



Review

Cite this article: Crombie EM, Cleverley K, Timmers HTM, Fisher EMC. 2024 The roles of TAF1 in neuroscience and beyond. *R. Soc. Open Sci.* **11**: 240790.

<https://doi.org/10.1098/rsos.240790>

Received: 13 May 2024

Accepted: 5 August 2024

Subject Category:

Genetics and genomics

Subject Areas:

genetics, genomics, neuroscience

Keywords:

TATA box binding-protein (TBP)-associated factor 1, TAF1, X-linked dystonia–parkinsonism, X-linked intellectual disability, TFIID, RNA polymerase II

Author for correspondence:

Elizabeth M. C. Fisher

e-mail: elizabeth.fisher@ucl.ac.uk

The roles of TAF1 in neuroscience and beyond

Elisa M. Crombie¹, Karen Cleverley¹, H. T. Marc

Timmers^{2,3,4} and Elizabeth M. C. Fisher¹

¹Department of Neuromuscular Diseases, UCL Institute of Neurology, University College London, London WC1N 3BG, UK

²German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany

³German Cancer Consortium (DKTK), partner site Freiburg, a partnership between the DKFZ, Germany

⁴Department of Urology, Medical Center-University of Freiburg, Breisacher Straße 66, Freiburg, 79106, Germany

EMC, 0000-0003-4806-6140; KC, 0009-0007-2517-687X; HTMT, 0000-0001-7062-1417; EMCF, 0000-0003-2850-9936

The transcriptional machinery is essential for gene expression and regulation; dysregulation of transcription can result in a range of pathologies, including neurodegeneration, cancer, developmental disorders and cardiovascular disease. A key component of RNA polymerase II-mediated transcription is the basal transcription factor IID, which is formed of the TATA box-binding protein (TBP) and 14 TBP-associated factors (TAFs), the largest of which is the TAF1 protein, encoded on the X chromosome (Xq13.1). *TAF1* is dysregulated in X-linked dystonia–parkinsonism and congenital mutations in the gene are causative for neurodevelopmental phenotypes; TAF1 dysfunction is also associated with cardiac anomalies and cancer. However, how TAF1 contributes to pathology is unclear. Here, we highlight the key aspects of the *TAF1* gene and protein function that may link transcriptional regulation with disorders of development, growth and adult-onset disorders of motor impairment. We highlight the need to experimentally investigate the full range of TAF1 messenger RNA variants and protein isoforms in human and mouse to aid our understanding of TAF1 biology. Furthermore, the X-linked nature of *TAF1*-related diseases adds complexity to understanding phenotypes. Overall, we shed light on the aspects of TAF1 biology that may contribute to disease and areas that could be addressed for future research and targeted therapeutics.

1. Introduction

TATA box binding-protein (TBP)-associated factor 1 (TAF1) plays a key role in the initiation of RNA polymerase II (pol II)-dependent transcription. Human TAF1 is encoded by a large gene located on the X chromosome (Xq13.1), with multiple splice variants of mostly unexplored function and tissue specificity. TAF1 is best known for being part of a protein complex with TBP and other TAFs, forming the basal transcription factor IID (TFIID), which is an essential component of the RNA pol II initiation complex [1]. Similarly to all other TAFs, TAF1 is a highly conserved protein, which has undergone only limited changes throughout eukaryotic evolution [2]. This testifies to the central role TAF1 plays in eukaryotic transcription regulation.

Since the *TAF1* gene is located on the q arm of the human X chromosome, genetic conditions affecting TAF1 function are X-linked and predominantly affect males. Missense mutations in the *TAF1* gene can cause X-linked syndromic mental retardation-33 (MRXS33) also known as X-linked intellectual disability (XLID), which presents with heterogeneous clinical features [3]. Moreover, *TAF1* is dysregulated in X-linked dystonia-parkinsonism (XDP), a progressive neurodegenerative condition, arising from the insertion of a retrotransposon into intron 32 of the gene [4–6]. XDP brains have striatal atrophy and patients manifest late-onset motor impairment similar to that in Huntington's disease (HD) [7,8], in which reduced TAF1 protein expression has been reported [9]. However, why dysfunction of ubiquitously expressed *TAF1* predominantly affects the brain and leads to neurological disorders when mutated is unclear. Much remains to be discovered about this large, fundamentally important protein including the roles of its multiple splicing isoforms.

1.1. TAF1 protein function

The canonical TAF1 protein comprises 1873 amino acids with a molecular mass of 250 kDa. TAF1 is the largest component of the human TFIID complex, which acts as a basal transcription factor for all pol II-mediated transcription in eukaryotes. As part of TFIID, TAF1 acts together with TBP and other TBP-associated factors (18–140 kDa), including TAF2, TAF3, TAF4, TAF4B, TAF5, TAF6, TAF7, TAF8, TAF9, TAF9B, TAF10, TAF11, TAF12 and TAF13 ([10]; figure 1*a*).

TFIID is organized into three structural lobes A, B and C. TBP is bound in an inhibited form to the A lobe (TFIID-A) [14]. Nine TAFs heterodimerize at an interface of their histone-fold domains, specifically TAF4–TAF12, TAF6–TAF9, TAF3–TAF10, TAF8–TAF10 and TAF11–TAF13 [15]. The core TFIID TAFs (TAF4/TAF4B, TAF5, TAF6, TAF9/TAF9B and TAF12) are present in two copies across A and B lobes [14]. TAF1 is the core scaffold protein of the TFIID complex and is located within the C lobe (TFIID-C) [14]. TAF1 is involved in nucleating complex assembly and, as part of TFIID, is shown to form a promoter DNA-binding subcomplex, together with TAF7 and TAF2 [10]. A trimeric TAF2–8–10 complex also forms in the cytoplasm, where TAF8 promotes TAF2 incorporation into TFIID [16]. As the first basal transcription factor, TFIID recognizes the core promoter irrespective of a TATA box (figure 1*b*), promoting the assembly of the pre-initiation complex (PIC), which comprises pol II and other basal transcription factors (TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH) (figure 1*c*). It has been shown by cryoelectron microscopy (cryoEM) that PIC complexes assemble on three types of core promoters: TATA-less, TATA-only and TATA-DBE (consisting of an upstream TATA box and downstream TFIID-binding elements [DBEs]). There are two tracks of promoter assembly where the PIC is directly deposited on TATA-less and TATA-only promoters compared with the stepwise approach for TATA-DBE [10]. Further detail on TFIID complex assembly has been described in multiple recent reviews [11,13].

TAF1 has multiple functions including forming various protein–protein interactions and has DNA-binding activities (figure 2). TAF1 functions as a scaffold, contributing to the assembly and maintaining the structural integrity of TFIID [22], and is involved in the co-translational assembly of several TAFs and TFIID [14,15,23]. TAF1 has been shown to bind the initiator element (INR) directly, which lies at the core promoter downstream (–2 to +4 bp of the transcription start site) of any TATA box present (–31 to –26 bp) [24]. TAF1 also interacts with DNA downstream of the PIC, such as at the motif ten element (MTE)/downstream promoter element (DPE) (+28 to +32 bp) ([10,11,25]; figure 1*b*).

At the TAF1 N-terminal domain (TAND) (figure 2), TAF1 binds TBP to inhibit DNA binding in its non-promoter-bound state [17,26,27]. At its TAF7-interacting domain, TAF1 forms a stable interaction with TAF7, and a smaller winged-helix domain within this region has been shown to bind promoter DNA [18]. TAF1 also contains a conserved Zinc knuckle domain, which is involved in the recruitment of TFIID to endogenous promoters [19]. Towards the C-terminal end, TAF1 contains a unique double

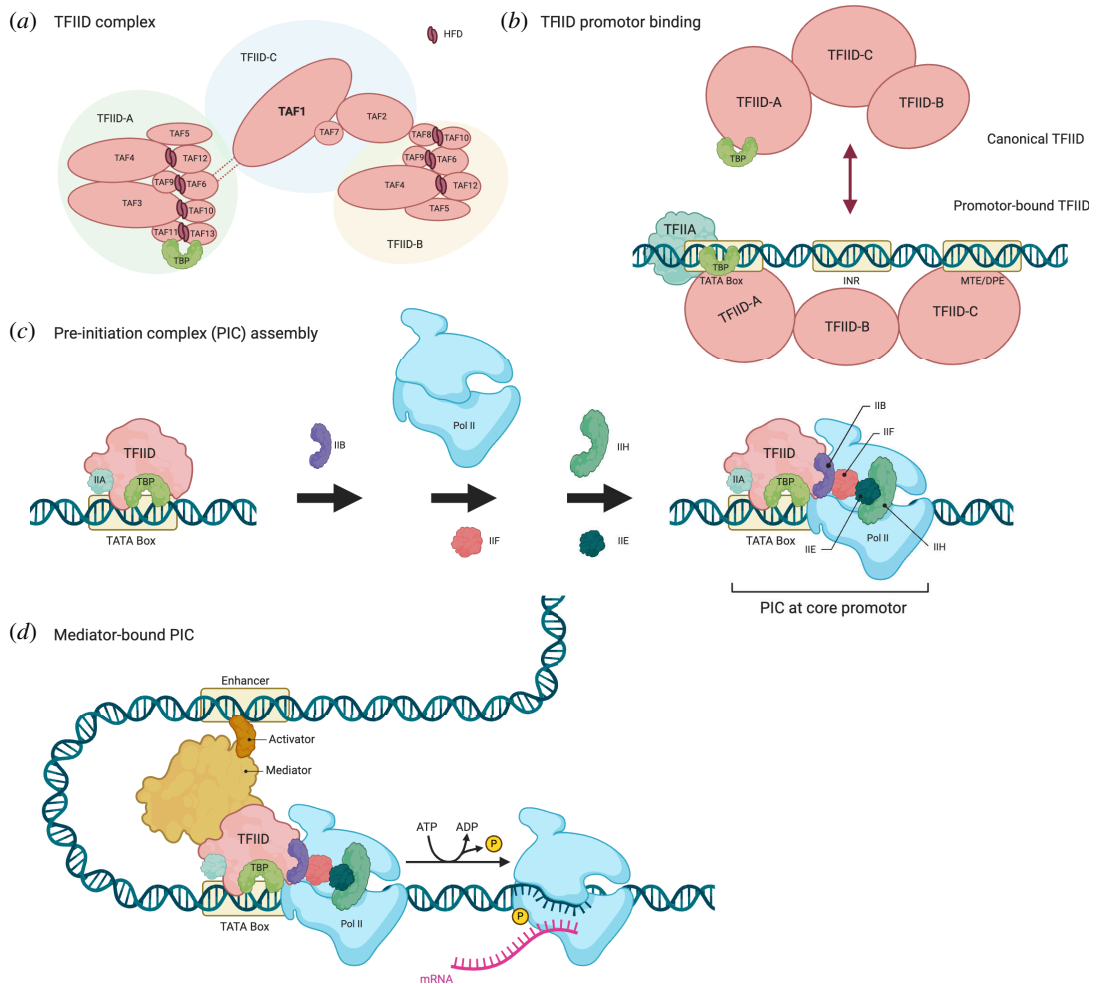


Figure 1. Transcription initiation complex assembly. (a) TFIID complex. TFIID is organized into three lobes: lobes A, B and C (TFIID-A, -B and -C). Certain TAFs are present in both lobes A and B and several heterodimerize through their histone-folding domains (HFDs). The cryo-electron microscopy (cryoEM)-visible parts of TAF1 locate/within TFIID-C i.e. within the C lobe. (b) TFIID exists in two major conformations: ‘canonical’ or ‘promoter-bound’ TFIID. Canonical TFIID exists in a relatively compact conformation, whereas promoter-bound TFIID binds to the TATA box via TBP together with TFIIA, stabilizing the complex at the core promoter. Aside from the TATA box, eukaryotic core promoters comprise multiple elements including initiator (INR), motive ten element (MTE) and downstream core promoter element (DPE) [11]; for (c) and (d), only the TATA box is shown for simplicity. TFIID structure and conformational changes are based on cryoEM data [10,12] and are described in Malik & Roeder [13]. (c) Pre-initiation complex (PIC) assembly. TFIID recognizes the core promoter of a gene to be transcribed, with TBP and TFIIA, then further recruits TFIIB, RNA polymerase II (pol II), TFIIF, TFIIB and TFIIE to assemble the PIC. (d) Mediator-bound PIC. Basal (a.k.a. general) transcription factors, pol II, and the coactivator mediator are recruited to the PIC [11]. Mediator facilitates interactions between the TFIID and transcriptional activators. Current models are reviewed in detail in Malik & Roeder [13]. Size of DNA and proteins not to scale. TAF, TBP-associated factor; TBP, TATA box-binding protein; (TF)IIA, B, D, E, F, H, basal transcription factor 2A, B, D, E, F, H. Created with [BioRender.com](https://www.biorender.com).

bromodomain, which recognizes post-translational modifications on histones and transcription factors including acetylated, butyrylated and crotonylated lysines [20,21,28]. Besides these functions, several enzymatic activities have been attributed to TAF1 in the past, which could not be confirmed in recent TFIID and/or TAF1 studies. These include kinase, histone acetylation and ubiquitination activities [29]. These studies could not be reproduced by others and much of the related work by Sauer *et al.* has been retracted [1]. Overall, TAF1 can regulate transcription through a variety of mechanisms and is probably responsive to epigenetic changes in chromatin.

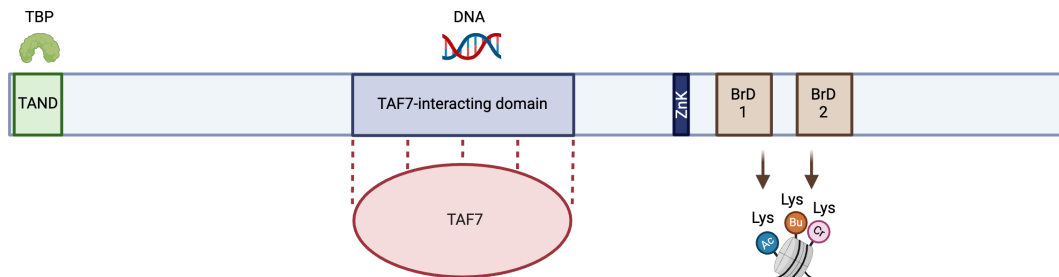


Figure 2. TAF1 protein domains and functions. DNA and protein interactions are shown. The TAF1 N-terminal (TAND) domain interacts with TBP to inhibit DNA binding [17]. The TAF7-interacting domain binds TAF7 to form a stable complex via a heterodimeric triple barrel (not shown); within this region is a DNA-binding domain, which forms DNA–protein interactions via a winged-helix domain [18]. Towards the C-terminal lies a Zinc knuckle domain (ZnK), which aids in the recruitment of TFIIID to endogenous promoters [19]. TAF1 contains tandem bromodomains (BrD1 and BrD2) that bind acetyl-lysines and acyl-lysines, where the second bromodomain of TAF1 binds to butyryl (Bu)- and crotonyl (Cr)-lysine residues on histone tails [20,21]. The diagram shows the relative size of each protein domain, semi-to-scale. Created with [BioRender.com](https://www.biorender.com).

1.2. *TAF1* messenger RNA (mRNA) variants and protein isoforms

1.2.1. Alternative splicing of *TAF1*

The human *TAF1* gene lies in Xq13.1 and produces a canonical transcript (*cTAF1*) of 38 exons (Ensembl 2024, https://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000147133;r=X:71366222-71532374). *cTAF1*¹ is called ‘*TAF1*–204’ in the Ensembl database and is a 7599 bp transcript encoding 1873 amino acids, a 3bp stop codon, and a surprisingly short 5′-untranslated region (UTR) of 18 bp and a 3′-UTR of 1959 bp (figure 3; table 1). Ensembl describes the canonical variant as ‘the most conserved, highly expressed, has the longest coding sequence, and is represented in other key resources, such as NCBI and UniProt’ (Ensembl 2024). *TAF1*–204 is also the Matched Annotation from NCBI and EMBL–EBI (MANE) Select transcript variant, which is the ‘default transcript per human gene that is represented by biology, well-supported, expressed and highly conserved’ (Ensembl 2024).

Currently, we do not know how many *TAF1* mRNA variants exist. In humans, *TAF1* has several internal exons that are found within alternatively spliced mRNA variants in various transcript databases and the literature. Ensembl describes 27 alternatively spliced transcripts (ENSG00000147133), and the NCBI database refers to 13 such transcripts (NG_012771.2 RefSeqGene) derived from multiple tissues including brain, foetal and HeLa cells. At least 10 splice variants are reported in the literature from predominantly brain samples, in which *TAF1* is abundant [6,30]. In figure 3*a,b*, we show only those eight transcripts from Ensembl and the equivalent three from NCBI that are described as potentially protein-coding, i.e. in that carry an open reading frame (ORF) throughout with (mostly) complete 5′- and 3′-UTRs. Without further analysis of full-length *TAF1* transcripts in different tissues, it is impossible to know how many of these mRNAs are real. However, *TAF1*–204 (*cTAF1*), *TAF1*–228, *TAF1*–226, *TAF1*–203 and *TAF1*–221 appear *bona fide* because they have a full-length ORF and 5′- and 3′-UTRs. *TAF1*–201, while incomplete, potentially encodes the neuronal-specific form of *TAF1* (N-*TAF1* or *TAF1*–34²), that has been characterized at the protein level and probably corresponds to the NCBI transcript variant 3. NCBI transcript variant 1 corresponds to the canonical form of *TAF1* (*TAF1*–204) and variant 2 is *TAF1*–203. Thus, in total, six full-length probably *bona fide* protein-coding transcripts have been described in Ensembl and NCBI for human *TAF1* (table 1).

With respect to the two other *TAF1* transcripts described in Ensembl as ‘protein coding’: *TAF1*–217 (described in foetal brain and thymus) and *TAF1*–209 (origin unknown), only contain four and two canonical *TAF1* exons, respectively, plus additional alternative exons that are not described in the literature. Neither transcript is full-length and therefore it is currently impossible to know if they represent functional mRNAs. It is important for the field to determine the full complexity of *TAF1* alternative transcripts, presumably each undertaking a different function, within each cell type, and

¹*cTAF1*, canonical human *TAF1* transcript; *nTAF1*, neuronal human *TAF1* transcript; *cTaf1*, canonical mouse *Taf1* transcript; *nTaf1*, neuronal mouse *Taf1* transcript; C-*TAF1*, canonical protein *TAF1* isoform; N-*TAF1*, neuronal protein *TAF1* isoform.

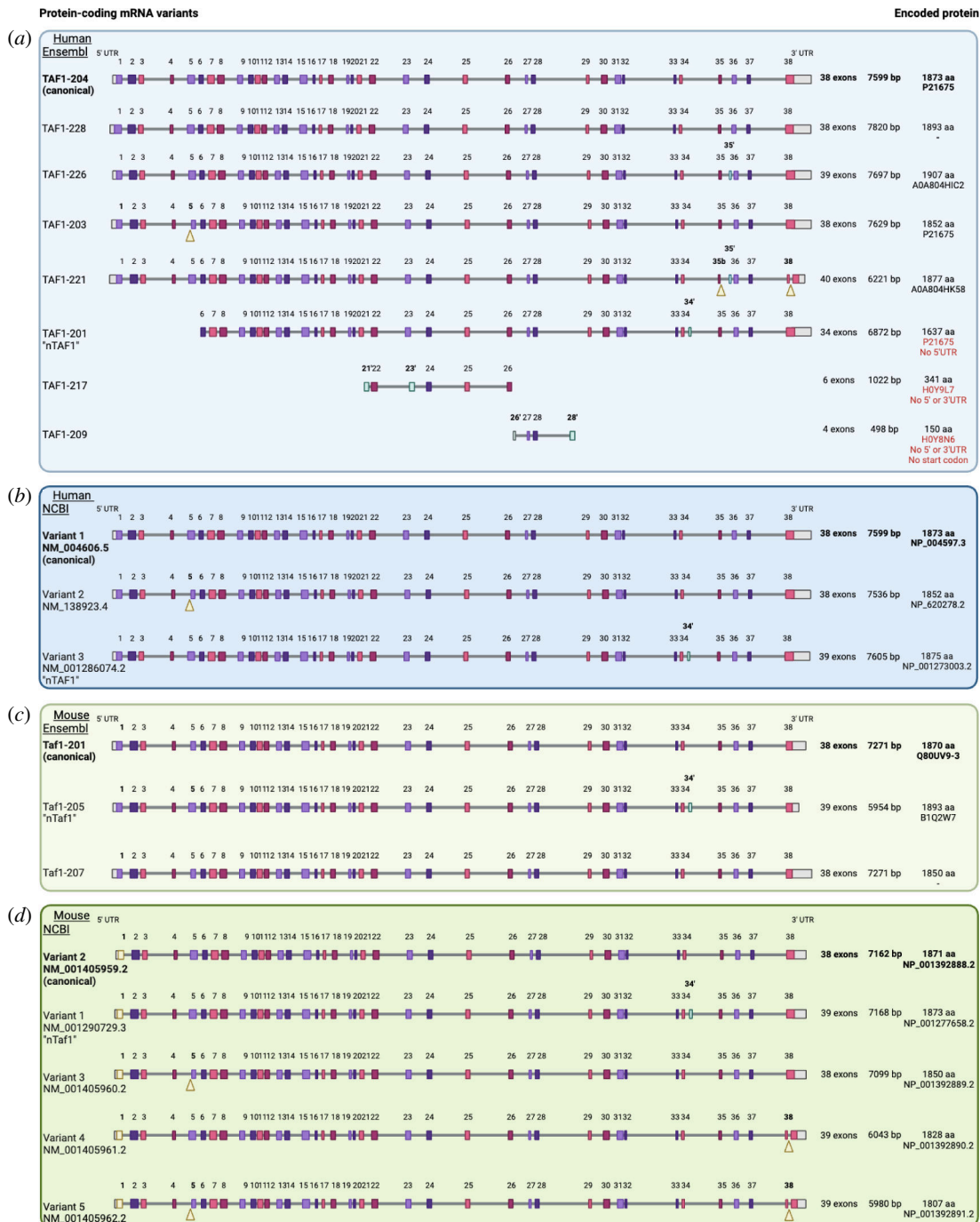


Figure 3. Genomic architecture (exon–intron structure) of protein-coding mRNA variants for human and mouse *TAF1*. (a) Eight protein-coding variants are shown for human *TAF1* as described in Ensembl and (b) three are defined by NCBI. (c) Three protein-coding variants are shown for mouse *Taf1* as categorized by Ensembl and (d) five are categorized by NCBI. The canonical variant (*cTAF1/cTaf1*; bold) and *nTAF1/nTaf1* are indicated. Exons are shown as coloured boxes and introns as grey lines. Grey boxes indicate 5'- and 3'-UTRs. Blue boxes indicate additional exons to the human canonical *TAF1* variant. The total number of exons, amino acid (aa) length, and their amino acid translation by UniProt ID are shown on the right. Yellow arrowheads and bold writing on exon number indicate the absence of nucleotides in exons compared with the same exons in the Ensembl canonical variant of respective human or mouse *TAF1*; yellow boxes and bold writing on exon number indicates sequence variation in exons to the respective human or mouse canonical *TAF1* variant. Red writing of UniProt ID indicates that mRNA is unlikely to be translated to protein owing to missing structural features of a complete open reading frame as indicated. Created with [BioRender.com](https://www.biorender.com).

furthermore, it would be very useful to determine relative levels of *TAF1* transcripts in different tissues, conditions or at different developmental stages.

Table 1. Table of likely human and mouse protein-coding transcripts as named in Ensembl, NCBI and the literature.

	Ensembl variant	NCBI variant	alternative name
human	<i>TAF1</i> –204	variant 1	<i>cTAF1</i>
	<i>TAF1</i> –226		
	<i>TAF1</i> –228		
	<i>TAF1</i> –203	variant 2	
	<i>TAF1</i> –221		
	<i>TAF1</i> –201	variant 3	<i>nTAF1</i> , <i>TAF1</i> –34'
mouse	<i>Taf1</i> –205	variant 1	<i>nTaf1</i> , <i>Taf1</i> –34'
	<i>Taf1</i> –201	variant 2	<i>cTaf1</i>
	<i>Taf1</i> –207		
		variant 3	
		variant 4	
		variant 5	

TAF1 transcripts in online databases mostly differ in alternative exon usage, including smaller exons, primarily towards the 3' end of the *TAF1* gene (figure 3a,b). For example, microexon 34' (6 bp) is present in *TAF1*–201/variant 3/*nTAF1*, whereas exon 35' (102 bp) is present in *TAF1*–221 [30,31]. Alternative exon 35b, is slightly shorter than exon 35a (126 bp versus 117 bp) because it does not contain the nine 3' terminal base pairs [31]. Exon 35b is present in *TAF1*–221 and maintains the reading frame.

We note that some nucleotides are missing from the canonical exons in some *TAF1* variants in the databases, such as 63 bp 3' of exon 5 in *TAF1*–203/variant 2. Similarly, *TAF1*–221 shows sequence variation within exon 38 from the canonical exon 38 and is missing 1081 bp from its 3'-UTR. It is not clear whether these sequence changes are *bona fide* or artefacts of sequencing and/or complementary DNA (cDNA) cloning, however, the changes in nucleotide numbers in the coding regions are multiple of 3 and so would maintain the ORF.

Since most alternative splicing events reported in the literature occur towards the 3'-end of the *TAF1* gene, in a recent paper Capponi *et al.* amplified the region between exons 30 and 38 in cDNA derived from male and female adult human prefrontal cortex and presented the 10 most abundant transcripts ([30]; figure 4b,c). Although it is speculative how these published transcripts relate to the full-length transcripts described in the Ensembl and NCBI databases, one transcript appears to be derived from *cTAF1* (*TAF1*–204, variant 1) or from *TAF1*–203 (variant 2) (which are identical apart from exon 5, which is not included in the sequencing), one is potentially derived from *nTAF1* (*TAF1*–201, variant 3), one is probably from *TAF1*–226, and one is probably from *TAF1*–221 (figure 4b,c). By nanopore long-read sequencing in HeLa cells, Capponi *et al.* showed that approximately 70% of *TAF1* transcripts are *cTAF1*, but there is a large inter-individual variation in relative levels of expression of different variants within the human prefrontal cortex [30]. Aside from these 10 most abundant variants in the brain, they state that remaining low abundance variants are less than 1% individually and make up less than 5% on average (maximum 12%) of the remainder. Other mRNA variants include the incorporation of alternative exon 32', which is 22 bp in length and would cause a translational frameshift leading to a stop codon in exon 34.

Older literature states that *TAF1* may function together with downstream genes such as *ING2* (antisense strand) and *OGT* (sense strand), in what has been referred to as a 'multiple transcript system' (MTS) [31]. This MTS has been described to also include five exons, referred to as exons d₁–d₅ ('downstream' exons 1–5), located 3' to canonical exon 38 ([31,32]; figure 4a). Exons d₁–d₅ were reported as spliced with *TAF1* exons (26–37 but not exon 38) to produce at least eight variants in the human brain ('variants 2a–i') (figure 4d). However, while 13 variants containing d exons are presented in the NCBI database, they are described as non-coding RNA [31–33].

The prevalence and function of possible 'd' mRNA variants are not known [30]. While the 38 canonical exons are highly conserved across vertebrates and invertebrates, the five d exons are less conserved and—if real—predominantly occur in primates [32]. Expression of d₂–d₃ and d₃–d₄ was reported to be higher in the brain than in other tissues, where disease-related sequence changes in d

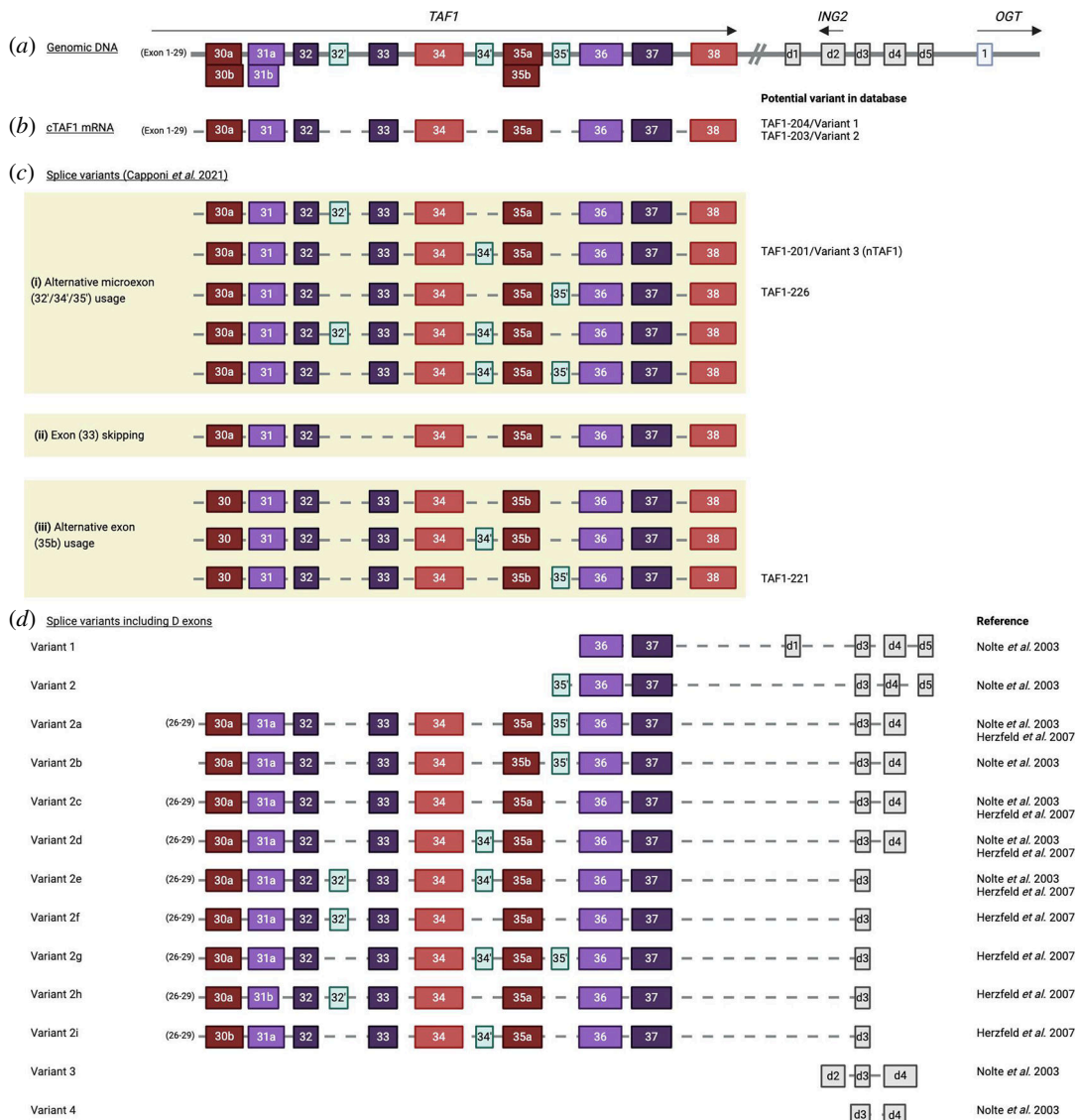


Figure 4. mRNA splice variants of human *TAF1* in the literature. (a) *TAF1* genomic DNA region (Xq13.1) including human *TAF1* exons 30–38 that have been found in mRNAs (5' exons 1–29 are not shown). Alternative exons 32', 34' and 35' and exons 35 a/b are shown in genomic DNA at their positions relative to canonical exons. *TAF1* is transcribed on the sense strand and *ING2* is transcribed downstream of *TAF1* in an antisense orientation. *OGT* is transcribed in a sense orientation downstream. In older literature, d exons 1–5, shown in grey, were described as located downstream of *TAF1* [31] but whether they are spliced with canonical *TAF1* exons is yet to be robustly demonstrated. (b,c) mRNA variants described by Capponi *et al.* involving exon 30–38 taken from sequenced brain transcripts [30]. (b) Canonical *TAF1* (*cTAF1*, *TAF1*–204). (c) Nine further *TAF1* variants resulting from alternative splicing where (i) alternative exons 32', 34' and/or 35' are incorporated; (ii) exon 33 is skipped in the *cTAF1* transcript or (iii) alternative exon 35b is incorporated into the *TAF1* transcript instead of 35a. Exons and introns are not to scale. Data were obtained from long-range reverse transcription-polymerase chain reaction (RT-PCR) in which the amplicon spanned between exons 30 and 38. Whether canonical *TAF1* exons 1–29 are present upstream of exons 30–38 cannot be confirmed, because of lack of upstream sequencing data from the transcripts other than *cTAF1*. We speculate on the presence of *TAF1*–203/variant 2 in NCBI (which is identical to *cTAF1* apart from exon 5 which is not included in the sequencing), *nTAF1* (*TAF1*–201/variant 3), *TAF1*–226 and *TAF1*–221. (d) mRNA variants shown are based on phage-cloned DNA and RT-PCR analyses. Variants 1–2i include *TAF1* exons, whereas variants 3 and 4 are transcribed independently of *TAF1*. Nolte *et al.* investigated transcripts using primers designed between exon 30 and d4, whereas Herzfeld *et al.* investigated transcripts using primers designed between *TAF1* exon 26 and d3. Variants identified by Nolte *et al.* and Herzfeld *et al.* were identified in the human foetal brain and caudate nucleus [31,32]. Created with [BioRender.com](https://www.biorender.com).

exon expression were reported to disturb dopamine transport and metabolism [33]. However, recently, Aneichyk *et al.* could not confirm the presence of these transcripts and did not find MTS exons to be transcribed with *TAF1* exons using *de novo* transcript assembly in fibroblasts and neural cells [4]. Together, it is unlikely that d exons produce functional mRNAs.

1.2.2. Human TAF1 protein isoforms

While many *TAF1* transcript variants exist in the databases and the literature, not all may be *bona fide* mRNAs that are translated into a functional protein. For example, some variants may include retained introns from partially processed mRNAs, potentially resulting in nonsense-mediated decay (NMD) [34,35]. However, as shown in figure 3*a,b*:

- (i) five Ensembl 'protein-coding' mRNA variants (*TAF1*-204 [*cTAF1*], -228, -203, -206 and -221) have intact 5'- and 3'-UTRs, start and stop codons, and a full-length ORF; and
- (ii) the three protein-coding transcript variants in NCBI share homology with three Ensembl variants (*TAF1*-204, variant 1; *TAF1*-203, variant 2; *TAF1*-201, variant 3) (table 1). The sequence homology of the three full-length Ensembl/NCBI transcripts supports their legitimacy.

However, differences in total transcript lengths are shown, such as for *TAF1*-203/variant 2, due to variable 5'- and 3'-UTR lengths probably owing to different sources providing the sequencing data. Furthermore, protein lengths in online databases vary: for example, the canonical form of human TAF1 is 1893 amino acids in UniProt (P21675) but 1873 amino acids would be translated from the Ensembl *TAF1*-204 (canonical) mRNA transcript. The updated sequence on NCBI is 1873 amino acids (NP_004597.3), corrected from the previous version (NP_004597.2). The additional 20 amino acids in UniProt were previously reported, but it now appears that this may have been caused by an error in the prediction software of UniProt. The extra 20 amino acids would result from an ATG, that lies in the 5'-UTR and is upstream from cognate ATG start. Furthermore, the second ATG in the genetic sequence (but not the first upstream) lies in the consensus Kozak sequence, which would direct translation initiation to the second methionine. Therefore, the current evidence suggests that the correct length of human *cTAF1* is 1873 amino acids.

Transcripts containing alternative exon 32' have been published [30]. However, exon 32' is not in any of the protein-coding transcripts in published databases but is present in non-coding transcripts such as *TAF1*-218 (NMD, ENST00000485087.6), *TAF1*-211 (retained intron, ENST00000468167.6), *TAF1*-223 (NMD, ENST00000683358.1) and *TAF1*-220 (NMD, ENST00000682124.1). We note that exon 32' is 22 bp in length and would produce a frameshift and early stop codon resulting in a truncated protein. Finally, transcripts containing d exons do not fulfil the criteria for encoded proteins, since many have incomplete 5'-UTRs and are probably degraded by NMD (ENSG00000147133).

1.2.3. Neuron-specific TAF1

The relative abundance of each of the various transcripts in different tissues and their function have not been comprehensively described and, importantly, other splice isoforms may yet be found in different tissues. However, one transcript, *nTAF1/TAF1*-201 (Ensembl)/variant 3 (NCBI), is of particular interest because it appears to be neuron-specific [6]. *TAF1*-201 fulfils the criteria for 'protein-coding' but has no 5'-sequence before exon 6, which is presumably an artefact, whereas the homologous NCBI variant 3 transcript has an identical coding sequence; both mRNAs include the 6 bp microexon 34', which encodes an alanine and a lysine residue (figure 3*a,b*). While direct protein expression data on this neuron-specific isoform are scarce, in part because no specific antibody is currently available, homology has been found in the mouse in which transcript *Taf1*-205/variant 1 incorporates the same microexon (figure 3*c,d*), indicating that both species produce functional N-TAF1 protein. Note that, confusingly, the mouse and human transcript numbers do not correspond, i.e. human *TAF1*-201 is not the homologous transcript to mouse *Taf1*-201.

An antibody made with an epitope across canonical exons 34–35 (towards the C-terminal), detected (by Western blot) the N-TAF1 isoform in human 293 T cells that were transiently transfected with green fluorescent protein (GFP)-*TAF1*-34' cDNAs [30,36]. A mouse protein containing this epitope was detected in the cortex and striatum of the mouse brain by immunohistochemical staining using the same *TAF1*-34'-specific antibody [36]. In addition, long-range reverse transcription-polymerase chain reaction (RT-PCR) showed that the identical 6 bp microexon 34' sequence is incorporated into mouse *Taf1* transcripts at this conserved region [37]. Long-range PCR methods used to generate these data involved first-strand synthesis from RNAs by long RT and subsequently fragment PCR by use of the long RT products (i.e. cDNA) as a template [6,37]. Thus, whether other microexon 34' containing transcripts are amplified by these methods is uncertain. However, since discrimination of microexon-containing mRNAs from canonical mRNAs is challenging owing to the small size of such

microexons, Capponi *et al.* used *in situ* hybridization to discriminate between *cTAF1* and *nTAF1* using the BaseScope method with specific probes designed against the 6 bp microexon 34' or the sequence spanning the flanking exons [36]. This revealed the presence of the microexon experimentally, but this RNA-based analysis did not analyse the sequences 5' and 3' of the microexon-containing region and thus does not demonstrate the transcript length.

Capponi *et al.* provided evidence for microexon 34' incorporation into the *nTAF1* transcript by neuron-specific splicing factor Serine/Arginine Repetitive Matrix 4 (SRRM4) [30,36] and Cirnaru *et al.* reported that in rat striatum, micro RNA (miRNA) knockdown (KD) of *nTaf1* by intracerebroventricular adeno-associated virus (AAV) injection successfully reduced N-TAF1 expression as shown by Western blot of tagged GFP [38]. It remains unclear whether microexon 34' is incorporated into other transcripts/protein isoforms.

1.2.4. Mouse *Taf1* splice variants and protein isoforms

cTAF1 protein sequences are highly conserved between mouse and human, with an amino acid sequence identity of approximately 96% (UniProt: Q80UV9 versus P21675, Release 2024_01). Mouse genes tend to have fewer transcripts than their human orthologues; in Ensembl, mouse *Taf1* has seven transcripts, and NCBI indicates that there are five *Taf1* transcripts (figure 3c,d). Of the Ensembl transcripts, three are probably full-length *bona fide* transcripts, having 5'- and 3'-UTRs and a long ORF and appear to correspond to human *cTAF1* (mouse *Taf1*-201) and *nTAF1* (mouse *TAF1*-205) (table 1; figure 3a,c). In Ensembl, compared with the human canonical transcript (*TAF1*-201), mouse transcripts show sequence variation at exon 1 and exon 5 (*Taf1*-201 and -207). Meanwhile in NCBI, mouse transcript variants 3 and 5 contained the shorter form of exon 5 present in human *TAF1*-203/variant 2, and mouse variants 4 and 5 contained the sequence variation in exon 38 that is present in human *TAF1*-221 (figure 3d). Conservation of these sequences between mice and humans supports the legitimacy of these findings.

2. TAF1 and development and disease

2.1. TAF1 function and experimental loss of function

A large body of data suggests that TAF1 is essential for embryonic development [3,39,40]. Moreover, TAF1 and other TFIID components may play key roles in cell proliferation and growth, such as TAF1 in the G1 phase of the cell cycle [41,42]. Effects may be cell state specific, e.g. expression of TAF1 and other TFIID components (TAF4 and TBP) is higher in myoblasts (muscle cell progenitors) compared with myotubes (differentiated muscle cells) [43]. Stemness seems to be regulated by TFIID TAFs [44]. In addition, TAF1 has been reported to bind to PAX3 (regulator of myogenesis), leading to its proteasomal degradation [45]. This is similar to another canonical TFIID complex component, TAF4, for which expression in rat embryonic cortical neural stem cells (NSCs) was strong and decreased with neuronal (but interestingly not glial cell) differentiation [46].

In mice, TAF4 is required for mid-stage gestation and the expression of paralogues TAF4A and TAF4B overlaps at early embryonic stages [47]. TAF4B is also highly expressed in embryonic stem cells (ESCs) and is downregulated during differentiation, but in contrast to TAF4, maintains rather than inhibits ESC proliferation and cell cycle progression [48]. Knockout (KO) of *Taf4a* in mouse ESCs prevented the completion of differentiation into glutamatergic neurons and cardiomyocytes owing to impaired PIC formation at the promoters of critical differentiation genes [47]. These findings highlight the roles of TFIID components in regulating stemness and cell differentiation—key processes in embryonic development.

Homozygous KO of *taf1* in zebrafish is embryonically lethal and affects neurological, musculoskeletal and cardiac development. Null embryos have upregulation of cardiac and muscle cell differentiation as shown by RNA-seq analysis [40]. To understand the role of TAF1 in development including in the brain, *taf1* gene expression was knocked down in zebrafish by injection of the embryo at three days post fertilization with a morpholino to *taf1* [3]. Similar to full KO, *taf1* KD was embryonically lethal, and both full KO and KD fish had neurodevelopmental defects and reduced area of the optic tectum (midbrain) indicating microencephaly [3,40]. In the *taf1* KD zebrafish, overexpression of wild-type(WT) human *TAF1* mRNA restored the area of the optic tectum to WT levels, and overexpression alone did not induce a phenotype that was significantly different from that of controls [3]. Furthermore, in

neonatal mouse ventricular cardiomyocytes (immature cardiac tissue), chemical TAF1 bromodomain inhibition led to transcriptional changes in genes regulating mitochondrial function, cell proliferation, cell senescence, apoptosis, cell polarity, cell differentiation and congenital malformations [49]. These findings indicate that TAF1 is essential to embryo viability and has an important function in growth that may involve the regulation of stemness.

Gene editing of *Taf1* at postnatal day three in rats using CRISPR/Cas9 technology produced behavioural and motor defects at neonatal and juvenile periods of development (days 21, 23 and 35 after birth) with histopathological effects in the cerebellum and cerebral cortex [50,51]. The abnormal neuronal morphology and Purkinje cell apoptosis were found to be owing to decreased Cav3.1 T-type channel expression, and neuropathology and behavioural defects in *Taf1*-edited rats were rescued by induction of the T-type calcium channel enhancer SAK3 (injected at postnatal day 21) [50,52]. Similarly, early-stage AAV-miRNA-induced *Taf1* KD in mice and in rats produced motor deficits, with a stronger phenotype from *Taf1* KD at postnatal stage P0 compared with at three weeks after birth [38], highlighting a role of TAF1 in early development. These data indicate that complete loss of function *TAF1* mutations is embryonically lethal, whereas less detrimental missense or KD mutations may allow survival with neurodevelopmental and cardiac defects.

It is possible that different TAF1 protein isoforms play specific roles at different stages of development. For example, in mice, nTAF1 has higher expression postpartum than in early embryonic development, whereas *Taf1* mRNA variants without microexon 34' are most highly expressed during early development [37]. Moreover, nTAF1 showed higher relative expression in brain tissue compared with neuroblastoma [37] and microexon 34' incorporation into mRNAs is reported to be highest in post-mitotic neurons [36], suggesting that nTAF1 plays a greater role in differentiated neurons than in proliferating cells.

Finally, we have recently submitted a report on the generation and characterization of the first conditional *Taf1* loss of function model in mice [53]. Similar to zebrafish KO/KD models, *Taf1* KO was embryonically lethal, producing no hemizygous males or homozygous females. Furthermore, *Taf1* was essential for development at an early embryonic stage, as blastocysts with complete *Taf1* KO were able to survive to stage E3.5 but not E9.5. Intriguingly, female (carrier) mice heterozygous for the *Taf1* KO allele were produced with equal *Taf1* expression to WT female mice but showed behavioural deficits and increased body weight as adults (12 months old). These findings indicate that heterozygous *Taf1* KO may have skewed X-inactivation and that possibly having one functional copy of the gene may affect brain function, even though gene expression appears to be maintained overall; perhaps this is owing to a subset of neurons or cell types affected by *Taf1* deficiency.

2.2. Human *TAF1* mutations and development

In humans, males with congenital hemizygous mutations in *TAF1* (maternally inherited or *de novo*) show cardinal facial dysmorphologies including microencephaly, with intellectual disability from a young age [3,40]. Comparable microencephaly and intellectual disability have been reported for hypomorphic mutations in other canonical TFIID components such as TAF2 (8q24.12) [54] and TAF13 (1p13.3) [55], which are both inherited in an autosomal-recessive manner. Furthermore, missense variants in *TAF1* cause congenital heart disease [56], indicating that TAF1 function is required for cardiac development.

The severity of neurodevelopmental effects from mutations in *TAF1* could depend on the location of the mutation in the gene, and thus the protein function affected, but missense mutations occur throughout the protein sequence in people with XLID (or suspected XLID) and no single common mutation or domain is affected. However, fewer mutations are reported at the C-terminal than at the central portions of the protein, where none have been reported to occur after the second bromodomain (BrD) (figure 5; [3,40,57–62]). The reason that no affected individuals with C-terminal mutations have been described may be because missense mutations towards the end of the gene might not affect overall protein function and would not result in a clinical phenotype. In line with this, the region C-terminal to the BrDs lacks both structure and evolutionary conservation [2]. Alternatively, this portion of TAF1 may be essential to its function and the viability of embryos, and thus few living patients with such mutations have been reported owing to early lethality. Further mutational analyses are required to demonstrate the effect of *TAF1* mutations on each protein domain in humans.

The predicted regions of the TAF1 protein affected by congenital *TAF1* mutations in XLID patients are shown in figure 5. For example, O'Rawe *et al.* describe eight missense *TAF1* variants leading to XLID, four of which change conserved residues reported as regions important for TAF7 binding [3].

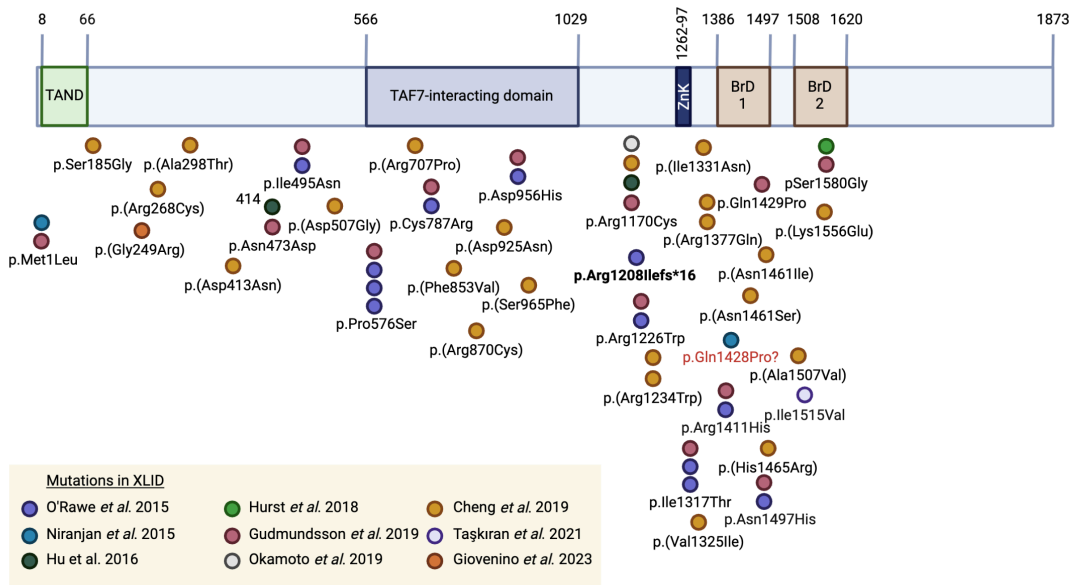


Figure 5. TAF1 protein-coding variants in human XLID. Amino acids (aa) in TAF1 are numbered marking the start and end of each major protein domain. Congenital mutations in *TAF1* identified in XLID are shown, where aa locations in relation to the protein domains are indicated; colour indicates the source reference. Of note, mutations related to other phenotypes such as congenital heart disease are not shown. Protein domains are scaled to align with corresponding exons, where aa numbers are defined by the most updated accession reference for the human TAF1 protein on NCBI (NP_004597.3). The only reported frameshift mutation is shown in bold text. One reported mutation does not match with the aa in the protein sequence and is shown in red text (p.Gln1428Pro). Of note, only male patients are shown for consistency with phenotypic effects of XLID. TAND, TAF1 N-terminal domain; TAF7, TATA-Box binding protein associated factor 7; Zn, Zinc knuckle domain; BrD, bromodomain. Created with [BioRender.com](https://www.biorender.com).

Three separate substitutions (p.Cys787Arg, p.Pro576Ser or p.Asp956His) in different families occurred within the central TAF7-interacting domain (amino acids 566–1029). For example, the p.Cys787Arg substitution, where the polar cysteine residue is replaced with the large basic arginine, is predicted to destabilize the heterotrimeric triple-barrel-fold comprising TAF1 and TAF7, altering the interaction between these two proteins. Similarly, p.Asp956His leads to an acidic-to-basic residue change, disrupting the glycine-rich motif and probably altering TAF7 binding. Interestingly, Cheng *et al.* report the difference in Gibbs free energy change for XLID-associated mutations in the TAF1-TAF7 binding domain compared with the WT molecule, where many mutations such as p.Gly680Asp (reported in one female only), p.Phe853Val and p.Arg870Cys showed no differences, whereas p.Arg707Pro shows a positive change indicating that it destabilizes the protein structure [58]. Gudmundsson *et al.* reported that the pArg1170Cys (c3568C>T) variant is probably disease-causing and deleterious owing to its effects on a conserved base, which has been reported in four separate XLID studies [40,58,60,61].

Cheng *et al.* described variants in the TAF1 double BrD, and a large proportion of mutations have been located here by multiple XLID studies [3,40,57,58,60]. However, generally missense mutations in this region do not appear to change the structural stability of TAF1, and, therefore, whether these mutations produce *TAF1* loss of function effects and how they impact development and produce clinical phenotypes is difficult to predict [58]. Details of other XLID-associated *TAF1* mutations and their predicted effects on protein structure and function are described in O'Rawe *et al.* [3] and Cheng *et al.* [58].

Finally, upon screening patients with missense mutations in *TAF1* who have congenital heart disease and neurodevelopmental defects, the most deleterious missense variants usually occurred within the functional domains of TAF1, including the N-terminal domain, TAF7-interacting domain, DNA-binding domain and BrD (in 12 of 16 cases) [49]. Interestingly, similarly to XLID patients, no *TAF1* mutations related to congenital heart disease were reported in the gene after exon 31 (p.Lys1576Glu), indicating that mutations at the C-terminal end were not found in patients with this phenotype [56]. In such instances where TAF1 protein function and protein-protein interactions are altered, gene expression during development is likely to be perturbed to produce XLID and cardiac phenotypes.

3. TAF1 in neurodegeneration

Defects in RNA pol II-mediated transcription can lead to multiple types of motor disorders and other features and these diseases have been called ‘transcription syndromes’ [63]. For example, spinocerebellar ataxia 17 is an inherited progressive movement disorder primarily affecting the cerebellum, which is caused by an expanded polyglutamine tract in TBP [64], altering its interaction with TFIID and dysregulating expression of its target genes [65,66]. Mutations causative for one form of familial amyotrophic lateral sclerosis (ALS)—an adult-onset neurodegenerative disorder leading to loss of upper and lower motor neurons—lie within the RNA/DNA-binding protein fused in sarcoma/translocated in liposarcoma (FUS/TLS) [67,68], which can interact with TBP and may influence RNA polymerase II and the TFIID complex assembly, among other roles in transcription [69,70]. Of note, TBP is also known to regulate pol I- and III-mediated transcription [28]. Thus, we cannot say for certain that these motor disorders are directly caused by TFIID defects. Many defects in TFIID and other components of the basal transcription machinery result in neurodegeneration, but these are beyond the scope of this review as we focus here on TAF1.

Neurodegenerative diseases implicating TAF1 dysfunction in their pathogenesis predominantly exhibit atrophy of the striatum or other regions of the basal ganglia with motor defects. These include, but may not be limited to, XDP, HD and Parkinson’s disease (PD).

3.1. X-linked dystonia–parkinsonism

XDP is an adult-onset neurodegenerative disorder affecting individuals with ancestry from the Philippines. XDP predominantly affects males owing to its X-linked mode of inheritance and it arises from a presumed founder mutation with a prevalence of 0.57 per 100 000 on the island of Panay, Philippines [71]. In most patients, XDP is characterized initially by dystonia (93% of cases) that generalizes within 5–10 years from the age of onset in males, which eventually transitions into parkinsonism with a mean age of death of 55.6 years (range: 33–81) [71,72]. Relatively few XDP female cases (approx. 14 symptomatic to date) have been reported with similar XDP features to affected males but with a later age of onset (males: approx. 39 years (range: 12–64 years) versus females: approx. 52 years (range: 26–75 years)) [8,71,73,74]. Similar to other major movement disorders, the basal ganglia are affected in XDP. In the striatum, medium spiny neurons are preferentially lost from the caudate nucleus and putamen [7,75] in a similar manner to that shown in HD [76] and cerebral ischaemia [77]. Additionally, it has been reported that neural progenitor cells are lost from the subventricular zone [75]. Thus, although XDP is a rare inherited form of dystonia and parkinsonism, it may share a similar neuropathological mechanism to other late-onset neurodegenerative conditions.

In 2003, Nolte *et al.* first identified a 300 kb interval of Xq13.1 associated with XDP, described as the MTS [31]. This region encoded eight genes, including *TAF1*, and sequencing within the affected haplotype showed five single-nucleotide changes and a 48 bp deletion [31]. Further analysis identified an approximately 2.6 kb SINE-VNTR-Alu (SVA) type F retrotransposon insertion in intron 32 of *TAF1* [6]. Of these seven sequence variants, three fall within *TAF1* introns, while the remaining four are localized to an intergenic region containing the MTS 3’ to *TAF1*. Domingo *et al.* narrowed the disease locus in the XDP haplotype to a 294 kb region that included four genes: *TAF1*, *OGT*, *ACRC* and *CXCR3* [78]. Subsequently, Aneichyk *et al.* revealed 47 novel variants that segregated with XDP, which were confined to a narrower minimal critical region of 203.6 kb, encompassing only *TAF1* [4].

The SVA insertion in intron 32 of *TAF1* appears to be causative for XDP. This is primarily shown by the inverse correlation between the number of hexameric repeats ((CCCTCT)_n, n = 30–55 repeats) within the SVA and both the age of disease onset [5,79] and with *TAF1* expression [79]. Repeat expansion in somatic tissues has been shown to be repeat length-dependent and may be tissue-specific, with greater levels of expansion in the brain than in blood [80]. Differences in hexameric repeat numbers between brain regions may also play a role in selective neurodegeneration [81].

The SVA insertion in *TAF1* in XDP leads to aberrant splicing of *TAF1* isoforms ([5,6]; figure 6). Early analysis of an XDP brain showed reduced transcript levels of n*TAF1* as detected by Taqman probes [6,82]. However, a recent study indicated that the presence of the SVA may not be causal for reduced microexon 34’ incorporation into *TAF1* transcripts in XDP, since levels of microexon 34’-containing transcripts were no different between control and XDP brains [30]. Nonetheless, mice injected postnatally (P0) with miRNA knocking down n*Taf1* expression showed motor defects at two and four months of age compared with age-matched controls, with similar but fewer effects when the same experiments were performed in rats. No increase in deficits was seen in a model knocking down both

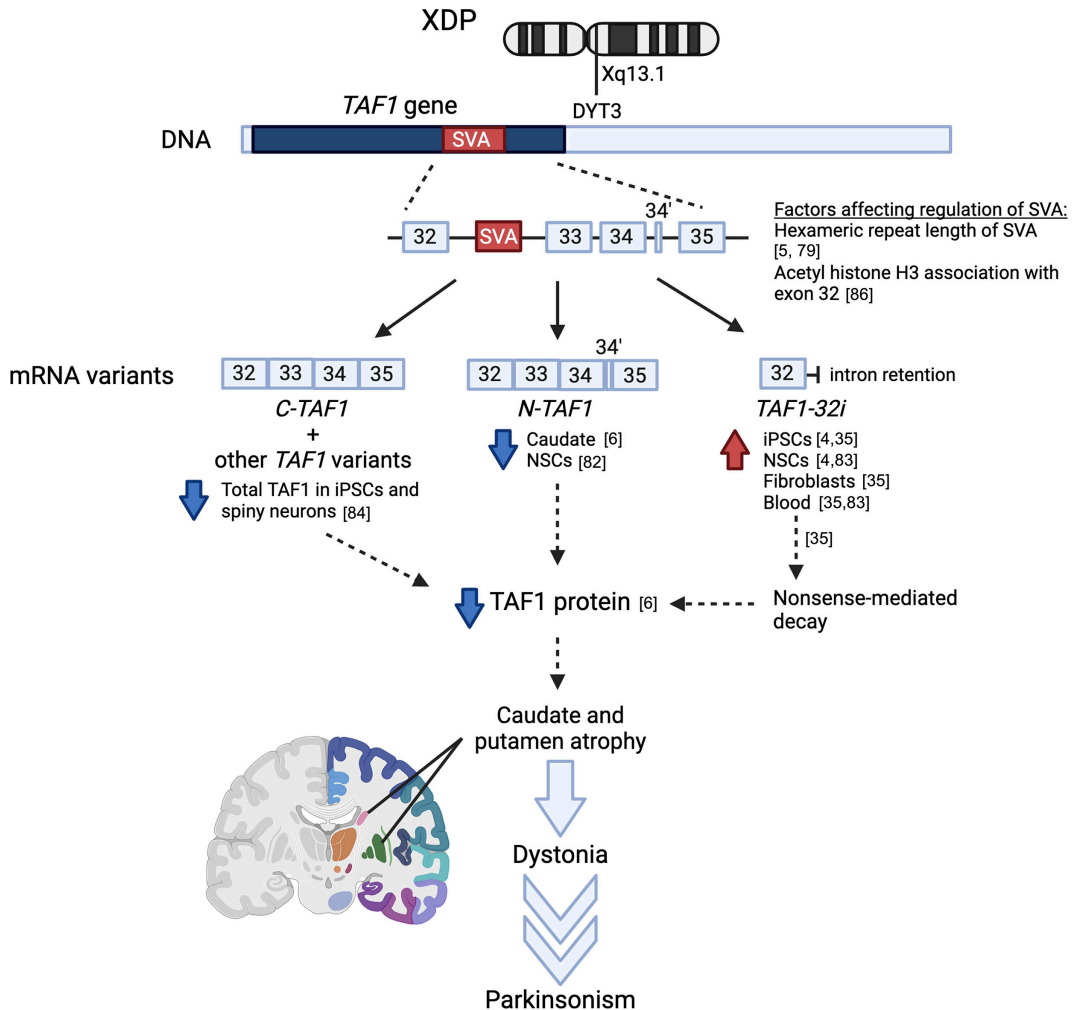


Figure 6. Evidence linking *TAF1* to XDP pathogenesis. The region of Xq13.1 affected in the XDP allele is at the MTS, in which *TAF1* lies. *TAF1* from XDP patients or XDP carriers contains a novel SVA within intron 32, and *cTAF1*, *nTAF1* and *TAF1-32i* show differential transcript levels in XDP brain compared with that without XDP. Changes to transcript levels may affect *TAF1* protein levels in yet unknown ways (dotted arrow), including NMD of the transcript, preventing synthesis of *TAF1* proteins. Other mechanisms of differential *TAF1* expression may be owing to the length of the hexameric repeat in the SVA and the association of acetyl histone H3 with exon 32. Nonetheless, eventually the caudate nucleus and putamen regions of the neostriatum show atrophy, which is thought to be linked to the onset of dystonia and parkinsonism observed in XDP patients. Figure (introns and exons) not drawn to scale. Created with BioRender.com.

cTaf1 and *nTaf1*, indicating that *cTaf1* plays a smaller role in this phenotype, with milder effects in rats injected at later stages (three weeks old), alongside cholinergic striatal dysregulation [38].

Conversely, a novel transcript named *TAF1-32i* identified by Aneichyk *et al.* is elevated in XDP owing to the activation of a cryptic exon (32i) in intron 32, which lies just 5' of the SVA insertion [4,83]. Misincorporation of this intronic exon causes disruption of the ORF and premature translation termination leading to NMD of the transcript [35]. Finally, the SVA has been proposed to physically impede RNA pol II elongation by forming G-quadruplexes [5], which may interfere with transcriptional elongation or cause premature termination that could affect the expression of *TAF1* variants and splicing particularly of intron 32 [4]. However, owing to the lack of XDP patient samples and ethnically matched control subjects, other pathological transcript variants of *TAF1* may have not been identified yet and they could contribute to XDP.

Although XDP is not thought to be owing to a simple loss of function of *TAF1*, reduced total *TAF1* at the mRNA [84,85] and protein [6] levels has also been reported in XDP. However, this is not consistent across studies, tissues, brain regions and cell types, and possibly changes with disease duration [4,84,85]. Nonetheless, excision of the SVA has been shown to rescue *TAF1* expression, intron retention and aberrant splicing [4,84] in human induced pluripotent stem cells (iPSCs) and NSCs. These studies indicate the XDP-SVA modifies the expression of *TAF1* variants, although owing to small

patient numbers and the limitations with *post-mortem* samples, it has been a challenge to link variant expression with disease onset in a temporal manner.

In addition to gene expression changes, the effects of the SVA could be related to the chromatin state of the *TAF1* gene. For example, acetylation of histone H3 (AcH3) was reported to be modified in XDP-derived fibroblasts, such as increased AcH3 association with *TAF1* exon 32 which is upstream to the SVA, as well as downstream exon 38, and increased AcH3 association was reported at exon 17 [86]. Nonetheless, AcH3 association at exon 32 could be normalized by CRISPR/Cas9-mediated excision of the SVA [86]. However, while *TAF1* loss of function may partially explain the XDP phenotype, how dysregulation of the chromatin state of the *TAF1* locus would cause XDP pathogenesis remains to be elucidated. Interestingly, alterations in histone acetyltransferase and histone deacetylase levels and activity have been reported in progressive neurodegenerative conditions in human patients and animal models of HD [87], Alzheimer's disease [88], PD [89], ALS [90] and spinal muscular atrophy [91].

3.2. Huntington's disease

Similar to XDP, HD is a monogenic basal ganglia disorder with the first signs of pathology in the striatum [92]. Motor symptoms of HD include chorea as well as dystonia; chorea is rarely present in XDP [71]. HD is caused by a polyglutamine (polyQ) tract expansion of a CAG repeat of the huntingtin (*HTT*) gene [93]. Expansion of the HD CAG repeat is negatively correlated with age of disease onset and there is strong genetic evidence to suggest that the somatic expansion of the *HTT* CAG repeat specifically drives the timing of HD onset [94,95]. There is some overlap in genes (e.g. *MSH3*, *PMS2*) that have been implicated as genetic modifiers of both HD [95] and XDP [96]. These modifiers are implicated in repeat instability indicating a possible mechanistic link between XDP and HD. The interaction of mutant huntingtin with SP1 and TAF4 reduces the expression of dopamine receptor D2 (*DRD2*) [97], a gene that may also be downregulated in XDP [6]. Interestingly, a recent paper has shown that TAF1 expression is downregulated in the striatum of HD brains [9]. These findings indicate that both repeat expansion and dysregulation of TAF1 may have common functional features with HD, which connect these two neostriatal neurodegenerative conditions. Additionally, TAF7, which is a direct interactor of TAF1 within TFIID, is differentially expressed in HD patients [98]. However, whether TAF1 function is dysregulated in HD has not yet been determined. Additionally, while HD pathology is caused by protein aggregation of mutant HTT [99], whether aggregation of protein isoforms of TAF1 or other proteins in XDP occurs is not known.

3.3. Parkinson's disease

PD is a movement disorder, in which degeneration of nigral dopaminergic neurons occurs, leading to widespread neurodegeneration and diffuse Lewy body deposition [100,101]. Unlike XDP and HD, PD is not usually an inherited disorder, although genetic forms of the disease exist [102]. Degeneration of substantia nigra, which is well known to occur in PD, has not been reported at the Parkinsonian stage of XDP [7]. However, XDP, HD and PD are neurodegenerative conditions sharing a similar disease manifestation and heterogeneous age of onset most frequently in middle age. A recent study identified a rare frameshift mutation (c.29_53dupGGA(CAG)₂CTAC-CATCA(CTG)₂C; p.A19Dfs*50) in *TAF1* in two male patients from unrelated families with PD in a Chinese population [103]. Whether similar mutations in *TAF1* exist or splice variant abundance changes in other populations with PD has not been apparent so far. Currently, there is limited evidence for differential regulation of *TAF1* in PD.

4. TAF1 in other diseases

Diseases of TAF1 dysfunction are largely but not exclusively limited to the brain and heart and, therefore, *TAF1* expression is mostly investigated in these tissues. However, *TAF1* expression is also clearly required for early growth and development, potentially within different stem cells. Thus, TAF1 may be differentially expressed in clinical conditions such as cancer and there may be cell type-specific roles of TAF1 and its isoforms.

4.1. Cancer

Multiple TAFs have been shown to be mutated or dysregulated in cancer, including TAF1, TAF2, TAF4/TAF4B, TAF6, TAF9, TAF10 and TAF12 [104]. TAF1 plays key roles in processes that are tumorigenic when disrupted, such as regulation of the cell cycle. For example, it has been reported that TAF1 phosphorylates P53 (at Thr55) *in vivo* to induce G1 progression [41]. TAF1 was also linked to phosphorylation of the protooncogene MDM2 [105], leading to the degradation of P53 [41]. Downregulation of TAF1 activity leads to P53 phosphorylation (at Ser15) and activation of a DNA damage response and cell cycle arrest [106]. Thus, the transcriptional machinery has been linked with cell cycle regulation through the DNA damage response, in which TAF1 is thought to play a key role in the progression of the G1 phase of the cell cycle.

A characteristic of cancer cells is an evasion of apoptosis and TAF1 has been linked to the regulation of oxidative and genotoxic stress-induced apoptosis through control of p27^{Kip1} expression [107] indicating that TAF1 inhibition may be oncogenic in some cases. Overall, the majority of studies, albeit limited, indicate that TAF1 has a pro-tumorigenic rather than a tumour-suppressive function in cancer.

Findings from clinical data are consistent with the role(s) of TAF1 in tumorigenesis. For example, TFIID has been shown to be overexpressed in human lung and breast carcinomas [108]. Frameshift mutations have been found in *TAF1* in colorectal cancer (3.8%) with high microsatellite instability in the Korean population [109]. Mutations acquired following neoadjuvant chemotherapy are known as ‘cancer driver genes’, which are absent from paired pre-treatment samples. In oesophageal adenocarcinoma, these included a mutation in *TAF1* (c.1237G>C/p.Asp413His) [110]. TAF1 is implicated in prostate cancer pathogenesis owing to its regulation of the androgen receptor (AR), which represents a sterol receptor that requires multiple components of the transcription machinery to regulate its target genes in the prostate. TAF1 was found to increase with the duration of androgen withdrawal in patient samples indicating that it plays a role in castration-resistant prostate cancer [111]. TAF1 binds to the AR resulting in enhancement of AR transcriptional activity [111]. In uterine serous carcinoma, seven *TAF1* mutations were identified in unmatched tumours, most of which lay in the TAF7-interacting region, but the functional outcome of these mutations was not investigated [112]. Thus, TAF1 is likely to play roles in a variety of cancers although mechanisms are yet to be elucidated. Furthermore, TAF1 activity may affect tumorigenesis in a cell type- and tissue-specific manner, which is yet to be defined.

5. Knowns and unknowns with TAF1 activity

While many of the genetic and protein functional consequences of TAF1 expression and dysfunction have been reported as described above, there is still much to uncover regarding the relationship between TAF1 and disease. Since *TAF1* is located on the X chromosome, X-linked diseases in males have been described, whereas there are fewer studies on carrier females of *TAF1* mutations and XDP. Is TAF1 function compensated in heterozygous females? Can the testis-specific paralogue of *TAF1*, *TAF1L* (encoding TAF1L; TBP-associated factor 1 like; TAF(II)210), substitute for its function in some cases despite its restricted expression pattern under normal circumstances? Moreover, *TAF1* expression in different cell and tissue types has not been thoroughly investigated. Current knowledge and implications of TAF1 regulation in different conditions are discussed in the following sections.

5.1. X-inactivation in carrier females of TAF1 diseases

In both XLID and XDP, the primary affected individuals are hemizygous males, owing to the X-linked recessive nature of these rare conditions. The majority of the few female cases of XDP reported were heterozygous for the XDP allele and only one homozygous female has been reported to date ([74]; table 2). Most XDP carrier females do not have severe symptoms, possibly owing to skewed X-inactivation of the mutated allele, and many mild cases were only identified owing to genetic testing of female relatives of affected males. However, some heterozygous XDP females are highly affected with predominantly parkinsonism symptoms. Furthermore, it has been reported that X-inactivation can be skewed in favour of the XDP allele in some cases [74], where heterozygous females can show a similar phenotype to that of hemizygous males [114]. However, many more unreported cases of XDP may exist in females with mild symptoms and it has even been proposed that XDP may contribute to the apparently relatively higher proportion of PD in females in the Philippines (1 : 1.07 males :

Table 2. Females with mutations in the XDP disease-specific region or the *TAF1* gene and their associated phenotypes. (For XDP, patients are listed in order of age of onset, whereas patients with congenital *TAF1* mutations are listed in order of genetic changes (location of mutation along the amino acid sequence). For XDP, most cases are heterozygous (or presumed if not stated), whereas there is one case with homozygous XDP alleles as stated. For the cases with congenital *TAF1* mutations, most mutations are missense mutations (described), whereas there is one case of frameshift mutation as stated.)

predicted genotype	genetic changes/X-inactivation status	clinical presentation	age of onset (years)	affected family members	reference
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	mild chorea	26 y	mother, maternal uncle, aunt, great-grandfather	[74]
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: upper limb action and postural tremor, mild breakdown of limb rapid alternating movements (RAMs), mild retropulsion	35 y	1 brother	[74]
XDP	no data	leg cramps, face dystonia, generalized in 2 years	37 y	1 father, 2 brothers	[71]
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	perioral tremor, mild impairment of tandem gait	42 y	mother, 3 brothers	[74]
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: breakdown of limb RAMs, shuffling gait, cervical dystonia	42 y	mother, 4 brothers	[74]
XDP	no data	leg dystonia, generalized within 4 years	47 y	1 sister	[71]
XDP	no data	slurred speech, gait problems, then dystonia with parkinsonism	48 y	uncles on both sides, brothers	[71]
XDP	no data	leg dystonia, generalized within 5 years, parkinsonism features after 16 years with dystonia	49 y	1 sister	[71]
heterozygous XDP	missense mutation DSC3 change (C>T) X-chromosome monosomy in a subset of cells	abnormal putamen and caudate, leg and hand dystonia, then parkinsonism with some dystonia Turner syndrome (short stature, drooping ears)	50 y	3 brothers, maternal grandfather, mother (?)	[71,73]

(Continued.)

Table 2. (Continued.)

predicted genotype	genetic changes/X-inactivation status	clinical presentation	age of onset (years)	affected family members	reference
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: shuffling gait, breakdown of limb RAMs, mild retropulsion leg chorea that generalized	51 y	maternal grandfather, 1 brother, 1 sister, 1 daughter	[113]
heterozygous XDP	5 disease-specific single-nucleotide changes 1 48 bp deletion 1 SVA retrotransposon insertion skewed X-inactivation (98:2%), only C>T expressed (no other changes)	parkinson symptoms caudate atrophy	57 y		[114]
XDP	no data	left foot dystonia	59 y		[113]
homozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: stooped posture, breakdown of RAMs of limbs, shuffling gait, retropulsion chorea in arm	72 y	2 sons	[74]
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: shuffling gait, breakdown of limb RAMs, hypomimia, stooped posture, micrographia, absent arm swing bilaterally, tremor effects	75 y	1 daughter, 3 sons	[74]
heterozygous XDP	3 XDP-specific mutant alleles at DXS8030, DXS8101, and DXS559	parkinsonism: breakdown of RAMs, shuffling gait	75 y	1 daughter, 4 sons	[74]
heterozygous XDP	disease-specific haplotype SVA retrotransposon insertions	no neurological defects	NA	1 brother	[115]
heterozygous XDP	disease-specific haplotype SVA retrotransposon insertions	no neurological defects	NA	1 brother	[115]
heterozygous <i>TAF1</i> mut	missense mutation (c.745G > A; p.Gly249Arg) skewed X-inactivation (95:5)	global developmental delay, delayed psychomotor development	12 mo	1 brother, mother, paternal grandmother with unidentified psychiatric pathology	[62]

(Continued.)

Table 2. (Continued.)

predicted genotype	genetic changes/X-inactivation status	clinical presentation	age of onset (years)	affected family members	reference
heterozygous <i>TAF1</i> mut	missense mutation (c.1786C > T; p.Pro596Ser)	no data	NA	3 sons	[3]
heterozygous <i>de novo TAF1</i> mut	missense mutation (c.2039G > A p.Gly680Asp) skewed X-inactivation (> 90:10%)	dysmorphism in the face, hands, feet and neck, global developmental delay cardiovascular issues (heart murmur, pulmonary artery atresia and hypoplasia)	birth	none	[58]
<i>de novo TAF1</i> mut	missense mutation (c.2039G > A; p.Gly680Asp)	congenital heart disease (pulmonary artery atresia)		none	[56]
<i>TAF1</i> mut	missense mutation (g.23335G > T; exon18: c.2774G > A;p.G925D) assumed <i>de novo</i> (not confirmed)	at birth: jaundice, microcephaly, gastroesophageal reflux, hypotonia and congenital cardiopathy. during childhood: delayed developmental milestones, fine motor incoordination, delayed speech, moderate intellectual disability. Asthma, pneumonia dysmorphic features	birth	none confirmed	[116]
<i>de novo TAF1</i> mut	missense mutation (c.2933C > T; p.Thr978Met)	congenital heart disease (Ebstein, ventricular septal defect (VSD), hypoplastic aortic arch)		none	[56]
heterozygous <i>de novo TAF1</i> mut	missense mutation (c.3035C > T p.Thr1012Ile) skewed X-inactivation	learning disability, natal oral cleft, ventricular septal defect, swirling pigmentary disturbances, postnatal growth retardation, global developmental delay, generalized hypotonia, hypertelorism and other facial	birth	none	[58]

(Continued.)

Table 2. (Continued.)

predicted genotype	genetic changes/X-inactivation status	clinical presentation	age of onset (years)	affected family members	reference
		developmental abnormalities, sensorineural hearing impairment			
<i>de novo TAF1</i> mut	missense mutation (c.3035C > T; pThr1012Ile)	congenital heart disease (VSD, PA stenosis)		none	[56]
heterozygous <i>TAF1</i> mut	missense mutation (c.3568C > T; p.Arg1190Cys)	no symptoms	NA	2 sons	[117]
heterozygous <i>TAF1</i> mut	frameshift mutation (c.3708A > G; p.(arg1228llefs*16) alternative transcript (splice site) with 28 bp deletion, resulting in a frameshift and a premature stop codon	no symptoms	NA	1 son	[3]
heterozygous <i>TAF1</i> mut	missense mutation (c.4010T>C) skewed X-inactivation (99:1%)	no symptoms (healthy)	NA	2 sons	[3]

females) compared with globally (1 : 0.52 males : females) [114,118]. However, one cannot rule out ethnicity-specific effects at this moment [118].

Common features among affected females with non-XDP *TAF1* mutations include intellectual disability and facial abnormalities [58], as well as congenital heart disease in a similar manner to their male counterparts ([56,116]; table 2). However, the proportion of affected versus healthy females with *TAF1* mutations is difficult to estimate because many female cases without symptoms, or with mild presentations, may have escaped identification. In females with *TAF1* mutations, missense mutations are most common, although a frameshift mutation in *TAF1* has also been identified [3]. As in XDP, skewed X-inactivation of *TAF1* has been reported in females carrying *TAF1* mutations, which presented with a spectrum of clinical symptoms or were asymptomatic, while they had affected sons [3,57,58]. Thus, X-inactivation status in heterozygous female carriers of XDP or other *TAF1* mutations may contribute to the highly heterogeneous clinical features.

5.2. Functional compensation of TAF1 activity

In conditions of reduced *TAF1* expression, it is intriguing to speculate that a paralogous gene may compensate for *TAF1* activity. *TAF1L* is a human paralogue of *TAF1* that is located on chromosome 9 at an intronless locus (9p21.1) (Ensembl: ENSG00000122728). *TAF1L* probably arose from a retrotransposon gene insertion and can be translated into an intact protein (Uniprot: 1826 aa, Q8IZX4; Antonova S, Timmers HTM, unpublished results [2024]), which shows approximately 95% amino acid homology with *TAF1* [119]. *TAF1L* has just one transcript (6216 bp) and is composed of the 38 canonical *TAF1* exons including exon 35' (excluding exons d₁– d₅) [32].

TAF1L has been proposed to compensate for *TAF1* in spermatocytes, where *TAF1L* is exclusively expressed [119]. In these meiotic cells, autosomes are transcriptionally active, but X and Y chromosomes are silenced owing to their condensation into heterochromatin in a transcriptionally inactive compartment. Therefore, autosomal genes may functionally substitute activities of genes on the sex chromosomes during this period [120], such as initiation of RNA pol II-dependent transcription by *TAF1L*-containing TFIID in spermatocytes replacing *TAF1*. *TAF1L* is also dysregulated in a number

of cancers, indicating that it may have a similar function to TAF1 in the regulation of tumorigenesis [109,121].

5.3. Tissue-specific roles of TAF1

As highlighted above, *TAF1* and its isoforms may be differentially expressed between cell types and tissues. TAFs, including TAF1, have activator-dependent transcription functions, which are both promoter- and tissue-specific [122]. For example, while TAF4B is expressed in many different tissues, deletion of this gene results in viable but infertile offspring owing to its requirement for folliculogenesis [123]. Additionally, *TAF1* may have a specific role in cardiac function, where loss of function causes congenital heart disease, owing to the epigenetic modulatory role of its double BrD in regulating foetal myosins [49]. In this manner, TAF1 could potentially be involved in the repression of the atrial-specific SMYHC3 promoter leading to a cardiac-specific effect of its variants, or alternatively just results as a secondary tissue to be affected by the TAF1 dysfunction.

Cell type-specific differences in TAF1 expression could potentially differentially influence tissues affected by pathology such as the brain and specific stages of development and in neurodegenerative states. For example, the promoter of the sense transcript of *TAF1*-MTS exons d_2 , d_3 and d_4 has been described as under the control of an *Ikaros* element, which is specifically expressed in the striatum of the developing brain [32]. Furthermore, the most commonly known driver of microexon inclusion (e.g. 34') is neuron-specific splicing factor SRRM4, which incorporates brain-specific microexons into *TAF1* mRNAs [36,124]. Finally, the *TAF1*-32i variant, that is elevated in XDP, is primarily in dividing cells rather than differentiated neurons [4], indicating a role for this variant in specific cell types. For example, *TAF1*-32i was higher in XDP fibroblasts, iPSCs and NSCs, with low levels in cortical neurons, GABAergic neurons, and NSC-derived neurons compared with control cells [4]. Higher *TAF1*-32i levels were also reported in XDP fibroblasts and iPSCs compared with blood [35], indicating tissue-specific levels of this transcript.

Whether TAF1 expression, splicing and isoform abundance changes with ageing is not known yet. Diseases of TAF1 dysfunction such as XDP and XLID predominantly affect the brain, and postnatal depletion of *Taf1* in rodents takes several months to induce a phenotype of neurocognitive decline [38] indicating that TAF1 dysfunction plays a larger role with increasing age or that symptoms have delayed manifestation. Elucidation of the age-specific expression of TAF1 will be key to targeting TAF1-induced diseases.

6. Conclusions and further remarks

In conclusion, TAF1 has multiple functions governed by its activity and interacting partners that ultimately regulate RNA pol II-dependent transcription of many genes. *TAF1* is a key gene for development and complete deletion results in embryonic lethality in zebrafish and probably humans and other animals (https://gnomad.broadinstitute.org/gene/ENSG00000147133?dataset=gnomad_r4). Congenital hypomorphic mutations in *TAF1* in humans result in aberrant development of the brain with corresponding intellectual disability and can also cause congenital heart disease. *TAF1* dysregulation in XDP results in striatal neurodegeneration and consequential loss of motility in middle age, similar to that in HD and PD. However, while much is known about the *TAF1* gene and its protein structure and function, and diseases involving TAF1 dysregulation have been identified, there is still much to be uncovered to mechanistically connect this gene with disease causation.

We have provided various examples of how *TAF1* variant expression may be linked to the neurodegenerative condition XDP, such as increased levels of *TAF1*-32i. However, the cell type-specificity of these variants and their abundance in various tissues has not been comprehensively explored. Tissue-specific splicing factors may be key to uncovering the regulation of *TAF1* transcripts and their relationship to pathogenesis, such as why the striatum is predominantly affected in XDP, with application to HD and other motor disorders.

Functional aspects of TAF1 biology may play essential roles in pathogenesis. We highlight that *TAF1* is primarily expressed in dividing rather than differentiated cells. TAF1 has also been linked to tumorigenesis, and we speculate it might be differentially regulated by cancer cells as evidenced in multiple types of cancers. This hypothesis would make sense given that the TFIID complex regulates basal transcription, which is required for cell growth. However, how different mutations affect the function of TAF1 in disease is not yet clear.

Genetic animal models are important for understanding systems biology and stages of pathogenesis, particularly for neurodegenerative and other progressive conditions. While *TAF1* KO models have been developed in organisms such as for zebrafish, embryonic lethality in these studies highlights the key role of TAF1 function in development but these models do not allow the investigation of *TAF1* loss of function effects in full KOs. Additionally, rodent studies with postnatal *Taf1* KD support the role of *TAF1* loss of function in motor decline, while the relevant molecular and cellular mechanisms remain to be elucidated. Modelling of XDP is a challenge because the sequence of intron 32 is not conserved between mice and humans. Thus, humanizing the mouse *Taf1* gene to introduce the SVA into mice will be necessary to effectively model this disease in animals. In the last 10 years, the body of research describing the role of *TAF1* in XDP has increased exponentially, but new techniques will be required to elucidate the mechanistic role of this gene in this complex disease.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. This article has no additional data.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. E.M.C.: conceptualization, data curation, formal analysis, investigation, validation, visualization, writing—original draft, writing—review and editing; K.C.: conceptualization, writing—review and editing; H.T.M.T.: funding acquisition, investigation, resources, writing—review and editing; E.M.C.F.: conceptualization, funding acquisition, project administration, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The research of H.T.M.T. was supported by the Deutsche Forschungsgemeinschaft (DFG) through projects 192904750-SFB 992 and TI688/I-1, and by the Collaborative Center for X-linked Dystonia–Parkinsonism, Massachusetts General Hospital (no. 239295). E.M.C.F., E.M.C. and K.C. are supported by the Collaborative Center for X-linked Dystonia–Parkinsonism (no. 239295 to E.M.C.F.).

References

1. Timmers HTM. 2021 SAGA and TFIIID: friends of TBP drifting apart. *Biochim. Biophys. Acta Gene Regul. Mech.* **1864**, 194604. (doi:10.1016/j.bbagr.2020.194604)
2. Antonova SV, Boeren J, Timmers HTM, Snel B. 2019 Epigenetics and transcription regulation during eukaryotic diversification: the saga of TFIIID. *Genes Dev.* **33**, 888–902. (doi:10.1101/gad.300475.117)
3. O'Rawe JA *et al.* 2015 TAF1 variants are associated with dysmorphic features. *Intellect. Disabil. Neurol. Manifest. Am. J. Hum. Genet.* **97**, 922–932. (doi:10.1016/j.ajhg.2015.11.005)
4. Aneichyk T *et al.* 2018 Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* **172**, 897–909. (doi:10.1016/j.cell.2018.02.011)
5. Bragg DC *et al.* 2017 Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc. Natl Acad. Sci. USA* **114**, E11020–E11028. (doi:10.1073/pnas.1712526114)
6. Makino S *et al.* 2007 Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am. J. Hum. Genet.* **80**, 393–406. (doi:10.1086/512129)
7. Goto S *et al.* 2005 Functional anatomy of the basal ganglia in X-linked recessive dystonia-parkinsonism. *Ann. Neurol.* **58**, 7–17. (doi:10.1002/ana.20513)
8. Lee LV *et al.* 2002 The natural history of sex-linked recessive dystonia parkinsonism of Panay, Philippines (XDP). *Parkinsonism Relat. Disord.* **9**, 29–38. (doi:10.1016/s1353-8020(02)00042-1)
9. Hernández IH, Cabrera JR, Santos-Galindo M, Sánchez-Martín M, Domínguez V, García-Escudero R, Pérez-Álvarez MJ, Pintado B, Lucas JJ. 2020 Pathogenic SREK1 decrease in huntington's disease lowers TAF1 mimicking X-linked dystonia parkinsonism. *Brain* **143**, 2207–2219. (doi:10.1093/brain/awaa150)
10. Chen X *et al.* 2021 Structural insights into preinitiation complex assembly on core promoters. *Science* **372**, eaba8490. (doi:10.1126/science.aba8490)
11. Chen X, Xu Y. 2022 Structural insights into assembly of transcription preinitiation complex. *Curr. Opin. Struct. Biol.* **75**, 102404. (doi:10.1016/j.sbi.2022.102404)
12. Louder RK, He Y, López-Blanco JR, Fang J, Chacón P, Nogales E. 2016 Structure of promoter-bound TFIIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609. (doi:10.1038/nature17394)
13. Malik S, Roeder RG. 2023 Regulation of the RNA polymerase II pre-initiation complex by its associated coactivators. *Nat. Rev. Genet.* **24**, 767–782. (doi:10.1038/s41576-023-00630-9)
14. Bernardini A *et al.* 2023 Hierarchical TAF1-dependent co-translational assembly of the basal transcription factor TFIIID. *Nat. Struct. Mol. Biol.* **30**, 1141–1152. (doi:10.1038/s41594-023-01026-3)

15. Kamenova I *et al.* 2019 Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* **10**, 1740. (doi:10.1038/s41467-019-09749-y)
16. Trowitzsch S *et al.* 2015 Cytoplasmic TAF2–TAF8–TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6**, 6011. (doi:10.1038/ncomms7011)
17. Kokubo T, Yamashita S, Horikoshi M, Roeder RG, Nakatani Y. 1994 Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding. *Proc. Natl Acad. Sci. USA* **91**, 3520–3524. (doi:10.1073/pnas.91.9.3520)
18. Wang H, Curran EC, Hinds TR, Wang EH, Zheng N. 2014 Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Res.* **24**, 1433–1444. (doi:10.1038/cr.2014.148)
19. Curran EC, Wang H, Hinds TR, Zheng N, Wang EH. 2018 Zinc knuckle of TAF1 is a DNA binding module critical for TFIID promoter occupancy. *Sci. Rep.* **8**, 4630. (doi:10.1038/s41598-018-22879-5)
20. Flynn EM, Huang OW, Poy F, Oppikofer M, Bellon SF, Tang Y, Cochran AG. 2015 A subset of human bromodomains recognizes butyryllysine and crotonyllysine histone peptide modifications. *Structure* **23**, 1801–1814. (doi:10.1016/j.str.2015.08.004)
21. Jacobson RH, Ladurner AG, King DS, Tjian R. 2000 Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422–1425. (doi:10.1126/science.288.5470.1422)
22. Singh MV, Bland CE, Weil PA. 2004 Molecular and genetic characterization of a TAF1p domain essential for yeast TFIID assembly. *Mol. Cell. Biol.* **24**, 4929–4942. (doi:10.1128/MCB.24.11.4929-4942.2004)
23. Bernardini A, Tora L. 2024 Co-translational assembly pathways of nuclear multiprotein complexes involved in the regulation of gene transcription. *J. Mol. Biol.* **436**, 168382. (doi:10.1016/j.jmb.2023.168382)
24. Chalkley GE, Verrijzer CP. 1999 DNA binding site selection by RNA polymerase II tafs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J.* **18**, 4835–4845. (doi:10.1093/emboj/18.17.4835)
25. Patel AB *et al.* 2018 Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872. (doi:10.1126/science.aau8872)
26. Liu D, Ishima R, Tong KI, Bagby S, Kokubo T, Muhandiram DR, Kay LE, Nakatani Y, Ikura M. 1998 Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**, 573–583. (doi:10.1016/s0092-8674(00)81599-8)
27. Burley SK, Roeder RG. 1998 TATA box mimicry by TFIID: autoinhibition of pol II transcription. *Cell* **94**, 551–553. (doi:10.1016/s0092-8674(00)81596-2)
28. Bhuiyan T, Timmers HTM. 2019 Promoter recognition: putting TFIID on the spot. *Trends Cell Biol.* **29**, 752–763. (doi:10.1016/j.tcb.2019.06.004)
29. Wassarman DA, Sauer F. 2001 TAF(II)250: a transcription toolbox. *J. Cell. Sci.* **114**, 2895–2902. (doi:10.1242/jcs.114.16.2895)
30. Capponi S *et al.* 2021 Dissection of TAF1 neuronal splicing and implications for neurodegeneration in X-linked dystonia-parkinsonism. *Brain Commun.* **3**, fcab253. (doi:10.1093/braincomms/fcab253)
31. Nolte D, Niemann S, Müller U. 2003 Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism. *Proc. Natl Acad. Sci. USA* **100**, 10347–10352. (doi:10.1073/pnas.1831949100)
32. Herzfeld T, Nolte D, Müller U. 2007 Structural and functional analysis of the human TAF1/DYT3 multiple transcript system. *Mamm. Genome* **18**, 787–795. (doi:10.1007/s00335-007-9063-z)
33. Herzfeld T, Nolte D, Grznarova M, Hofmann A, Schultze JL, Müller U. 2013 X-linked dystonia parkinsonism syndrome (XDP, lubag): disease-specific sequence change D5C3 in TAF1/DYT3 affects genes in vesicular transport and dopamine metabolism. *Hum. Mol. Genet.* **22**, 941–951. (doi:10.1093/hmg/dd5499)
34. Kurosaki T, Popp MW, Maquat LE. 2019 Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **20**, 406–420. (doi:10.1038/s41580-019-0126-2)
35. Pozojevic J *et al.* 2022 Transcriptional alterations in X-linked dystonia-parkinsonism caused by the SVA retrotransposon. *Int. J. Mol. Sci.* **23**, 2231. (doi:10.3390/ijms23042231)
36. Capponi S *et al.* 2020 Neuronal-specific microexon splicing of TAF1 mRNA is directly regulated by SRRM4/nSR100. *RNA Biol.* **17**, 62–74. (doi:10.1080/15476286.2019.1667214)
37. Jambalдорж J, Makino S, Munkhbat B, Tamiya G. 2012 Sustained expression of a neuron-specific isoform of the TAF1 gene in development stages and aging in mice. *Biochem. Biophys. Res. Commun.* **425**, 273–277. (doi:10.1016/j.bbrc.2012.07.081)
38. Cirnaru MD, Creus-Muncunill J, Nelson S, Lewis TB, Watson J, Ellerby LM, Gonzalez-Alegre P, Ehrlich ME. 2021 Striatal cholinergic dysregulation after neonatal decrease in X-linked dystonia parkinsonism-related TAF1 isoforms. *Mov. Disord.* **36**, 2780–2794. (doi:10.1002/mds.28750)
39. Walker AK, Shi Y, Blackwell TK. 2004 An extensive requirement for transcription factor IID-specific TAF-1 in *Caenorhabditis elegans* embryonic transcription. *J. Biol. Chem.* **279**, 15339–15347. (doi:10.1074/jbc.M310731200)
40. Gudmundsson S *et al.* 2019 TAF1, associated with intellectual disability in humans, is essential for embryogenesis and regulates neurodevelopmental processes in zebrafish. *Sci. Rep.* **9**, 10730. (doi:10.1038/s41598-019-46632-8)
41. Li HH, Li AG, Sheppard HM, Liu X. 2004 Phosphorylation on Thr-55 by TAF1 mediates degradation of p53: a role for TAF1 in cell G1 progression. *Mol. Cell* **13**, 867–878. (doi:10.1016/s1097-2765(04)00123-6)
42. Suzuki-Yagawa Y, Guermah M, Roeder RG. 1997 The ts13 mutation in the TAF(II)250 subunit (CCG1) of TFIID directly affects transcription of D-type cyclin genes in cells arrested in G1 at the nonpermissive temperature. *Mol. Cell. Biol.* **17**, 3284–3294. (doi:10.1128/MCB.17.6.3284)
43. Deato MDE, Tjian R. 2007 Switching of the core transcription machinery during myogenesis. *Genes Dev.* **21**, 2137–2149. (doi:10.1101/gad.1583407)

44. Pijnappel WW *et al.* 2013 A central role for TFIIID in the pluripotent transcription circuitry. *Nature* **495**, 516–519. (doi:10.1038/nature11970)
45. Boutet SC, Biresi S, Iori K, Natu V, Rando TA. 2010 Taf1 regulates Pax3 protein by monoubiquitination in skeletal muscle progenitors. *Mol. Cell* **40**, 749–761. (doi:10.1016/j.molcel.2010.09.029)
46. Brunkhorst A, Karlén M, Shi J, Mikolajczyk M, Nelson MA, Metsis M, Hermanson O. 2005 A specific role for the TFIIID subunit TAF4 and ranbpm in neural progenitor differentiation. *Mol. Cell. Neurosci.* **29**, 250–258. (doi:10.1016/j.mcn.2005.02.015)
47. Langer D, Martianov I, Alpern D, Rhinn M, Keime C, Dollé P, Mengus G, Davidson I. 2016 Essential role of the TFIIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. *Nat. Commun.* **7**, 11063. (doi:10.1038/ncomms11063)
48. Bahat A, Kedmi R, Gazit K, Richardo-Lax I, Ainbinder E, Dikstein R. 2013 TAF4b and TAF4 differentially regulate mouse embryonic stem cells maintenance and proliferation. *Genes Cells* **18**, 225–237. (doi:10.1111/gtc.12030)
49. Leigh RS, Välimäki MJ, Kaynak BL, Ruskoaho HJ. 2023 TAF1 bromodomain inhibition as a candidate epigenetic driver of congenital heart disease. *Biochim. Biophys. Acta Mol. Basis Dis.* **1869**, 166689. (doi:10.1016/j.bbadis.2023.166689)
50. Janakiraman U, Yu J, Moutal A, Chinnasamy D, Boinon L, Batchelor SN, Anandhan A, Khanna R, Nelson MA. 2019 TAF1-gene editing alters the morphology and function of the cerebellum and cerebral cortex. *Neurobiol. Dis.* **132**, 104539. (doi:10.1016/j.nbd.2019.104539)
51. Janakiraman U, Dhanalakshmi C, Yu J, Moutal A, Boinon L, Fukunaga K, Khanna R, Nelson MA. 2020 The investigation of the T-type calcium channel enhancer SAK3 in an animal model of TAF1 intellectual disability syndrome. *Neurobiol. Dis.* **143**, 105006. (doi:10.1016/j.nbd.2020.105006)
52. Dhanalakshmi C, Janakiraman U, Moutal A, Fukunaga K, Khanna R, Nelson MA. 2021 Evaluation of the effects of the T-type calcium channel enhancer SAK3 in a rat model of TAF1 deficiency. *Neurobiol. Dis.* **149**, 105224. (doi:10.1016/j.nbd.2020.105224)
53. Crombie EM *et al.* 2024 Taf1 knockout is lethal in embryonic male mice and heterozygous females show weight and movement disorders. *Dis. Model. Mech.* **17**, dmm050741. (doi:10.1242/dmm.050741)
54. Hellman-Aharony S *et al.* 2013 Microcephaly thin corpus callosum intellectual disability syndrome caused by mutated TAF2. *Pediatr. Neurol.* **49**, 411–416. (doi:10.1016/j.pediatrneurol.2013.07.017)
55. Tawamie H *et al.* 2017 Hypomorphic pathogenic variants in TAF13 are associated with autosomal-recessive intellectual disability and microcephaly. *Am. J. Hum. Genet.* **100**, 555–561. (doi:10.1016/j.ajhg.2017.01.032)
56. Morton SU *et al.* 2020 Congenital heart defects due to TAF1 missense variants. *Circ. Genom. Precis. Med.* **13**, e002843. (doi:10.1161/CIRCGEN.119.002843)
57. Hurst SE *et al.* 2018 A novel variant in TAF1 affects gene expression and is associated with X-linked TAF1 intellectual disability syndrome. *Neur. Signal.* **2**, S20180141. (doi:10.1042/NS20180141)
58. Cheng H *et al.* 2019 Missense variants in TAF1 and developmental phenotypes: challenges of determining pathogenicity. *Hum. Mutat.* **41**, 1075. (doi:10.1002/humu.23936)
59. Niranjan TS, Skinner C, May M, Turner T, Rose R, Stevenson R, Schwartz CE, Wang T. 2015 Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. *PLoS One* **10**, e0116454. (doi:10.1371/journal.pone.0116454)
60. Taşkıran EZ, Karaosmanoğlu B, Koşukcu C, Ürel-Demir G, Akgün-Doğan Ö, Şimşek-Kiper PÖ, Alikışıoğlu M, Boduroğlu K, Utine GE. 2021 Diagnostic yield of whole-exome sequencing in non-syndromic intellectual disability. *J. Intellect. Disabil. Res.* **65**, 577–588. (doi:10.1111/jir.12835)
61. Hu H *et al.* 2016 X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol. Psychiatry* **21**, 133–148. (doi:10.1038/mp.2014.193)
62. Giovenino C *et al.* 2023 Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation for X-linked genes. *Eur. J. Hum. Genet.* **31**, 1228–1236. (doi:10.1038/s41431-023-01324-w)
63. Aso T, Shilatfard A, Conaway JW, Conaway RC. 1996 Transcription syndromes and the role of RNA polymerase II general transcription factors in human disease. *J. Clin. Invest.* **97**, 1561–1569. (doi:10.1172/JCI118580)
64. Koide R, Kobayashi S, Shimohata T, Ikeuchi T, Maruyama M, Saito M, Yamada M, Takahashi H, Tsuji S. 1999 A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum. Mol. Genet.* **8**, 2047–2053. (doi:10.1093/hmg/8.11.2047)
65. Friedman MJ, Shah AG, Fang ZH, Ward EG, Warren ST, Li S, Li XJ. 2007 Polyglutamine domain modulates the TBP-TFIIIB interaction: implications for its normal function and neurodegeneration. *Nat. Neurosci.* **10**, 1519–1528. (doi:10.1038/nn2011)
66. Shah AG, Friedman MJ, Huang S, Roberts M, Li XJ, Li S. 2009 Transcriptional dysregulation of TrkA associates with neurodegeneration in spinocerebellar ataxia type 17. *Hum. Mol. Genet.* **18**, 4141–4152. (doi:10.1093/hmg/ddp363)
67. Kwiatkowski TJ Jr *et al.* 2009 Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208. (doi:10.1126/science.1166066)
68. Vance C *et al.* 2009 Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208–1211. (doi:10.1126/science.1165942)
69. Tan AY, Manley JL. 2010 TLS inhibits RNA polymerase III transcription. *Mol. Cell. Biol.* **30**, 186–196. (doi:10.1128/MCB.00884-09)
70. Yang L, Embree LJ, Hickstein DD. 2000 TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins. *Mol. Cell. Biol.* **20**, 3345–3354. (doi:10.1128/MCB.20.10.3345-3354.2000)
71. Lee LV *et al.* 2011 The unique phenomenology of sex-linked dystonia parkinsonism (XDP, DYT3, “Lubag”). *Int. J. Neurosci.* **121**, 3–11. (doi:10.3109/00207454.2010.526728)
72. Lee LV, Munoz EL, Tan KT, Reyes MT. 2001 Sex linked recessive dystonia parkinsonism of Panay, Philippines (XDP). *MP Mol. Pathol.* **54**, 362–368.

73. Westenberger A *et al.* 2013 X-linked dystonia-parkinsonism manifesting in a female patient due to atypical Turner syndrome. *Mov. Disord.* **28**, 675–678. (doi:10.1002/mds.25369)
74. Evidente VGH, Nolte D, Niemann S, Advincula J, Mayo MC, Natividad FF, Müller U. 2004 Phenotypic and molecular analyses of X-linked dystonia-parkinsonism ('lubag') in women. *Arch. Neurol.* **61**, 1956–1959. (doi:10.1001/archneur.61.12.1956)
75. Goto S, Kawarai T, Morigaki R, Okita S, Koizumi H, Nagahiro S, Munoz EL, Lee LV, Kaji R. 2013 Defects in the striatal neuropeptide Y system in X-linked dystonia-parkinsonism. *Brain* **136**, 1555–1567. (doi:10.1093/brain/awt084)
76. Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP. 1985 Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* **44**, 559–577. (doi:10.1097/00005072-198511000-00003)
77. Francis A, Pulsinelli W. 1982 The response of GABAergic and cholinergic neurons to transient cerebral ischemia. *Brain Res.* **243**, 271–278. (doi:10.1016/0006-8993(82)90250-5)
78. Domingo A *et al.* 2015 New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3). *Eur. J. Hum. Genet.* **23**, 1334–1340. (doi:10.1038/ejhg.2014.292)
79. Westenberger A *et al.* 2019 A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. *Ann. Neurol.* **85**, 812–822. (doi:10.1002/ana.25488)
80. Champion LN *et al.* 2022 Tissue-specific and repeat length-dependent somatic instability of the X-linked dystonia parkinsonism-associated CCTCT repeat. *Acta Neuropathol. Commun.* **10**, 49. (doi:10.1186/s40478-022-01349-0)
81. Reyes CJ *et al.* 2021 Brain regional differences in hexanucleotide repeat length in X-linked dystonia-parkinsonism using nanopore sequencing. *Neurol. Genet.* **7**, e608. (doi:10.1212/NXG.0000000000000608)
82. Ito N *et al.* 2016 Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells. *Dis. Model. Mech.* **9**, 451–462. (doi:10.1242/dmm.022590)
83. Al Ali J *et al.* 2021 TAF1 transcripts and neurofilament light chain as biomarkers for X-linked dystonia-parkinsonism. *Mov. Disord.* **36**, 206–215. (doi:10.1002/mds.28305)
84. Rakovic A *et al.* 2018 Genome editing in induced pluripotent stem cells rescues TAF1 levels in X-linked dystonia-parkinsonism. *Mov. Disord.* **33**, 1108–1118. (doi:10.1002/mds.27441)
85. Domingo A *et al.* 2016 Evidence of TAF1 dysfunction in peripheral models of X-linked dystonia-parkinsonism. *Cell. Mol. Life Sci.* **73**, 3205–3215. (doi:10.1007/s00018-016-2159-4)
86. Petrozziello T *et al.* 2020 SVA insertion in X-linked dystonia parkinsonism alters histone H3 acetylation associated with TAF1 gene. *PLoS One* **15**, e0243655. (doi:10.1371/journal.pone.0243655)
87. Valor LM, Guiretti D, Lopez-Atalaya JP, Barco A. 2013 Genomic landscape of transcriptional and epigenetic dysregulation in early onset polyglutamine disease. *J. Neurosci.* **33**, 10471–10482. (doi:10.1523/JNEUROSCI.0670-13.2013)
88. Gonzalez-Zuñiga M, Contreras PS, Estrada LD, Chamorro D, Villagra A, Zanlungo S, Seto E, Alvarez AR. 2014 C-abl stabilizes HDAC2 levels by tyrosine phosphorylation repressing neuronal gene expression in Alzheimer's disease. *Mol. Cell.* **56**, 163–173. (doi:10.1016/j.molcel.2014.08.013)
89. Park G, Tan J, Garcia G, Kang Y, Salvesen G, Zhang Z. 2016 Regulation of histone acetylation by autophagy in Parkinson disease. *J. Biol. Chem.* **291**, 3531–3540. (doi:10.1074/jbc.M115.675488)
90. Cudkowicz ME *et al.*, Northeast ALS and National VA ALS Research Consortiums. 2009 Phase 2 study of sodium phenylbutyrate in ALS. *Amyotroph. Lateral Scler.* **10**, 99–106. (doi:10.1080/17482960802320487)
91. Kernochan LE, Russo ML, Woodling NS, Huynh TN, Avila AM, Fischbeck KH, Sumner CJ. 2005 The role of histone acetylation in SMN gene expression. *Hum. Mol. Genet.* **14**, 1171–1182. (doi:10.1093/hmg/ddi130)
92. Walker FO. 2007 Huntingtons disease. *Lancet* **369**, 218–228. (doi:10.1016/S0140-6736(07)60111-1)
93. Macdonald M. 1993 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983. (doi:10.1016/0092-8674(93)90585-E)
94. Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, Wheeler VC. 2009 Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **18**, 3039–3047. (doi:10.1093/hmg/ddp242)
95. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. 2019 CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell* **178**, 887–900. (doi:10.1016/j.cell.2019.06.036)
96. Laabs BH *et al.* 2021 Identifying genetic modifiers of age-associated penetrance in X-linked dystonia-parkinsonism. *Nat. Commun.* **12**, 3216. (doi:10.1038/s41467-021-23491-4)
97. Zhai W, Jeong H, Cui L, Krainc D, Tjian R. 2005 *In vitro* analysis of huntingtin-mediated transcriptional repression reveals multiple transcription factor targets. *Cell* **123**, 1241–1253. (doi:10.1016/j.cell.2005.10.030)
98. Borovecki F *et al.* 2005 Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proc. Natl Acad. Sci. USA* **102**, 11023–11028. (doi:10.1073/pnas.0504921102)
99. Scherzinger E, Sittler A, Schweiger K, Heiser V, Lurz R, Hasenbank R, Bates GP, Lehrach H, Wanker EE. 1999 Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc. Natl Acad. Sci. USA* **96**, 4604–4609. (doi:10.1073/pnas.96.8.4604)
100. Schneider JA, Arvanitakis Z, Yu L, Boyle PA, Leurgans SE, Bennett DA. 2012 Cognitive impairment, decline and fluctuations in older community-dwelling subjects with lewy bodies. *Brain* **135**, 3005–3014. (doi:10.1093/brain/aws234)
101. Lees AJ, Hardy J, Revesz T. 2009 Parkinson's disease. *Lancet* **373**, 2055–2066. (doi:10.1016/S0140-6736(09)60492-X)

102. Bloem BR, Okun MS, Klein C. 2021 Parkinson's disease. *Lancet* **397**, 2284–2303. (doi:10.1016/S0140-6736(21)00218-X)
103. Zeng Q *et al.* 2022 Association study of TAF1 variants in Parkinson's disease. *Front. Neurosci.* **16**, 846095. (doi:10.3389/fnins.2022.846095)
104. Ribeiro JR, Lovasco LA, Vanderhyden BC, Freiman RN. 2014 Targeting TBP-associated factors in ovarian cancer. *Front. Oncol.* **4**, 45. (doi:10.3389/fonc.2014.00045)
105. Allende-Vega N, McKenzie L, Meek D. 2008 Transcription factor TAFII250 phosphorylates the acidic domain of Mdm2 through recruitment of protein kinase CK2. *Mol. Cell. Biochem.* **316**, 99–106. (doi:10.1007/s11010-008-9816-3)
106. Buchmann AM, Skaar JR, DeCaprio JA. 2004 Activation of a DNA damage checkpoint response in a TAF1-defective cell line. *Mol. Cell. Biol.* **24**, 5332–5339. (doi:10.1128/MCB.24.12.5332-5339.2004)
107. Kimura J, Nguyen ST, Liu H, Taira N, Miki Y, Yoshida K. 2008 A functional genome-wide RNAi screen identifies TAF1 as a regulator for apoptosis in response to genotoxic stress. *Nucleic Acids Res.* **36**, 5250–5259. (doi:10.1093/nar/gkn506)
108. Wada C, Kasai K, Kameya T, Ohtani H. 1992 A general transcription initiation factor, human transcription factor IID, overexpressed in human lung and breast carcinoma and rapidly induced with serum stimulation. *Cancer Res.* **52**, 307–313.
109. Oh HR, An CH, Yoo NJ, Lee SH. 2017 Frameshift mutations in the mononucleotide repeats of TAF1 and TAF1L genes in gastric and colorectal cancers with regional heterogeneity. *Pathol. Oncol. Res.* **23**, 125–130. (doi:10.1007/s12253-016-0107-0)
110. Findlay JM *et al.* 2016 Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat. Commun.* **7**, 11111. (doi:10.1038/ncomms11111)
111. Tavassoli P, Wafa LA, Cheng H, Zoubeidi A, Fazli L, Gleave M, Snoek R, Rennie PS. 2010 TAF1 differentially enhances androgen receptor transcriptional activity via its N-terminal kinase and ubiquitin-activating and -conjugating domains. *Mol. Endocrinol.* **24**, 696–708. (doi:10.1210/me.2009-0229)
112. Zhao S *et al.* 2013 Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl Acad. Sci. USA* **110**, 2916–2921. (doi:10.1073/pnas.1222577110)
113. Waters CH, Faust PL, Powers J, Vinters H, Moskowitz C, Nygaard T, Hunt AL, Fahn S. 1993 Neuropathology of lubag (x-linked dystonia parkinsonism). *Mov. Disord.* **8**, 387–390. (doi:10.1002/mds.870080328)
114. Domingo A, Lee LV, Brüggemann N, Freimann K, Kaiser FJ, Jamora RDG, Rosales RL, Klein C, Westenberger A. 2014 Woman with x-linked recessive dystonia-parkinsonism: clue to the epidemiology of parkinsonism in Filipino women? *JAMA Neurol.* **71**, 1177–1180. (doi:10.1001/jamaneurol.2014.56)
115. Deng H, Le WD, Jankovic J. 2008 Genetic study of an American family with DYT3 dystonia (lubag). *Neurosci. Lett.* **448**, 180–183. (doi:10.1016/j.neulet.2008.10.049)
116. Vianna EQ *et al.* 2020 Understanding the landscape of X-linked variants causing intellectual disability in females through extreme X chromosome inactivation skewing. *Mol. Neurobiol.* **57**, 3671–3684. (doi:10.1007/s12035-020-01981-8)
117. Okamoto N, Arai H, Onishi T, Mizuguchi T, Matsumoto N. 2020 Intellectual disability and dysmorphic features in male siblings arising from a novel *taf1* mutation. *Congenit. Anom. (Kyoto)* **60**, 40–41. (doi:10.1111/cga.12330)
118. Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA, Nelson LM. 2003 Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. *Am. J. Epidemiol.* **157**, 1015–1022. (doi:10.1093/aje/kwg068)
119. Wang PJ, Page DC. 2002 Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum. Mol. Genet.* **11**, 2341–2346. (doi:10.1093/hmg/11.19.2341)
120. Wang PJ. 2004 X chromosomes, retrogenes and their role in male reproduction. *Trends Endocrinol. Metab.* **15**, 79–83. (doi:10.1016/j.tem.2004.01.007)
121. Zhong S, Yan H, Chen Z, Li Y, Shen Y, Wang Y, Li L, Sheng S, Wang Y. 2019 Overexpression of TAF1L promotes cell proliferation, migration and invasion in esophageal squamous cell carcinoma. *J. Cancer* **10**, 979–989. (doi:10.7150/jca.26504)
122. Chen X, Hiller M, Sancak Y, Fuller MT. 2005 Tissue-specific tafs counteract polycomb to turn on terminal differentiation. *Science* **310**, 869–872. (doi:10.1126/science.1118101)
123. Freiman RN, Albright SR, Zheng S, Sha WC, Hammer RE, Tjian R. 2001 Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* **293**, 2084–2087. (doi:10.1126/science.1061935)
124. Calarco JA *et al.* 2009 Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* **138**, 898–910. (doi:10.1016/j.cell.2009.06.012)