mark Wainwright Analytical Centre at UNSW Australia for their training and support for the microscopy techniques. PC-M and CPC are recipients of a CONICYT-Becas Chile scholarship from the Government of Chile.

Compliance with ethical standards

Funding This work was supported by the National Health and Medical Research Council of Australia (Project Grant 1049639).

Conflict of interest The authors declare no conflict of interest.

References

Antonell A, Del Campo M, Magano LF, Kaufmann L, de la Iglesia JM, Gallastegui F, Flores R, Schweigmann U, Fauth C, Kotzot D, Perez-Jurado LA (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. doi:10.1136/jmg.2009.071712

Bass-Zubek AE, Godsel LM, Delmar M, Green KJ (2009) Plakophilins: multifunctional scaffolds for adhesion and signaling. Curr Opin Cell Biol 21:708–716. doi:10.1016/j.ceb.2009.07.002

Bayarsaihan D, Ruddle FH (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

Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, Jensen LJ (2014) COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database (Oxford) 2014:bau012. doi:10.1093/database/bau012

Canzio D, Larson A, Narlikar GJ (2014) Mechanisms of functional promiscuity by HP1 proteins. Trends Cell Biol 24:377–386. doi:10.1016/j.tcb.2014.01.002

Caraveo G, van Rossum DB, Patterson RL, Snyder SH, Desiderio S (2006) Action of TFII-I outside the nucleus as an inhibitor of agonist-induced calcium entry. Science 314:122–125. doi:10.1126/science.1127815

Chen X, Bonne S, Hatzfeld M, van Roy F, Green KJ (2002) Protein binding and functional characterization of plakophilin 2. Evidence for its diverse roles in desmosomes and beta-catenin signaling. J Biol Chem 277:10512–10522. doi:10.1074/jbc.M108765200

Denham M, Dottori M (2011) Neural differentiation of induced pluripotent stem cells. Methods Mol Biol 793:99–110. doi:10.1007/978-1-61779-328-8_7

Denham M, Parish CL, Leaw B, Wright J, Reid CA, Petrou S, Dottori M, Thompson LH (2012) Neurons derived from human embryonic stem cells extend long-distance axonal projections through growth along host white matter tracts after intra-cerebral transplantation. Front Cell Neurosci 6:11. doi:10.3389/fncel.2012.00011

Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, Bouteiller D, Verloes A, LeGuern E, Leboyer M, Brice A (2007) Autism, language delay and mental retardation in a patient with 7q11 duplication. J Med Genet 44:452–458. doi:10.1136/jmg.2006.047092

Enkhmandakh B, Makeyev AV, Erdenechimeg L, Ruddle FH, Chimge NO, Tussie-Luna MI, Roy AL, 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. doi:10.1073/pnas.0811531106

Franke Y, Peoples RJ, Francke U (1999) Identification of GTF2IRD1, a putative transcription factor within the Williams–Beuren syndrome deletion at 7q11.23. Cytogenet Genome Res 86:296–304

Fujita N, Watanabe S, Ichimura T, Ohkuma Y, Chiba T, Saya H, Nakao M (2003) MCAF mediates MBD1-dependent transcriptional repression. Mol Cell Biol 23:2834–2843

Gietz RD, Woods RA (2002) Transformation of yeast by LiAc/SS carrier DNA/PEG Method. Methods Enzymol 35:87–96

Golbabapour S, Majid NA, Hassandarvish P, Hajrezaie M, Abdulla MA, Hadi AH (2013) Gene silencing and Polycomb group proteins: an overview of their structure, mechanisms and phylogenetics. OMICS 17:283–296. doi:10.1089/omi.2012.0105

Guemez-Gamboa A, Coufal NG, Gleeson JG (2014) Primary cilia in the developing and mature brain. Neuron 82:511–521. doi:10.1016/j.neuron.2014.04.024

Gunbin KV, Ruvinsky A (2013) Evolution of general transcription factors. J Mol Evol 76:28–47. doi:10.1007/s00239-012-9535-y

Hakimi MA, Dong Y, Lane WS, Speicher DW, Shiekhattar R (2003) A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. J Biol Chem 278:7234–7239. doi:10.1074/jbc.M208992200

Han YG, Kim HJ, Dlugosz AA, Ellison DW, Gilbertson RJ, Alvarez-Buylla A (2009) Dual and opposing roles of primary cilia in medulloblastoma development. Nat Med 15:1062–1065. doi:10.1038/nm.2020

Hatzfeld M, Haffner C, Schulze K, Vinzens U (2000) The function of plakophilin 1 in desmosome assembly and actin filament organization. J Cell Biol 149:209–222

Howard ML, Palmer SJ, Taylor KM, Arthurson GJ, Spitzer MW, Du X, Pang TY, Renoir T, Hardeman EC, Hannan AJ (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. doi:10.1016/j.nbd.2011.12.010

Issa LL, Palmer SJ, Guven KL, Santucci N, Hodgson VR, Popovic K, Joya JE, Hardeman EC (2006) MusTRD can regulate postnatal fiber-specific expression. Dev Biol 293:104–115. doi:10.1016/j.ydbio.2006.01.019

Jackson TA, Taylor HE, Sharma D, Desiderio S, Danoff SK (2005) Vascular endothelial growth factor receptor-2: counter-regulation by the transcription factors, TFII-I and TFII-IRD1. J Biol Chem 280:29856–29863. doi:10.1074/jbc.M500335200

Jiang W, Sordella R, Chen GC, Hakre S, Roy AL, Settleman J (2005) An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression. Mol Cell 17:23–35. doi:10.1016/j.molcel.2004.11.024

Lopez-Domenech G, Serrat R, Mirra S, D'Aniello S, Somorjai I, Abad A, Vitureira N, Garcia-Arumi E, Alonso MT, Rodriguez-Prados M, Burgaya F, Andreu AL, Garcia-Sancho J, Trullas R, Garcia-Fernandez J, Soriano E (2012) The Eutherian Armcx genes regulate mitochondrial trafficking in neurons and interact with Miro and Trak2. Nat Commun 3:814. doi:10.1038/ncomms1829

Merla G, Brunetti-Pierri N, Micale L, Fusco C (2010) Copy number variants at Williams–Beuren syndrome 7q11.23 region. Hum Genet 128:3–26. doi:10.1007/s00439-010-0827-2

O'Leary J, Osborne LR (2011) Global analysis of gene expression in the developing brain of Gtf2ird1 knockout mice. PLoS One 6:e23868. doi:10.1371/journal.pone.0023868

O'Mahoney J, Guven KL, Joya JE, Robinson S, Wade RP, Hardeman EC (1998) Identification of a novel slow-muslce-fiber enhancer binding protein, MusTRD1. Mol Cell Biol 18:6641–6652

Osborne LR (2010) Animal models of Williams syndrome. Am J Med Genet C Semin Med Genet 154C:209–219. doi:10.1002/ajmg.c.30257

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

Palmer SJ, Tay ES, Santucci N, Cuc Bach TT, Hook J, Lemckert FA, Jamieson RV, Gunnning PW, Hardeman EC (2007) Expression of Gtf2ird1, the Williams syndrome-associated gene, during mouse development. Gene Expr Patterns 7:396–404. doi:10.1016/j.modgep.2006.11.008

Palmer SJ, Santucci N, Widagdo J, Bontempo SJ, Taylor KM, Tay ES, Hook J, Lemckert F, Gunning PW, Hardeman EC (2010) Negative autoregulation of GTF2IRD1 in Williams–Beuren syndrome via a novel DNA binding mechanism. J Biol Chem 285:4715–4724. doi:10.1074/jbc.M109.086660

Palmer SJ, Taylor KM, Santucci N, Widagdo J, Chan YK, Yeo JL, Adams M, Gunning PW, Hardeman EC (2012) GTF2IRD2 from the Williams–Beuren critical region encodes a mobile-element-derived fusion protein that antagonizes the action of its related family members. J Cell Sci 125:5040–5050. doi:10.1242/jcs.102798

Pérez Jurado LA, Wang Y-K, Peoples R, Coloma A, Cruces J, Francke U (1998) A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. Hum Mol Genet 7:325–334

Polly P, Haddadi LM, Issa LL, Subramaniam N, Palmer SJ, Tay ES, Hardeman EC (2003) hMusTRD1a1 represses MEF2 activation of the troponin I slow enhancer. J Biol Chem 278:36603–36610

Proulx E, Young EJ, Osborne LR, Lambe EK (2010) Enhanced prefrontal serotonin 5-HT(1A) currents in a mouse model of Williams–Beuren syndrome with low innate anxiety. J Neurodev Disord 2:99–108. doi:10.1007/s11689-010-9044-5

Ring C, Ogata S, Meek L, Song J, Ohta T, Miyazono K, Cho KW (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. doi:10.1101/gad.963802

Roy AL (2012) Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later. Gene 492:32–41. doi:10.1016/j.gene.2011.10.030

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH Jr, Geschwind D, Roeder K, Devlin B, State MW (2011) Multiple recurrent de novo CNVs,

including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70:863–885. doi:10.1016/j.neuron.2011.05.002

Schneider T, Skitt Z, Liu Y, Deacon RM, Flint J, Karmiloff-Smith A, Rawlins JN, Tassabehji M (2012) Anxious, hypoactive phenotype combined with motor deficits in Gtf2ird1 null mouse model relevant to Williams syndrome. Behav Brain Res 233:458–473. doi:10.1016/j.bbr.2012.05.014

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. doi:10.1101/gr.1239303

Sleeman JE, Trinkle-Mulcahy L (2014) Nuclear bodies: new insights into assembly/dynamics and disease relevance. Curr Opin Cell Biol 28:76–83. doi:10.1016/j.ceb.2014.03.004

Somerville MJ, Mervis CB, Young EJ, Seo EJ, del Campo M, Bamforth S, Peregrine E, Loo W, Lilley M, Perez-Jurado LA, Morris CA, Scherer SW, Osborne LR (2005) Severe expressive-language delay related to duplication of the Williams–Beuren locus. N Engl J Med 353:1694–1701. doi:10.1056/NEJMoa051962

Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, Carey M, Garcia BA, Plath K (2013) Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1gamma in reprogramming to pluripotency. Nat Cell Biol 15:872–882. doi:10.1038/ncb2768

Tanikawa M, Wada-Hiraike O, Nakagawa S, Shirane A, Hiraike H, Koyama S, Miyamoto Y, Sone K, Tsuruga T, Nagasaka K, Matsumoto Y, Ikeda Y, Shoji K, Oda K, Fukuhara H, Nakagawa K, Kato S, Yano T, Taketani Y (2011) Multifunctional transcription factor TFII-I is an activator of BRCA1 function. Br J Cancer 104:1349–1355. doi:10.1038/bjc.2011.75

Tantin D, Tussie-Luna MI, Roy AL, Sharp PA (2004) Regulation of immunoglobulin promoter activity by TFII-I class transcription factors. J Biol Chem 279:5460–5469. doi:10.1074/jbc.M311177200

Tassabehji M, Carette M, Wilmot C, Donnai D, Read AP, 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. doi:10.1038/sj.ejhg.5200396

Tassabehji M, Hammond P, Karmiloff-Smith A, Thompson P, Thorgeirsson SS, Durkin ME, Popescu NC, Hutton T, Metcalfe K, Rucka A, Stewart H, Read AP, Maconochie M, Donnai D (2005) GTF2IRD1 in craniofacial development of humans and mice. Science 310:1184–1187. doi:10.1126/science.1116142

Tay ES, Guven KL, Subramaniam N, Polly P, Issa LL, Gunning PW, Hardeman EC (2003) Regulation of alternative splicing of Gtf2ird1 and its impact on slow muscle promoter activity. Biochem J 374:359–367. doi:10.1042/BJ20030189

Thompson PD, Webb M, Beckett W, Hinsley T, Jowitt T, Sharrocks AD, Tassabehji M (2007) GTF2IRD1 regulates transcription by binding an evolutionarily conserved DNA motif 'GUCE'. FEBS Lett 581:1233–1242. doi:10.1016/j.febslet.2007.02.040

Tipney HJ, Hinsley TA, Brass A, Metcalfe K, Donnai D, Tassabehji M (2004) Isolation and characterisation of GTF2IRD2, a novel fusion gene mapping to the Williams–Beuren syndrome critical region. Eur J Hum Genet 12:551–560

Torniero C, dalla Bernardina B, Novara F, Vetro A, Ricca I, Darra F, Pramparo T, Guerrini R, Zuffardi O (2007) Cortical dysplasia of the left temporal lobe might explain severe expressive-language delay in patients with duplication of the Williams–Beuren locus. Eur J Hum Genet 15:62–67. doi:10.1038/sj.ejhg.5201730

Tussie-Luna MI, Michel B, Hakre S, Roy AL (2002) The SUMO ubiquitin-protein isopeptide ligase family member Miz1/PIASx-beta/Siz2 is a transcriptional cofactor for TFII-I. J Biol Chem 277:43185–43193. doi:10.1074/jbc.M207635200

Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B, Destree A, Maystadt I, Mannik K, Kurg A, Reimand T, McMullan D, Oley C, Brueton L, Bongers EM, van Bon BW, Pfund R, Jacquemont S, Ferrarini A, Martinet D, Schrander-Stumpel C, Stegmann AP, Frints SG, de Vries BB, Ceulemans B, Kooy RF (2009) Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. Eur J Med Genet 52:94–100. doi:10.1016/j.ejmg.2009.02.006

Vullhorst D, Buonanno A (2003) Characterisation of general transcription factor 3, a transcription factor involved in slow muscle-specific gene expression. J Biol Chem 278:8370–8379. doi:10.1074/jbc.M209361200

Vullhorst D, 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. doi:10.1074/jbc.M500593200

Widagdo J, Taylor KM, Gunning PW, Hardeman EC, Palmer SJ (2012) SUMOylation of GTF2IRD1 regulates protein partner interactions and ubiquitin-mediated degradation. PLoS One 7:e49283. doi:10.1371/journal.pone.0049283

Yang W, Desiderio S (1997) BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. Proc Natl Acad Sci USA 94:604–609

Young EJ, Lipina T, Tam E, Mandel A, Clapcote SJ, Bechard AR, Chambers J, Mount HT, Fletcher PJ, Roder JC, Osborne LR (2008) Reduced fear and aggression and altered serotonin metabolism in Gtf2ird1-targeted mice. Genes Brain Behav 7:224–234. doi:10.1111/j.1601-183X.2007.00343.x