

Disorders of the Nervous System

# Variation in TAF1 Expression in Female Carrier-Induced Pluripotent Stem Cells and Human Brain Ontogeny Has Implications for Adult Neostriatum Vulnerability in X-Linked Dystonia Parkinsonism

Laura D'Ignazio,<sup>1,2</sup> Ricardo S. Jacomini,<sup>1,2</sup> Bareera Qamar,<sup>1</sup>  Kynon J. M. Benjamin,<sup>1,2,3</sup> Ria Arora,<sup>1,4</sup> Tomoyo Sawada,<sup>1,2</sup> Taylor A. Evans,<sup>1,2</sup> Kenneth E. Diffenderfer,<sup>5</sup> Aimee R. Pankonin,<sup>5</sup> William T. Hendriks,<sup>6,7</sup> Thomas M. Hyde,<sup>1,3</sup> Joel E. Kleinman,<sup>1,3</sup> Daniel R. Weinberger,<sup>1,2,3,8,9</sup> D. Christopher Bragg,<sup>6,7</sup> Apua C. M. Paquola,<sup>1,2</sup> and  Jennifer A. Erwin<sup>1,2,3,8</sup>

<https://doi.org/10.1523/ENEURO.0129-22.2022>

<sup>1</sup>Lieber Institute for Brain Development, Baltimore, MD 21205, <sup>2</sup>Department of Neurology, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, <sup>3</sup>Department of Psychiatry & Behavioral Sciences, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, <sup>4</sup>Department of Biology, Krieger School of Arts & Sciences, Johns Hopkins University, Baltimore, MD 21218, <sup>5</sup>Stem Cell Core, Salk Institute for Biological Studies, La Jolla, CA 92037, <sup>6</sup>Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, <sup>7</sup>The Collaborative Center for X-linked Dystonia-Parkinsonism, Massachusetts General Hospital, Charlestown, MA 02129, <sup>8</sup>Department of Neuroscience, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, and <sup>9</sup>McKusick-Nathans Department of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21205

## Abstract

X-linked dystonia-parkinsonism (XDP) is an inherited, X-linked, adult-onset movement disorder characterized by degeneration in the neostriatum. No therapeutics alter disease progression. The mechanisms underlying regional differences in degeneration and adult onset are unknown. Developing therapeutics requires a deeper understanding of how XDP-relevant features vary in health and disease. XDP is possibly due, in part, to a partial loss of *TAF1* function. A disease-specific SINE-VNTR-Alu (SVA) retrotransposon insertion occurs within intron 32 of *TAF1*, a subunit of TFIID involved in transcription initiation. While all XDP males are usually clinically affected, females are heterozygous carriers generally not manifesting the full syndrome. As a resource for disease modeling, we characterized eight iPSC lines from three XDP female carrier individuals for X chromosome inactivation (XCI) status and identified clonal lines that express either the wild-type X or XDP haplotype. Furthermore, we characterized XDP-relevant transcript expression in neurotypical humans, and found that SVA-F expression decreases after 30 years of age in the brain and that *TAF1* is decreased in most female samples. Uniquely in the caudate nucleus, *TAF1* expression is not sexually dimorphic and decreased after

## Significance Statement

X-linked dystonia-parkinsonism (XDP) is an inherited, X-linked, adult-onset movement disorder characterized by degeneration in the neostriatum. No therapeutics alter disease progression. Developing therapeutics requires a deeper understanding of how XDP-relevant features vary in health and disease. XDP is possibly because of a partial loss of *TAF1* function. While all XDP males are usually affected, females are heterozygous carriers generally not manifesting the full syndrome. As a resource for disease modeling, we characterized eight stem cell lines from XDP female carrier individuals. Furthermore, we found that, uniquely in the caudate nucleus, *TAF1* expression decreases after adolescence in healthy humans. We hypothesize that the decrease of *TAF1* after adolescence in human caudate, in general, may underlie the vulnerability of the adult neostriatum in XDP.

adolescence. These findings indicate that regional-specific, age-specific, and sex-specific mechanisms regulate *TAF1*, highlighting the importance of disease-relevant models and postmortem tissue. We propose that the decreased *TAF1* expression in the adult caudate may synergize with the XDP-specific partial loss of *TAF1* function in patients, thereby passing a minimum threshold of *TAF1* function, and triggering degeneration in the neostriatum.

**Key words:** induced pluripotent stem cells; movement disorder; neurodegeneration; repeat expansion; SVA retrotransposon; X-linked dystonia parkinsonism

## Introduction

X-linked dystonia-parkinsonism (XDP) is a rare movement disorder, predominantly affecting males descending from the island of Panay, in the Philippines. Disease onset typically occurs within the third or fourth decade of life, and clinical manifestations consist of hyperkinetic movements and focal dystonia shifting toward parkinsonian features in a temporal evolution spanning 5–10 years (Lee et al., 2011). Treatment is limited to symptomatic management, including oral medications, botulinum toxin, and pallidal deep brain stimulation (Bragg et al., 2019).

Neuropathological analyses of a few XDP postmortem brains demonstrated progressive neuronal loss within the neostriatum, particularly affecting medium spiny neurons (MSNs; Goto et al., 2005, 2013; Hanssen et al., 2018). Neuroimaging has detected other neostriatal and extra-neostriatal abnormalities (Brüggemann et al., 2016; Walter et al., 2017; Blood et al., 2018; Hanssen et al., 2018; Arasaratnam et al., 2021). Similar to many neurodegenerative disorders, the mechanisms underlying

XDP neurodegeneration and the differential vulnerability to pathology are unknown.

XDP is a Mendelian disorder and thirteen shared variants have been detected in all XDP probands: 11 single disease-specific nucleotide changes (Nolte et al., 2003; Aneichyk et al., 2018), a 48-bp deletion (Nolte et al., 2003), and an ~2.6-kb SVA-F type [short interspersed nuclear elements (SINE)-variable number of tandem repeats (VNTR)-Alu] retrotransposon insertion (Makino et al., 2007). All variants occur in intergenic or intronic regions within the *TAF1* gene [TATA-binding protein (TBP)-associated factor 1], encoding a core component of the TFIID complex that coordinates RNA polymerase II (RNA pol II) promoter initiation (Domingo et al., 2015).

XDP patient samples have decreased expression of the canonical full-length *TAF1* (*cTAF1*) transcript in dividing cells (Aneichyk et al., 2018; Rakovic et al., 2018), patient blood RNA (Al Ali et al., 2021), and a 2D cellular model of MSNs (Rakovic et al., 2018). The *cTAF1* decreases are caused by the SVA, which induces alternative splicing and intron retention (Aneichyk et al., 2018). In XDP cells, one of the most abundant aberrant transcripts caused by the pathogenic SVA is *TAF1-32i*, composed of canonical exon 32, spliced to a cryptic exon in intron 32, that terminates 5' to the SVA insertion site (Aneichyk et al., 2018).

SVAs are noncoding hominid-specific human retrotransposons. There are roughly 2700 SVA copies per human genome, with the SVA-F being the evolutionary youngest (Hancks and Kazazian, 2010). SVA elements are regulated by protein coding genes (Savage et al., 2013), but the dynamics of SVA expression during human brain lifespan have not been well elucidated.

XDP is an X-linked recessive disorder, and the vast majority of affected cases are hemizygous males. Heterozygous carrier females typically do not manifest the full syndrome (Evidente et al., 2004; Domingo et al., 2014). Female carriers without the full syndrome do demonstrate some pathology, including decreased caudate and putamen volumes compared with healthy controls (Blood et al., 2018). Understanding how XDP manifests in female carriers can reveal information about disease pathophysiology.

As a consequence of the dosage compensation mechanism X chromosome inactivation (XCI), females transcriptionally silence most genes on one X chromosome, including *TAF1* (Oliva et al., 2020; Patrat et al., 2020). Tissues of heterozygous female XDP carriers are a mosaic of two cell populations, with either the wild-type or mutant X chromosome randomly inactivated per cell (Migeon, 2020). XCI is a dosage compensation mechanism that achieves equivalent X-linked gene expression between XY males and XX females. Postimplantation,

Received March 24, 2022; accepted July 3, 2022; First published July 21, 2022.

The authors declare no competing financial interests.

Author contributions: L.D., A.C.M.P., and J.A.E. designed research; L.D., R.S.J., B.Q., K.J.M.B., R.A., K.E.D., A.R.P., A.C.M.P., and J.A.E. performed research; W.T.H., T.M.H., J.E.K., D.R.W., D.C.B., and J.A.E. contributed unpublished reagents/analytic tools; L.D., R.S.J., B.Q., K.J.M.B., R.A., T.S., T.A.E., A.C.M.P., and J.A.E. analyzed data; L.D., R.S.J., K.J.M.B., D.C.B., A.C.M.P., and J.A.E. wrote the paper.

This work was supported by the Lieber Institute for Brain Development, the Massachusetts General Hospital Collaborative Center for X-Linked Dystonia-Parkinsonism (J.A.E., A.C.M.P., D.C.B.), the Maryland Stem Cell Research Fund (L.D.), and the National Institutes of Health Grant T32MH015330 (to K.J.M.B.).

Acknowledgments: We thank Dr. Ronald McKay and Dr. Mike McConnell (Lieber Institute for Brain Development) for their thoughtful feedback throughout all aspects of the study; the contributions of the Office of the Chief Medical Examiner of the District of Columbia; the Office of the Chief Medical Examiner for Northern Virginia, Fairfax Virginia; and the Office of the Chief Medical Examiner of the State of Maryland, Baltimore, Maryland; the Medical University of Sofia Bulgaria; and Dr. H. Ronald Zielke, Robert D. Vigorito, and Robert M. Johnson of the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders, at the University of Maryland, for providing child and adolescent brain specimens. Finally, we are indebted to the generosity of the families of the decedents, who donated the brain tissue used in these studies.

Correspondence should be addressed to Jennifer A. Erwin at [jennifer.erwin@libd.org](mailto:jennifer.erwin@libd.org).

<https://doi.org/10.1523/ENEURO.0129-22.2022>

Copyright © 2022 D'Ignazio et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

one X chromosome is epigenetically silenced and coated by *XIST* on the inactive X chromosome. Female primed pluripotent stem cells (PSCs) generally harbor one active and one inactive X chromosome, but also demonstrate epigenetic variability in dosage compensation status, and some cultures demonstrate erosion of dosage compensation (Shen et al., 2008; Silva et al., 2008; Mekhoubad et al., 2012).

Here we perform extensive characterization of eight human induced PSC (iPSC) lines reprogrammed from fibroblasts derived from three XDP heterozygous female carriers (Ito et al., 2016; Aneichyk et al., 2018) to serve as a valuable tool to study XDP pathogenesis *in vitro* on differentiation into specific neuronal models.

Furthermore, the natural level of *TAF1* and *SVA* expression variation in humans is unknown. We hypothesize that age and tissue specific variation in *TAF1* and *SVA* expression may underlie the age and tissue-specific degeneration in XDP. Therefore, we performed extensive transcriptomic analyses of stem cell models and postmortem neurotypical brain regions to characterize *TAF1* and *SVA* expression in relevant human tissue and cells.

## Materials and Methods

### Cell culture of iPSC lines

Derivation and reprogramming from skin fibroblasts, as well as the initial characterization of iPSCs derived from a male patient with XDP, a male unaffected relative (control), and three XDP female carriers have previously been reported elsewhere (Ito et al., 2016; Aneichyk et al., 2018). The full list of iPSCs used in our study is reported in Extended Data Figure 1-1, including the XDP isogenic *SVA*-deleted iPSC line ( $\Delta$ *SVA*-XDP), previously generated through CRISPR/Cas9 genome editing (Aneichyk et al., 2018). All iPSCs were expanded and grown on Cultrex Growth factor Reduced BME (Bio-Techne) in mTeSR or mTeSR Plus medium (Stemcell). The iPSCs were passaged every 4–6 d with Versene solution (Invitrogen) and split 1:8–1:10 each time. Cells were routinely tested for mycoplasma contamination using a MycoAlert Mycoplasma Detection kit (Lonza).

### Ethics approval and consent

Performance of all experiments in this study strictly followed ethical guidelines, and any necessary Institutional Review Board (IRB) and/or ethics committee approvals were obtained. The cell lines used comprised de-identified iPSCs, which were derived previously (not specifically for the purpose of this study) from dermal punch fibroblasts at Massachusetts General Hospital (MGH; Boston, MA). These specimens, sample collection, and derivation, have been approved by MGH IRB and were de-identified by MGH. These samples were transferred to the Lieber Institute for Brain Development (LIBD) under a Material Transfer Agreement (MTA). Later, the Western IRB granted an exemption under 45 CFR 46.101(b)(4), because our study involved the use of retrospective data that does not contain patient identifying information.

We followed ethics guidelines set forth by LIBD and the Maryland Stem Cell Research Act.

Human postmortem brain tissues used for this study, and included in the BrainSeq Consortium, were collected at several sites. Most adult postmortem human brain samples were provided by the Clinical Brain Disorders Branch of the National Institute of Mental Health (NIMH) in collaboration with the Northern Virginia and District of Columbia Medical Examiners' Offices, following a protocol approved by the IRB of the National Institutes of Health (NIH). Samples were transferred to the LIBD under an MTA with the NIMH. Additional samples were collected at the LIBD according to a protocol approved by the IRB of the State of Maryland Department of Health and Mental Hygiene and the Western IRB. Child and adolescent brain tissue samples were provided by the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders via MTAs approved by IRB of the University of Maryland. Audiotaped informed consent to study brain tissue was obtained from the legal next-of-kin on every case collected at NIMH and LIBD. Further details about donation process and specimens handling have been described previously (Lipska et al., 2006). Neurotypical postmortem brain samples used in this study were nonpsychiatric non-neurologic controls with no known history of significant psychiatric or neurologic illnesses, including substance abuse. All necessary participant consent has been obtained and the appropriate institutional forms have been archived.

### RNA scope

An *in situ* hybridization assay was performed with RNAscope technology using the RNAscope Fluorescent Multiplex kit V2 (ACD) according to the manufacturer's instructions, and followed recommendations for cultured adherent cell preparations. Briefly, iPSCs were cultured in Cultrex-coated four-well chambered coverslips (ibidi), and when 70–80% confluent, fixed with a 10% neutral buffered formalin solution (Sigma-Aldrich) for 30 min at room temperature. Then, iPSCs were series dehydrated in ethanol, pretreated with hydrogen peroxide for 10 min at room temperature, and treated with protease III diluted 1:15 in 1× PBS for 10 min. Cells were incubated with probes for *XIST* (#311231-C1, ACD), *XACT* (#471031-C2, ACD), *HUWE1* (#526691-C2, ACD), and *DUSP6* (#405361-C3, ACD), and stored overnight at 4°C in 4× SSC (saline-sodium citrate) buffer. The following day, probes were fluorescently labeled with Opal Dyes (PerkinElmer) diluted in the TSA buffer. Specifically, Opal520 diluted 1:500 was assigned to *HUWE1* and *XACT*; Opal570 diluted 1:500 was assigned to *XIST*; Opal690 diluted 1:500 was assigned to *DUSP6*. To label nuclei, we used 4',6-diamidino-2-phenylindole (DAPI). Imaging was performed using a laser confocal microscope (LSM700, Zeiss) and a 63× objective lens.

### Immunofluorescence staining and microscopy

For immunofluorescence staining, iPSCs were cultured in Cultrex-coated imaging dishes with a polymer coverslip

bottom (ibidi). When colonies were 80–90% confluent, iPSCs were fixed with 4% paraformaldehyde (Sigma) at room temperature for 15 min, permeabilized and blocked with 0.3% Triton X-100 (Sigma) and 10% normal horse serum (Thermo Fisher Scientific) in D-PBS (Invitrogen) for 30 min at room temperature. Cells were then stained overnight at 4°C, with two primary antibodies: goat anti-NANOG (1:200, #AF1997, R&D Systems) and mouse anti-TRA-1-60 (1:200, #MAB4360, Millipore Sigma), both diluted in 5% normal horse serum/0.01% Tween 20/D-PBS. The following day, iPSCs were incubated with donkey anti-goat IgG-AF488 conjugated (1:250, #705-545-147, Jackson ImmunoResearch), and donkey anti-mouse IgG-Cy3 conjugated (1:250, #715-165-151, Jackson ImmunoResearch), diluted in 0.01% Tween 20/D-PBS for 2 h at room temperature. Cell nuclei were labeled with Hoechst 33342 (1:10 000, Invitrogen) for 10 min. Stained cells were imaged with laser confocal microscopy (LSM700, Zeiss) and processed with Zen software (Zeiss) and Adobe Photoshop CS4 (Adobe Systems).

### Genotyping and genome stability analysis

Genomic DNA was isolated from iPSCs with a DNeasy Blood and Tissue kit (QIAGEN) as recommended by the manufacturer. Genotyping array analysis was performed by Psomagen with an Infinium Omni 2.5–8 kit (Illumina). To detect SVA, long-range PCR was performed following the method previously described by (Ito et al., 2016). The PCR amplicons were resolved by gel electrophoresis to identify a 3229-bp product (corresponding to the ~2.6-kb SVA insertion), or a 599-bp product (without a SVA insertion).

### RNA extraction and real-time quantitative PCR analysis

Total RNA was extracted from iPSCs with a Direct-zol Miniprep kit (Zymo Research). For reverse transcription, we used SuperScript IV VILO Master Mix (Thermo Fisher Scientific). The expression of XIST and TAF1 was analyzed via qPCR with a QuantiTect SYBR Green PCR kit (QIAGEN) on a QuantStudio 3 (Applied Biosystems). The following 5 primer sets were used: XIST forward: 5'-TTGCCCTACTAGCTCCTCGGAC-3'; XIST reverse: 5'-TTCTCCAGATAGCTGGCAACC-3'; TAF1 forward: 5'-AGAGTCGGGAGAGCTTTCTG-3'; TAF1 reverse: 5'-CACAAATCTCCTGGGCAGTCT-3'; GAPDH forward: 5'-TGCACCACCAACTGCTTAGC-3'; GAPDH reverse: 5'-GGCATGGAAGTGTGGTCATGAG-3'; HPRT forward: 5'-TGACACTGGCAAACAATGCA-3'; HPRT reverse: 5'-GGT CCTTTTACCAGCAAGCT-3'; GUSB forward: 5'-GTCTGCGGCATTTTGTCCG-3'; GUSB reverse: 5'-CACACGATGGCATAGGAATGG-3'.

To quantify intron 32 retention, we used a method previously described (Aneichyk et al., 2018), as the authors generated a custom TaqMan primer/probe to detect the exon 32/intron 32 splice site (#AJWR28J, Thermo Fisher Scientific). When screening the XDP female carrier iPSC lines, 500 ng of RNA was reverse transcribed into cDNA using Superscript III First Strand Synthesis SuperMix

(Thermo Fisher Scientific) with oligo(dT). The resulting cDNA was then further amplified through nested PCR using Phusion 2× Hot Start Flex MasterMix (NEB), as recommended by the manufacturer, with the following primers: TAF1-nested forward: 5'-ACATCTCCAAGCA CAAGTATCA-3'; TAF1-nested reverse: 5'-GTAATGTA CCAATATAAATTTCTGGTTT-3'. The thermocycler conditions used for the nested PCR were: 98°C for 30 s; 5 cycles of 98°C for 10 s, 61°C for 30 s, 72°C for 30 s; 72°C for 5 min; and infinite hold at 4°C.

The nested PCR products were purified using a DNA clean/concentrator kit (Zymo Research) following the manufacturer's instructions. Two microliters (~25 ng/μl) of purified PCR products were used as template for subsequent qPCR performed on a QuantStudio 3 (Applied Biosystems) using TaqMan Fast Advanced Master Mix (Thermo Scientific), and the following conditions: 95°C for 20 s; 40 cycles of 95°C for 1 s, and 60°C for 20 s. The GUSB probe/primer (#Hs009627\_m1, Thermo Fisher Scientific) was used as a normalizer for gene expression analysis.

After qPCR analyses, for statistical purposes, one-way ANOVA followed by Dunnett's test was used for comparing multiple conditions to a control condition only; one-way ANOVA followed by Tukey's test was used for multiple pairwise condition comparisons. In all cases,  $p$ -values were determined as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

### Allele-specific transcriptomic analysis

Total RNA was extracted from the eight XDP female carrier-derived iPSC lines, and libraries were prepared using the TruSeq Stranded Total RNA Gold kit to remove both cytoplasmic and mitochondrial rRNA. Paired-end 150-bp Illumina sequencing was performed at a median depth of 30 million paired reads. Correction and filtering for technical errors (such as nonunique alignments, low-quality bases, overlapping mates and duplicate reads), were conducted to remove potential sequencing errors and to minimize false-positive single nucleotide polymorphism (SNP) identification. Reads were trimmed with Trimmomatic [0.39; (Bolger et al., 2014)]. To assess read quality before and after trimming, we performed FastQC (v0.11.8; <https://github.com/s-andrews/FastQC>). Then, we mapped the trimmed reads to GENCODE v32 (hg38/GRCh38) primary assembly using STAR aligner (2.7.3a; Dobin et al., 2013). We examined alignment and additional read quality metrics with RSeQC (3.0.1; L Wang et al., 2012).

To identify isogenic clonal pairs, single nucleotide variants (SNVs) were called from X chromosome-derived transcripts in all clones derived from the same XDP carrier individual (33360, 33811, or 33110), using GATK best practices for SNPs and indel calling on RNA-Seq data. More specifically, allele-specific expression was analyzed following the GATK framework and GATK ASEReadCounter. Clonal pairs with the majority of SNVs unique to each clone were identified as isogenic pairs, as different X chromosomes were inactive in the two distinct lines. Conversely, clone pairs sharing most

of SNVs were identified as iPSC lines with the same X chromosome inactivated.

### Global gene expression analysis

Gene counts were generated via featureCount [v1.5.3; (Liao et al., 2014)], a SubRead utility, for paired end, reversed stranded, reads in parallel (Tange, 2018), with four read filters: (1) uniquely mapped reads corresponding to a mapping quality of 60 for STAR BAM files; (2) reads aligned in proper pairs; (3) excluding chimeric reads; and (4) primary alignments. Raw counts were filtered for low expression by using filterByExpr from the edgeR Bioconductor package (Robinson et al., 2010; McCarthy et al., 2012), normalized for library size, followed by log2 counts per million (CPM) transformation.

### Differential expression analysis of the HipSci iPSC dataset

We downloaded FASTQ files associated with human iPSCs from (Kilpinen et al., 2017;  $n = 762$ ; #PRJEB7388). Read quality was examined with FastQC before alignment to GENCODE v32 with HISAT2 (v2.1.0; Kim et al., 2015) and Salmon (v1.2.1; Patro et al., 2017). For HISAT2 aligned reads, sorted BAM files were generated and indexed using SAMtools (v1.9). We generated gene counts with featureCount for paired end, reversed stranded reads similar to XDP gene counts using three read filters: (1) reads aligned in proper pairs; (2) excluding chimeric reads; and (3) primary alignments.

We selected samples based on two inclusion criteria: (1) British ethnicity of origin and (2) white ethnic group. This resulted in a total of 746 hiPSCs, comprising 438 females and 308 males, from 323 unique cell lines, with an age range from 25 to 79.

We performed quality control in two steps. In the first, we removed outliers by expression. To visualize outliers, we applied UMAP (uniform manifold approximation and projection) dimensional reduction to normalize Salmon-generated TPM (transcripts per million) log2 transformed expression. This resulted in two separated clusters, for which we dropped the smaller cluster, resulting in 662 hiPSCs. In the second quality control step, we performed RSeQC and aggregated the results of HISAT2 and FastQC using MultiQC (Ewels et al., 2016). We determined outliers based on read quality using principal component analysis (PCA) and centroid analysis. Specifically, we applied PCA to read quality metrics aggregated from MultiQC. From there, we calculated the distance from the centroid for all samples, and dropped samples that were outside of the 95th percentile. This resulted in a total of 628 hiPSCs, comprising 359 female and 269 male hiPSCs.

As there were multiple clones per cell line for many individuals, we performed differential expression analysis after randomly selecting one unique cell line. To include all samples, we performed this random sampling 100 times. For each differential expression analysis, we filtered out low expression via filterByExpr, normalized for library sizes, before applying voom normalization (Law et

al., 2014; Ritchie et al., 2015). Following voom normalization, we fitted a linear model (Eq. 1) to examine sex. Differentially expressed genes were identified using the eBayes function (Smyth, 2004) from limma. For plots, we calculated the median differential expression statistics and voom expression across iterations:

$$E(Y) = \beta_0 + \beta_1 \text{Sex} \quad (1)$$

### Differential expression analyses and life span analysis of human postmortem brains

We downloaded FASTQ files and phenotype information for the caudate nucleus, dorsolateral prefrontal cortex (DLPFC) and hippocampus from the BrainSeq Consortium (Benjamin et al., 2020), selecting samples based on (1) diagnosis of neurotypical control and schizophrenia; (2) postnatal; (3) African and European ancestries; and (4) RiboZero RNA library preparation. This resulted in 1236 samples associated with 554 unique individuals (females,  $n = 174$ ; males,  $n = 380$ ). We aligned FASTQ files to GENCODE v26 with HISAT2 and generated sorted and indexed BAM files with SAMtools. Following alignment and sorting, we generated gene, repeat, and exon counts using featureCounts. For gene and repeat counts, we created a customized GTF file of genes and repeats by combining GENCODE v26 with repeat masker (UCSC Table Browser; hg38) repeat annotation after we annotated repeats by strand. With this GTF file, we then generated counts with featureCounts using six parameters: (1) paired end; (2) reversed stranded reads; (3) primary alignments only; (4) excluding chimeric reads; (5) allowing for multimapping reads; and (6) one base as the minimum overlapping fraction in a read. For exon counts, we used GENCODE v26 with featureCounts for paired end, reversed stranded reads with four read filters: (1) uniquely mapped reads; (2) reads aligned in proper pairs; (3) excluding chimeric reads; and (4) primary alignments.

Similar to the hiPSC analysis, we assessed quality control based on read quality outliers determined via PCA and centroid analysis. We excluded samples that were outside of the 99th percentile (caudate nucleus and hippocampus) and 95th percentile (DLPFC), resulting in a total of 1203 samples: caudate nucleus ( $n = 422$ ), DLPFC ( $n = 412$ ), and hippocampus ( $n = 369$ ).

For life span differential expression analysis of human postmortem brains, we first selected neurotypical controls, resulting in a total of 768 samples (associated with 353 unique individuals): caudate nucleus ( $n = 269$ ), DLPFC ( $n = 260$ ), and hippocampus ( $n = 239$ ). After this sample selection, we then filtered out low expression with filterByExpr, and normalized for library sizes. For lifespan analysis, we used four age groups based on caudate developmental windows: (1) 0–15 years; (2) 16–30 years; (3) 31–50 years; and (4) 51+ years. For voom normalization, we fitted a linear model (Eq. 2) for age group, adjusting for sex, ancestry (SNP PC 1–3 and self-reported), RNA integrity number (RIN), mitochondria mapping rate, gene assignment rate, genome mapping rate, and any hidden variances with surrogate variable analysis (Leek et al., 2012). We identified differentially expressed

genes, repeats, and exons using eBayes. For plotting, we generated residualized data by regressing out covariates, and applying a z score transformation from the voom normalized counts.

$$E(Y) = \beta_0 + \beta_1 \text{Age.Group} + \beta_2 \text{Sex} + \beta_3 \text{Ethnicity} + \beta_4 \text{MitoRate} + \beta_5 \text{RIN} + \beta_6 \text{TotalAssignedGenes} + \beta_6 \text{OverallMappingRate} + \sum_{i=1}^3 \eta_i \text{snpPC}_i + \sum_{j=1}^k \gamma_j \text{SV}_j. \quad (2)$$

To examine sex differences for repetitive elements, we first selected adults (age > 17), resulting in a total of 1085 samples, associated with 479 unique individuals: caudate nucleus ( $n = 394$ ; 241CTL/153SZ; 120F/274 M; 206AA/188EA), DLPFC ( $n = 365$ ; 213CTL/152SZ; 120F/245 M; 202AA/163EA), and hippocampus ( $n = 326$ ; 196CTL/130SZ; 100F/226 M; 178AA/148EA). After this sample selection, we then performed differential expression analysis by sex, for genes and repeats together, after filtering for low expression, normalizing for library size, and voom normalization. For voom normalization, we fit a linear model (Eq. 3) for sex, adjusting for eight factors: diagnosis, age, ancestry (SNP PCs 1–3 and self-reported), RIN, mitochondria mapping rate, gene assignment rate, genome mapping rate, and hidden variance. For plotting, we generated residualized data using voom normalized counts by regressing out covariates and applying a z score transformation:

$$E(Y) = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{Diagnosis} + \beta_3 \text{Age} + \beta_4 \text{Ethnicity} + \beta_5 \text{MitoRate} + \beta_6 \text{RIN} + \beta_7 \text{TotalAssignedGenes} + \beta_8 \text{OverallMappingRate} + \sum_{i=1}^3 \eta_i \text{snpPC}_i + \sum_{j=1}^k \gamma_j \text{SV}_j. \quad (3)$$

All transcriptomic analyses of *TAF1* expression at the gene and exon levels have been conducted referring to the canonical isoform *TAF1-204* (ENST00000423759.6). The sex differential expression analysis of *TAF1* gene expression in human postmortem brain tissues shown in Extended Data Fig. 5-1 is a graphical representation of previously published data from the GTEX v8 collection (see Oliva et al., 2020; their Supplementary Table S3).

### Dosage compensation analysis

For dosage compensation, we calculated relative X expression (RXE) as previously described (Jue et al., 2013) with slight modifications to replace fragments per kilobase of transcript per million mapped reads (FPKM) with TPM (transcripts per million). We generated TPM from gene counts using effective length calculated from mean insert size, which was determined using the Picard tool CollectInsertSizeMetrics (<http://broadinstitute.github.io/picard/>). Genes with an effective length less than or equal to 1 were excluded. To compute RXE, we filtered TPM genes with low expression log<sub>2</sub> transformed before calculating the difference in the mean of

autosome chromosomes compared with the X chromosome (Eq. 4):

$$RXE = \log_2(X) - \log_2(A). \quad (4)$$

### SVA transcripts analysis

To detect SVA-derived transcripts in XDP female carrier iPSC lines, we aligned all reads to the XDP SVA sequence obtained from (Aneichyk et al., 2018) using HISAT2, and to calculate coverage at each base pair in SVA we used SAMtools.

### Graphics

We generated RXE scatter plots and boxplots using ggpubr (v0.4.0) in R.

### Code accessibility

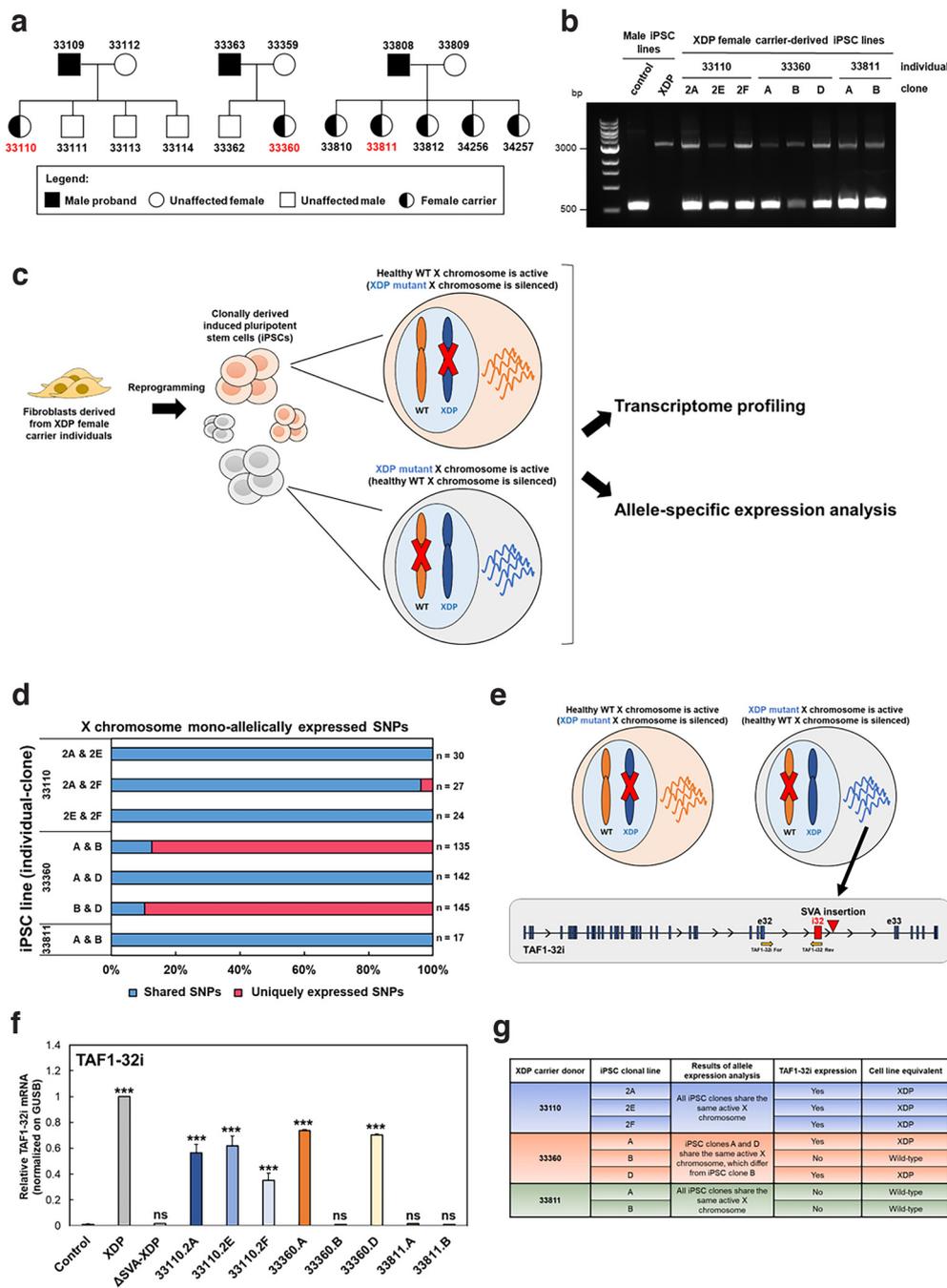
All data are presented in the manuscript and the Extended Data. Code and jupyter notebooks are available through GitHub at <https://github.com/apuapaquola>.

## Results

### Characterization of X-linked gene expression and XDP haplotype expression in female carrier-derived iPSCs

To model XDP using carrier derived iPSC lines, we characterized whether the mutant X chromosome or wild-type X chromosome are expressed in the eight iPSC lines previously derived from three heterozygous female carrier donors (Ito et al., 2016; Aneichyk et al., 2018; Fig. 1a). All eight XDP female carrier-derived iPSC lines carried the disease-specific haplotype on one of the X chromosomes, and we confirmed the 2.6-kb SVA retrotransposon insertion by long-range PCR amplification (Ito et al., 2016). The detection of two PCR products, of 0.6 kb and 3.2 kb, indicated that the samples were heterozygous for the control and XDP loci (Fig. 1b). Pluripotency analysis and BeadChIP genotyping array on all eight iPSC clones confirmed their primed pluripotent state, sample identity, and genomic integrity (Extended Data Fig. 1-2a-d).

To evaluate the XCI status of the eight XDP carrier iPSC lines, we analyzed gene expression profiles and allele-specific expression from stranded total RNA-seq (median 30 million paired-end reads per clone; Fig. 1c). To identify pairs of isogenic female iPSC lines with an identical genetic background, but expressing either the wild-type or XDP allele, we conducted an allele-specific expression analysis of expressed SNPs along the X chromosomes. For each set of clonal iPSC lines derived from the same female individual, we evaluated the percentage of X-linked monoallelically expressed SNPs that were shared or uniquely expressed between each clonal pair. When the same X chromosome is activated in two iPSC clones, most monoallelically expressed X-linked transcripts will be shared. In contrast, if different X chromosomes are activated, X-linked monoallelic transcripts will be expressed by different alleles in the two cell lines, and thus most SNPs will be uniquely expressed by each cell line.



**Figure 1.** Characterization of XDP female carrier-derived iPSCs for XCI status of the XDP haplotype. **a**, Family trees showing that the iPSC clones were derived by daughters/obligated carriers of XDP probands. **b**, Long-PCR products obtained from amplification of genomic DNA extracted from one control, one XDP mutant line, and all eight XDP female carrier-derived iPSC clones. Primers flanking the SVA insertion site were used to confirm the presence of the SVA retrotransposon in the genome. The control iPSC line shows only a product of ~599 bp (lower band), whereas the male XDP-derived iPSC line shows the predicted 3229 bp SVA product (upper band). Both PCR products were detectable in all eight XDP female carrier-derived iPSCs. **c**, Illustration of XDP female carrier-derived iPSCs. Because of clonal expansion of iPSCs and random XCI, female iPSC clones can inactivate two different X chromosomes. If the X chromosome carrying the healthy wild-type (WT) allele is active, the cell will express the canonical *TAF1* transcript; in contrast, if the X chromosome carrying the mutant XDP allele is active, the cell will express the mutant *TAF1* transcript, including the XDP haplotype. XDP female carrier-derived iPSCs were subjected to strand-specific RNA sequencing to perform gene expression analysis and allele-specific expression analysis. **d**, Bar plots of the fraction of allele-specific transcripts derived from the X chromosomes, which are shared between the different iPSC clones. Expression from the same or different alleles was determined by counting the number of shared and unique homozygous SNPs, respectively. The specific number of SNPs analyzed for each pair is shown to the right of the bar. **e**, Illustration of the mutant *TAF1* splice variant characterized by retention of intron 32 (*TAF1-32i*) close to the SVA insertion. We used a designed custom TaqMan primer/probe to detect the presence of the exon 32/intron 32 splice

continued

site (Aneichyk et al., 2018). Yellow arrows show that forward and reverse primers bind to exon 32 and intron 32, respectively, so that the probe spans the splice junction. **f**, The graph shows the relative expression of *TAF1-32i* in XDP female carrier-derived iPSCs in comparison to iPSCs derived from control, XDP, and relatively isogenic SVA-deleted lines ( $\Delta$ SVA-XDP). All values were normalized to the XDP sample. One-way ANOVA was performed on the mean and SEM from three independent experiments, and significance was determined as follows: \*\*\* $p \leq 0.001$ ; ns = not significantly different. **g**, Table summarizing features of the eight XDP female carrier-derived iPSCs used in this study. See also Extended Data Figures 1-1, 1-2, and 1-3.

**Figure contribution:** Ricardo S. Jacomini and Apua C. M. Paquola performed the allele-specific expression analysis. Laura D'Ignazio performed the long-PCR and TAF1-32i Taqman assay and analyzed the data.

We found that the three iPSC clones derived from the XDP female carrier 33360 were isogenic, with different active X chromosomes. Clones 33360.A and 33360.D shared the same active X chromosome, with all X-linked monoallelically expressed SNPs being shared. Clone 33360.B shared only 13% of monoallelically expressed SNPs with clone 33360.A and 10% with clone 33360.D, while 87% and 90% of X-linked monoallelically expressed SNPs were, respectively, uniquely expressed. This indicates 33360.B has a different active X chromosome compared with both clones 33360.A and 33360.D (Fig. 1d). Interestingly, clones 33360.A and 33360.D also exhibited high correlation with each other in the PCA analysis of normalized expression of X-linked genes from all XDP female carrier iPSC lines (Extended Data Fig. 1-3a).

In contrast, the three iPSC clones derived from carrier 33110 shared the same active X chromosome, and >96% of X-linked monoallelically expressed SNPs. Similarly, all iPSC clones from carrier 33811 had the same active X chromosome, with all X-linked monoallelically expressed SNPs being shared (Fig. 1d).

To further validate that allele-specific expression indicated an active X chromosome from XCI, we analyzed the allele-specific expression of chromosome 7, an autosome equivalently sized to the X chromosome. In all iPSCs, we observed expression from the same allele for chromosome 7-specific genes, indicating monoallelic expression because of eQTLs or imprinting (Extended Data Fig. 1-3b). Overall, the allele-specific expression analysis revealed that the iPSC clones derived from individual 33360 were isogenic XDP female carrier iPSC lines.

Next, we determined whether the active X chromosome in each XDP female carrier iPSC line represented the wild-type, or the XDP allele, by using total RNA-seq to quantify expression of the XDP-specific *TAF1-32i* transcript in the female carriers' transcriptomic profiles (Fig. 1e). No significant differences were detected among the eight clones, likely because of the few transcripts mapped to the region, and our use of the total RNA-seq method (compared with capture sequencing used previously; Aneichyk et al., 2018). Therefore, we used a *TAF1-32i*-specific TaqMan assay (Aneichyk et al., 2018) to measure the levels of this transcript in all female carrier iPSC lines. As expected, we detected *TAF1-32i* in XDP male iPSCs. We also detected *TAF1-32i* in five of the eight female carrier-derived iPSC lines, with levels of expression ~40–80% higher than those in the male control iPSCs ( $p < 0.001$ ), and ~35% lower than those in XDP patient-derived iPSCs. *TAF1-32i* was detected in clones 33360.A and 33360.D, but not in

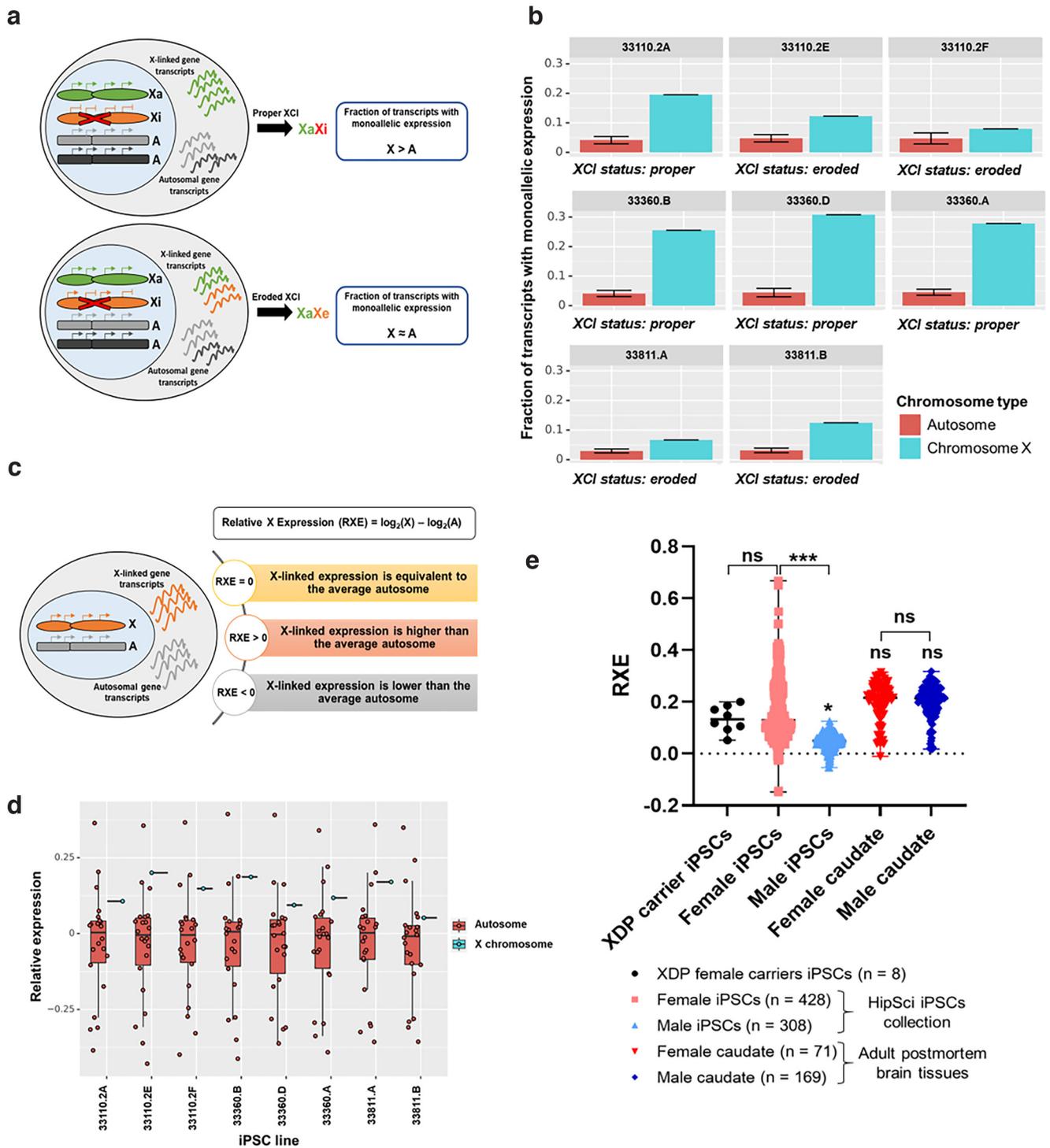
clone 33360.B, which was similar to the male control line (Fig. 1f). With this analysis, we defined the specific epigenetic identity of the XDP female carrier iPSC lines, with clone 33360.B being equivalent to wild-type cells, 33360.A and 33360.D resembling XDP male cells, all clonal lines from 33110 are XDP equivalent and all lines from individual 33811 are wild-type equivalent (Fig. 1g).

### Characterization of XCI in XDP female carrier-derived iPSCs

To assess epigenetic silencing of one X chromosome, we analyzed the total RNA-seq dataset for monoallelic and biallelic expression. Because of transcriptional silencing following X inactivation of one X chromosome, X-linked genes are expected to be transcribed mostly from one allele (monoallelic expression, with alternative allele frequency <0.1 or >0.9). In contrast, most autosomal genes are expressed biallelically from either maternal or paternal alleles, with alternative allele frequency  $\approx 0.5$ . Therefore, when XCI properly occurs, the fraction of transcripts with monoallelic expression is expected to be greater for X-linked genes than for autosomal genes.

We quantified the number of monoallelically expressed transcripts from the X chromosome compared with all autosomal chromosomes, as the fraction of homozygously expressed SNPs for a chromosome, defined as the ratio of the number of expressed SNPs with an allele frequency <0.1 and >0.9 from a chromosome over the total number of expressed SNPs for the chromosome (Fig. 2a). For all eight iPSC lines, the fraction of X-linked monoallelically expressed transcripts was greater than that of autosomal genes, confirming that XCI occurs. The fraction of X-linked monoallelically expressed genes varied significantly among iPSC lines. Four out of eight iPSC lines (33110.2A, 33360.A, 33360.B, and 33360.D) demonstrated 4- to 6-fold greater monoallelically expressed X-linked transcripts compared with autosomal transcripts, indicating transcription from a single active X (XaXi status). The remaining four iPSC lines (33110.2E, 33110.2F, 33811.A, and 33811.B) exhibited only a 1.5- to 2-fold increase in monoallelically expressed X-linked transcripts compared with autosomal transcripts, indicating partial erosion of the inactive X chromosome (XaXe status; Fig. 2b).

To investigate the expression dosage of X-linked genes on a chromosome-wide level, we compared the global transcriptional output of the X chromosome to that of autosomes. Briefly, we calculated the RXE, corresponding to chromosome-wide estimates of mean log-transformed TMP values of X-linked genes (X) compared



**Figure 2.** XDP female carrier-derived iPSCs have proper X chromosome dosage compensation. **a**, Diagram illustrating differences between “proper” and “eroded” XCI status. Proper XCI occurs when X-linked genes show higher monoallelic expression than do autosomal genes. Erosion of XCI occurs when the fraction of monoallelic transcripts is similar between the X chromosome and autosomes. **b**, Fraction of monoallelic expression from chromosome X compared with that from all autosomal chromosomes. An increased fraction of monoallelic expression indicates silencing of the X chromosome, and that proper XCI occurs in all XDP female carrier-derived iPSCs. The XCI status (proper or eroded) is reported for each iPSC clone. **c**, The relative expression of the X chromosome (RXE) is the difference between the log<sub>2</sub>-transformed mean TPM values of the X chromosome and all autosomes. If the expression of X and autosomes is equal, the RXE value will be 0, suggesting proper X dosage compensation. Positive RXE values indicate that the expression of X-linked genes is greater than that of autosomal genes; thus, dosage compensation is complete. Negative RXE values indicate that the expression of X-linked genes is lower than that of autosomes, and there is incomplete dosage compensation. **d**, Box plots showing log<sub>2</sub>-transformed relative expression of the X chromosome (RXE, blue) and all autosome

continued

chromosomes (red) for XDP female carrier-derived iPSCs. Red dots indicate the relative expression of each autosome pair (RGE) over all other chromosomes. **e**, RXE values of XDP female carrier-derived iPSC lines in comparison to female and male iPSC lines from public HipSci collection (Kilpinen et al., 2017) and adult female and male postmortem caudate nucleus samples from the BrainSeq Consortium (Collado-Torres et al., 2019; Benjamin et al., 2020). One-way ANOVA was performed on the mean and SEM calculated for each sample, and significance was determined as follows: \* $p \leq 0.05$ , \*\*\* $p \leq 0.001$ ; ns = not significant.

**Figure contribution:** Apua C. M Paquola and Ricardo S. Jacomini performed the allele-specific transcriptomic analysis. Apua C. M Paquola and Kynon J. M. Benjamin performed the dosage compensation analysis.

with autosomal genes (A; Duan et al., 2019). If the X and autosomal expression means are equal (RXE = 0), the average X-linked gene expression is comparable to the average expression from autosomal genes. If RXE > 0, the X chromosome is expressed at higher levels than the average autosome. RXE < 0 and > -1 indicated higher autosomal gene expression than X (Fig. 2c).

In all eight XDP female carrier-derived iPSCs, RXE values ranged from 0.05 to 0.20, suggesting similar levels of X-linked gene dosage in all lines (Fig. 2d). No statistically significant difference ( $p = 0.8399$ ) was observed when comparing the median RXE of the eight XDP female carrier iPSCs (median RXE = 0.1344) with the median RXE of the large publicly available HipSci collection of wild-type female iPSC lines (median RXE = 0.1662;  $n = 438$ ; Kilpinen et al., 2017). Aberrant eroded iPSC lines in the HipSci collection demonstrated an RXE value of up to 0.6, a value much higher than that of the XDP iPSCs. The RXE analysis demonstrated that the XDP carrier iPSC lines have X chromosome dosage compensation levels similar to those of most wild-type female iPSC lines. These HipSci lines maintain an inactive X (Bar et al., 2019). The median RXE of male iPSCs (median RXE = 0.0482;  $n = 308$ ; Kilpinen et al., 2017) was significantly lower than that of female iPSCs ( $p = 0.0428$  compared with XDP carrier iPSCs;  $p < 0.001$  compared with female HipSci iPSCs), because of genes escaping XCI.

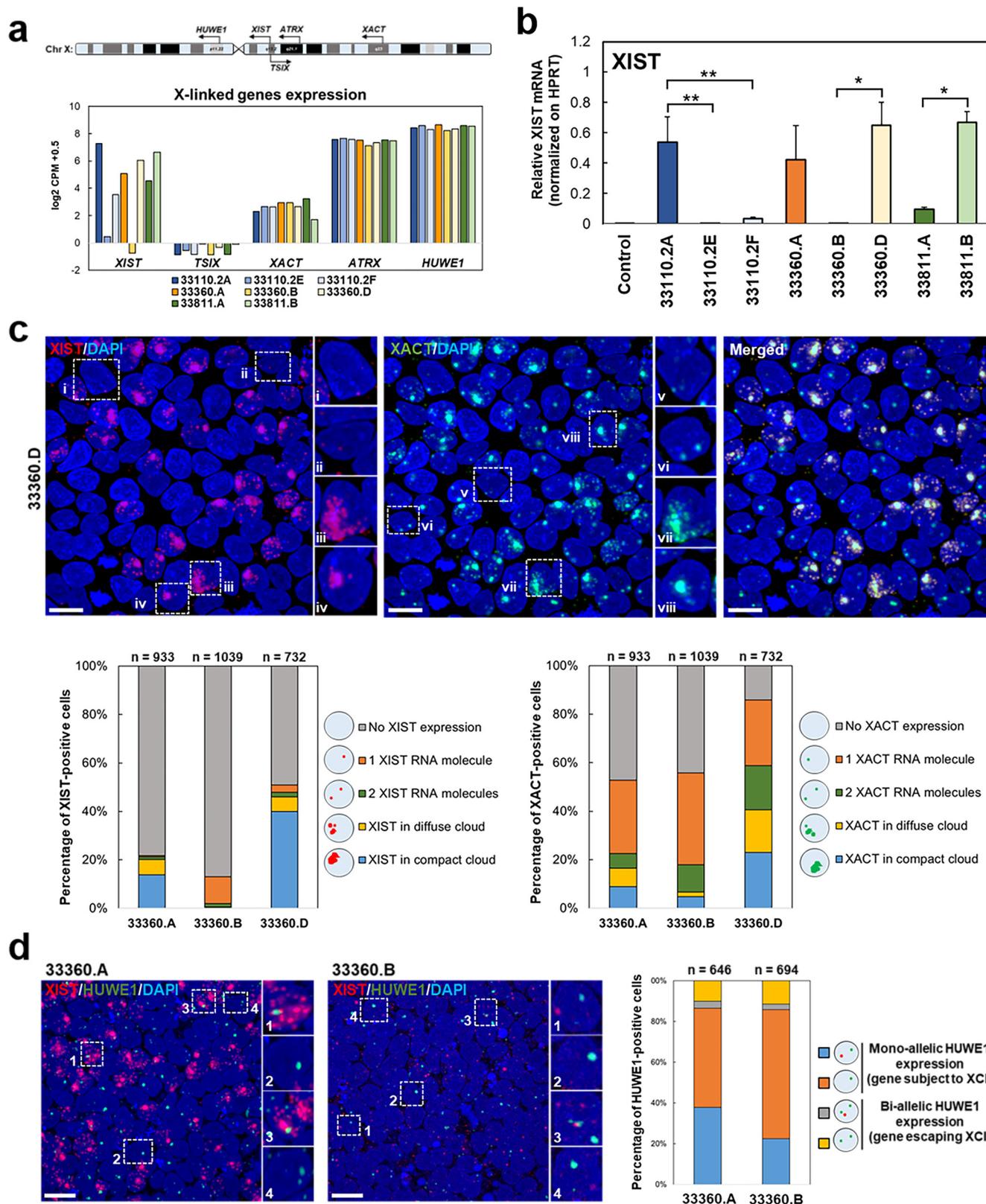
To measure dosage compensation in a postmortem brain tissue-specific context, we calculated the RXE for caudate nucleus samples (Collado-Torres et al., 2019; Benjamin et al., 2020), and found that median RXE values in both female (median RXE = 0.1985;  $n = 71$ ) and male (median RXE = 0.2048;  $n = 169$ ) caudate samples were slightly higher than those in wild-type female iPSCs from HipSci collection, without reaching statistical significance. No statistically significant difference was detected between female and male caudate samples ( $p = 0.9862$ ), suggesting full dosage compensation (Fig. 2e). In summary, the isogenic lines 33360.A, 33360.B, and 33360.D demonstrated proper XCI with monoallelic expression and dosage compensated X-linked gene expression.

### Isogenic XDP female carrier iPSC lines display variable XIST expression

Two long noncoding RNAs (lncRNAs), *XIST* and *XACT*, regulate human XCI. *XIST* (X-inactive coating transcript) coats the inactive X chromosome *in cis*, triggering gene silencing through epigenetic changes (Cantone and Fisher, 2017). *XACT* (X-active coating transcript) coats the active X chromosome in early preimplantation development (Vallot et al., 2013).

In the total RNA-seq data from the XDP female carrier-derived iPSCs, the X-linked transcripts *XACT*, *ATRAX* and *HUWE1* were expressed at similar levels among all iPSC lines (Fig. 3a). *TSIX*, the antisense transcript that negatively regulates *XIST* (Stavropoulos et al., 2001) in mice (but with unknown function in humans), was detected at extremely low levels in all iPSCs. Consistent with previous reports, *XIST* expression varied significantly, and even iPSC clones derived from the same XDP carrier individual exhibited statistically significant differences in *XIST* expression (Fig. 3a,b). Two out of eight iPSC lines, 33110.2E and 33360.B, exhibited minimal *XIST* mRNA expression that was comparable to that of a male iPSC line. The iPSC clones 33360.A and 33360.D, the mutant-equivalent cell lines of the isogenic set, showed *XIST* levels ~45–60% higher than the wild-type equivalent iPSC clone 33360.B. Our previous analyses demonstrated that monoallelic expression and dosage compensation occurred in all XDP female carrier-derived iPSC lines, despite low levels of *XIST* (Fig. 2). The *XIST*-primed iPSC state is thought to be an abnormal state that can proceed with erosion of X chromosome silencing; however, we demonstrate that global silencing of X was preserved, despite low levels of *XIST*.

While the transcriptomic data demonstrated that X-linked dosage compensation takes place at the population level of cells, we further assessed XCI at both single-cell and single-molecule resolution by using RNAscope *in situ* hybridization. Through canonical RNA FISH, *XIST* is usually visualized as small puncta or a single large cloud accumulating on the inactive X chromosome (Erwin and Lee, 2010). RNAscope technology allows for multiplex detection of target RNAs at a single-molecule resolution, with increased specificity and sensitivity compared with conventional multigene RNA-FISH that is detected using BAC probes (F Wang et al., 2012). Here, we identified four different patterns of *XIST* nuclear localization: (1) expression of one *XIST* RNA molecule when only a single *XIST*-specific fluorescent punctum was detectable; (2) expression of two *XIST* RNA molecules when two distinct *XIST*-specific fluorescent puncta were observed; (3) expression of multiple *XIST* RNA molecules forming a diffuse cloud when more than three dispersed *XIST*-specific fluorescent puncta were detected; and (4) expression of *XIST* as one condensed cloud (Fig. 3c; Extended Data Fig. 3-1a). Because of the high resolution of RNAscope, allowing single-molecule visualization of specific RNAs of interest, there are often additional dispersed *XIST* molecules in cells with one condensed cloud. For our quantitative analysis, these cells were included in the fourth category, one condensed cloud. *XIST*



**Figure 3.** Dosage compensation in female iPSCs inversely correlates with *XIST* expression. **a**, Simplified map of the *XIST*, *TSIX*, *HUWE1*, *ATRX*, and *XACT* loci on the X chromosome, and expression of X-linked gene transcripts in XDP carrier-derived iPSCs. Out of these genes, *XIST* shows higher variability among clones. **b**, RT-qPCR analysis of *XIST* expression in XDP female carrier-derived iPSCs compared with a control male iPSC line. Primers flanking exons 1 and 3 of *XIST* have been used. One-way ANOVA was performed on the mean and SEM from three independent experiments, and significance was determined as follows: \* $p \leq 0.05$ , \*\* $p \leq 0.01$ .

continued

**c**, Representative images from iPSC clone 33360.D showing localization of RNA scope-specific probes for *XIST* (red) and *XACT* (green) transcripts. DAPI was used to stain nuclei. Scale bar: 20  $\mu\text{m}$ . Based on the expression of those transcripts, different categories of nuclear localization were identified: (i/v) detection of one RNA molecule of *XIST/XACT*; (ii/vi) detection of two RNA molecules of *XIST/XACT*; (iii/vii) detection of several dispersed molecules of *XIST/XACT* transcripts forming a diffuse cloud; (iv/viii) detection of several molecules of *XIST/XACT* transcripts organized in a compact cloud. Bar charts report quantification of the four categories in the isogenic set of XDP female carrier-derived iPSCs. For each clone, >700 cells have been analyzed. The percentage of cells not expressing *XIST* or *XACT* transcripts is also reported. **d**, Representative images from iPSC clone 33360.A and 33360.B showing localization of RNA scope-specific probes for *XIST* (red) and *HUWE1* (green) transcripts. DAPI was used to stain nuclei. Scale bar: 20  $\mu\text{m}$ . The *HUWE1* gene is usually subject to XCI; thus, four different cellular conditions were identified: (1) monoallelic expression of *HUWE1* accompanied by *XIST* expression (in any form, i.e., single RNA molecules, multiple RNA molecules, diffuse clouds, compact cloud); (2) monoallelic expression of *HUWE1* in the absence of *XIST* expression; (3) biallelic expression of *HUWE1* accompanied by *XIST* expression (in any form, i.e., single RNA molecule, multiple RNA molecules, diffuse cloud, compact cloud); (4) biallelic expression of *HUWE1* in the absence of *XIST* expression. Bar graphs report the percentage of *HUWE1*-positive cells showing each of the four cellular conditions in the isogenic iPSC pair 33360.A and 33360.B. For each clone, >600 cells have been analyzed. See also Extended Data Figure 3-1.

**Figure contribution:** Kynon J. M. Benjamin analyzed the RNA-seq data. Laura D'Ignazio and Bareera Qamar performed and analyzed both the *XIST* qPCR analysis and the RNA scope.

expression and localization varied among the three isogenic iPSC clones, being detected in 22% of 33360.A iPSCs, 13% of 33360.B iPSCs, and 51% of 33360.D iPSCs. In iPSC clone 33360.B, 85% of cells exhibiting *XIST* expression expressed only one single *XIST* RNA molecule. In contrast, in iPSC clone 33360.A, 29% and 63% of *XIST*<sup>+</sup> cells exhibited *XIST* expression in diffuse clouds or compact clouds, respectively. Among cells showing *XIST* expression in iPSC clone 33360.D, 12% exhibited *XIST* expression in diffuse clouds, and 72% displayed *XIST* in compact clouds (Fig. 3c). These variations might be correlated with differences in the total *XIST* expression observed at the transcriptomic level.

Interestingly, we detected also four patterns of *XACT* nuclear localization: (1) one single RNA molecule; (2) two distinct RNA molecules; (3) multiple dispersed RNA molecules organized in a diffuse cloud; and (4) one large compact cloud was also detected using an *XACT*-specific fluorescent probe (Fig. 3c; Extended Data Fig. 3-1a). A small fraction of cells contained one condensed *XACT* cloud along with other dispersed smaller *XACT* molecules or one less intense *XACT* molecule. These cells were counted in the category of one condensed cloud. *XACT* coated 53% of 33360.A iPSCs, 56% of 33360.B iPSCs, and ~85% of 33360.D iPSCs. Nuclear colocalization of *XIST* and *XACT* was observed in all cells of iPSC clones 33360.A and 33360.D (Fig. 3c; Extended Data Fig. 3-1b). In contrast, in most cells from clone 33360.B, *XIST* and *XACT* had distinct nuclear localization (Extended Data Fig. 3-1b). Here, we report that the coaccumulation of *XIST* and *XACT* on the same X chromosome occurs not only in naive hESCs (Vallot et al., 2017), but also in human primed iPSC cultures.

To further evaluate X chromosome silencing at a single-cell resolution, we evaluated the single-molecule expression and localization of *HUWE1*, an X-linked gene subject to XCI, relative to *XIST* (Patel et al., 2017). If the X chromosome is silenced, *HUWE1* will be expressed from one X chromosome (monoallelic expression). Alternatively, if XCI mechanisms are impaired, *HUWE1* will be expressed by both X chromosomes (biallelic expression). For iPSC clones 33360.A and 33360.B, >80% of iPSCs demonstrated monoallelic expression of *HUWE1*, suggesting

proper XCI (Fig. 3d); only 13% of 33360.A iPSCs and 10% of 33360.B iPSCs displayed biallelic expression of *HUWE1*. ~49% of 33360.A iPSCs and 60% of 33360.B iPSCs monoallelically expressed *HUWE1* in the absence of *XIST* expression (Fig. 3d).

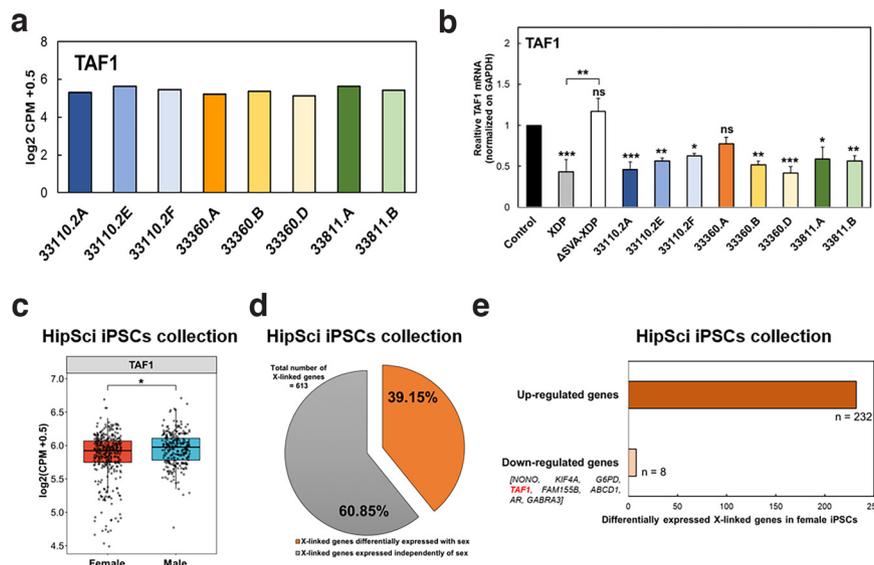
In summary, our results highlight the significant variations in *XIST* and *XACT* localization in human primed iPSC lines, as has been described previously. These findings demonstrate that X chromosome silencing and X-linked gene dosage compensation occur in isogenic XDP female carrier iPSC lines, independent of persistent *XIST* coating.

### Sexually dimorphic expression of *TAF1* in PSCs

Canonical *TAF1* expression was decreased in various XDP samples, including fibroblast, RNA from blood and NSCs, compared with control samples (Domingo et al., 2016; Ito et al., 2016; Aneichyk et al., 2018; Al Ali et al., 2021). To quantify the 3' region of canonical *TAF1* expression in XDP female carrier-derived iPSCs, we performed RT-qPCR with primers targeting *TAF1* exons 32 and 33 (Rakovic et al., 2018).

Unexpectedly, all eight XDP female carrier iPSC lines displayed similar levels of expression of the canonical full-length *TAF1* transcript (Fig. 4a), regardless of their expression of the mutant isoform *TAF1-32i* (Fig. 1g). Next, we compared *TAF1* expression in XDP female carrier lines to that in male control-derived and XDP-derived iPSCs, and XDP patient-derived iPSCs where the SVA was excised,  $\Delta\text{SVA-XDP}$  (Aneichyk et al., 2018; Fig. 4b). As expected, *TAF1* mRNA levels in XDP patient-derived iPSCs were ~50% lower than in control cells ( $p < 0.001$ ), and the expression normalized with SVA removal. Seven of the eight female carrier iPSC lines displayed a significant reduction in *TAF1* expression, compared with a control male iPSC line, with levels comparable to those of XDP iPSCs (Fig. 4b).

Decreased *TAF1* expression in carrier iPSCs could be caused by either an XDP female carrier iPSC molecular phenotype, or a sexually dimorphic expression in all female iPSCs. To distinguish between these two possibilities, we analyzed the publicly available iPSC transcriptome dataset from 301 individuals (Kilpinen et al., 2017). *TAF1* was significantly decreased in wild-



**Figure 4.** Sex-dependent expression of *TAF1* in female iPSCs. **a**, Normalized expression of the *TAF1* gene in XDP carrier-derived iPSCs. **b**, RT-qPCR analysis of *TAF1* expression in XDP female carrier-derived iPSCs compared with iPSCs derived from control, XDP, and a relatively isogenic SVA-deleted line ( $\Delta$ SVA-XDP). Primers flanking exons 32 and 33 of *TAF1* were used in this analysis. One-way ANOVA was performed on the mean and SEM from three independent experiments, and significance was determined as follows: \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ ; ns = nonsignificant. **c**, Box plots comparing normalized *TAF1* transcript expression levels in female (red) and male (blue) iPSCs from public HipSci collection (Kilpinen et al., 2017); \* $p \leq 0.05$ . **d**, Percentage of X-linked genes in female iPSCs (Kilpinen et al., 2017) differentially expressed with sex (orange) and expressed independently of sex (gray). **e**, Number of sex-differentially expressed X-linked genes that are upregulated or downregulated in female iPSCs (Kilpinen et al., 2017). *TAF1* is among the eight genes under-expressed in female iPSCs compared with male iPSCs.

*Figure contribution:* Laura D’Ignazio performed the *TAF1* qPCR analysis. Laura D’Ignazio and Kynon J.M. Benjamin analyzed the RNA-seq data from XDP female carrier-derived iPSCs and HipSci collection.

type female iPSCs compared with male cells ( $p = 0.0422$ ; Fig. 4c), indicating a sexually dimorphic expression of *TAF1*. X-linked genes are rarely decreased in females. A total of 240 X-linked genes were differentially expressed between sexes in human iPSCs (Fig. 4d). Eight genes, including *TAF1*, were downregulated in female iPSCs, while 232 of these differentially expressed genes were upregulated ( $\log_{2}FC = 0.10201$ ;  $p = 0.0422$ ; Fig. 4e). In summary, *TAF1* was decreased in control female iPSC lines, and decreased female expression was uncommon for X-linked genes.

### Striatal regions have equal levels of *TAF1* expression in males and females

To evaluate whether *TAF1* expression was decreased in the brain of adult females compared with males, we further investigated the recently reported sex expression differences in adult brains (Oliva et al., 2020). In the GTEx analysis, *TAF1* expression was significantly decreased (with LFSR  $< 0.05$ ) in 26 female tissues compared with males, including eight brain regions, such as amygdala, anterior cingulate cortex, cortex, frontal cortex, hippocampus, hypothalamus, spinal cord, and substantia nigra (Extended Data Fig. 5-1a). Importantly, no significant difference in *TAF1* expression between females and males was observed in the caudate nucleus, putamen or nucleus accumbens.

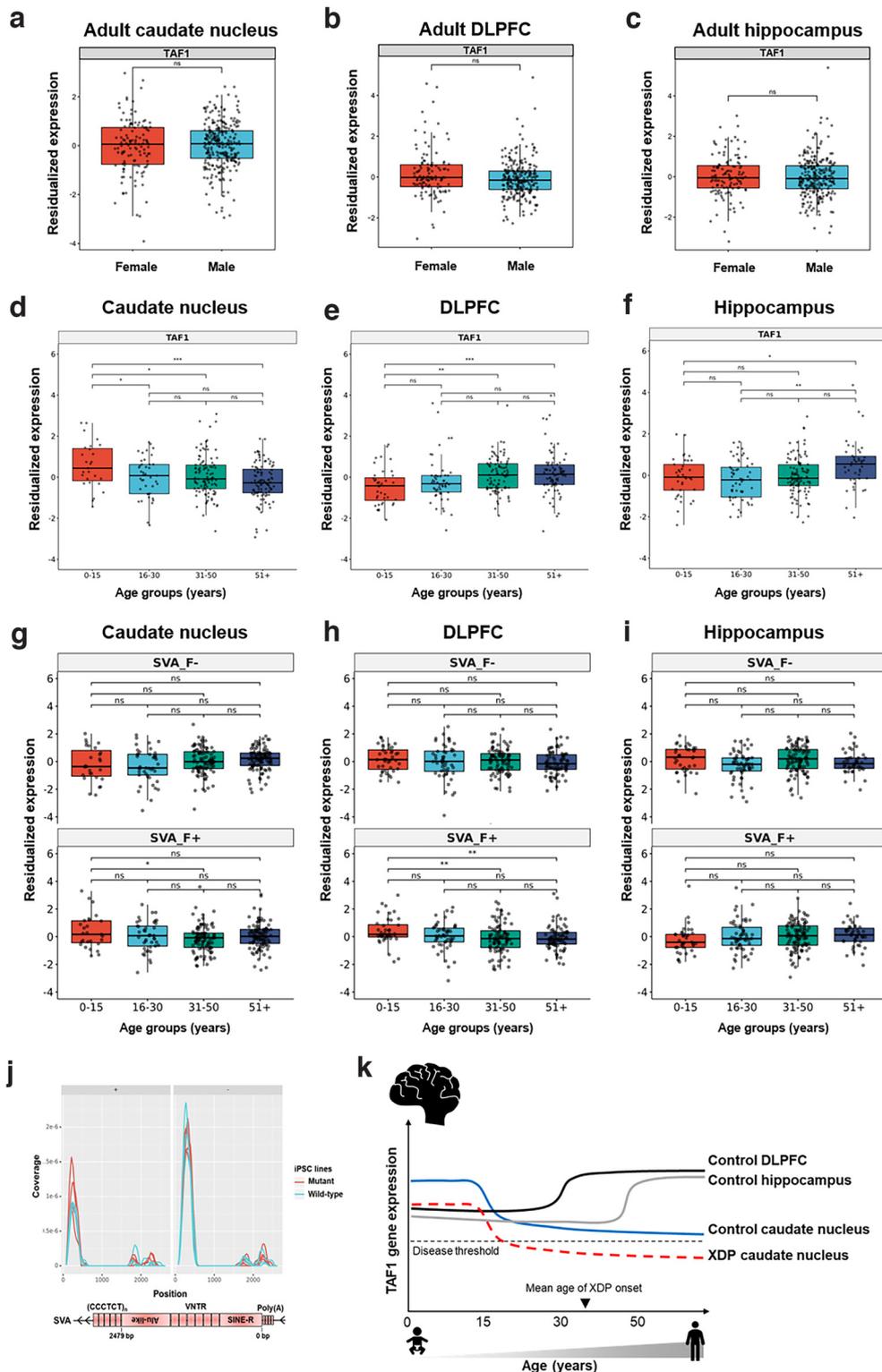
We further evaluated *TAF1* expression in neurotypical controls included in the BrainSeq Consortium dataset,

and confirmed that *TAF1* is not sexually dimorphic in the caudate nucleus (Fig. 5a). In the BrainSeq Consortium dataset, *TAF1* was not significantly sexually differentially expressed in the DLPFC (Fig. 5b) or hippocampus (Fig. 5c; a finding which contrasted with the decreased female expression in the GTEx samples, described in the preceding paragraph, possibly because of sample variation in age or analysis pipelines). As a positive control, a significant increase in *XIST* expression was consistently observed between female and male tissue samples ( $p < 0.001$ ; Extended Data Fig. 5-1b).

In summary, *TAF1* was decreased in most wild-type female cells and tissues compared with males, including wild-type iPSC cells, mutant carrier XDP iPSC cells and female postmortem brains. However, *TAF1* was expressed equally in males and females in caudate nucleus, putamen, and nucleus accumbens, indicating region-specific regulation of *TAF1*.

### *TAF1* is decreased in adult postmortem caudate compared with young individuals

Next, we investigated *TAF1* expression in postmortem brain tissue over the lifespan, using recent lifespan neurogenomic datasets from the BrainSeq Consortium (Fromer et al., 2016; GTEx Consortium et al., 2017; Collado-Torres et al., 2019). By analyzing the total RNA-seq dataset of caudate nuclei from neurotypical individuals aged 0–90 years (Collado-Torres et al., 2019; Benjamin et al., 2020), we found that



**Figure 5.** *TAF1* and SVA-F expression over the lifespan in the postmortem brain. **a–c**, Box plots comparing residualized gene expression of *TAF1* in **(a)** the caudate nucleus ( $n=394$ ), **(b)** DLPFC ( $n=379$ ), and **(c)** hippocampus ( $n=376$ ) of female (red) and male (blue) individuals, from the BrainSeq Consortium (Collado-Torres et al., 2019; Benjamin et al., 2020); ns = not significant. **d–i**, Box plots comparing residualized gene expression of individuals from the BrainSeq Consortium (Collado-Torres et al., 2019; Benjamin et al., 2020) for: **(d)** *TAF1* in the caudate nucleus, **(e)** *TAF1* in the DLPFC, **(f)** *TAF1* in the hippocampus, **(g)** SVA-F in the caudate nucleus, **(h)** SVA-F in the DLPFC, **(i)** SVA-F in the hippocampus. Samples were divided into four groups based on the age of the individuals: 0–15 years; 16–30 years; 31–51 years; and older than 51 years; ns = not significant; \*FDR  $\leq 0.05$ , \*\*FDR  $\leq 0.01$ , \*\*\*FDR  $\leq 0.001$ . **j**, SVA expression in iPSC cells. Plot shows normalized RNA-Seq coverage along the SVA sequence for both sense (+) and antisense (–) strands. Mutant trace (red)

*continued*

indicates female XDP carrier lines expressing the XDP allele: 33110.2A, 33110.2E, 33110.2F, 33360.A, 33360.D. Wild-type trace (light blue) indicates female XDP carrier lines expressing the wild-type X chromosome: 33360.B, 33811.A, 33811.B. The schematic illustration describes the structure of the human SVA retrotransposons. **k**, Diagram depicting our proposed model to explain the differential degeneration in the neostriatum of individuals with XDP. The decrease in *TAF1* expression after 15 years of age may synergize with a XDP-specific partial loss of *TAF1* function in the caudate nucleus, a brain region that seems to be selectively sensitive to *TAF1* levels. See also Extended Data Figures 5-1, 5-2, 5-3, and 5-4.

**Figure contribution:** Ria Arora performed the sex differential expression of *TAF1*. Kynon J. M. Benjamin performed the age-differential expression of *TAF1* across lifespan. Taylor Evans performed the SVA-F differential expression analysis across lifespan. Apua C. M. Paquola and Jennifer A. Erwin performed the SVA-derived transcript analysis on iPSCs.

*TAF1* gene expression significantly decreased after 15 years of age (Fig. 5d). Individuals aged 0–15 years had the highest level of *TAF1* expression compared with all other age groups [16–30 (FDR < 0.05), 31–50 (FDR < 0.05), and 51+ years (FDR < 0.0005); Extended Data Fig. 5-3]. Notably, increased *TAF1* expression occurred throughout the transcript. When examining *TAF1* expression at the exon level, 25 out of 38 exons located throughout the entire transcript were significantly increased in young individuals. The remaining 13 exons were excluded from this analysis because of their low expression (below the threshold for accurate quantification; Extended Data Fig. 5-4).

*TAF1* expression demonstrated an opposite age association in the DLPFC and hippocampus. In the DLPFC, there was higher expression in 31–50 year olds and 51+ compared with 0–15 year olds (Fig. 5e). In the hippocampus, *TAF1* expression was higher in the 51+ age group compared with younger groups (Fig. 5f). Considering the reported onset of XDP at approximately age 35, the specific decrease in *TAF1* expression after 15 years of life in the caudate nucleus precedes the age of onset of symptoms (Fig. 5k). Taken together, these findings support the existence of age-specific, tissue-specific, and sex-specific fine regulatory mechanisms controlling *TAF1* expression.

### Evaluating SVA-F expression

In addition to variation in *TAF1* expression, regulation of SVA expression in the brain may also influence the XDP phenotype. While neurotypical brains do not have the patient-specific SVA-F insertion in *TAF1* intron 32, other SVA-F copies with identical sequences are present in all humans and these sequences are regulated by cellular host factors in trans. Therefore, we evaluated SVA-F expression in general in postmortem brain samples from BrainSeq, which employs a stranded total RNA-seq protocol. Because SVA-F elements are repetitive and have identical sequences, we quantified all transcripts mapping to any SVA-F location in the genome, the human-specific subfamily involved in XDP (Hancks and Kazazian, 2010).

The expression of the SVA-F retrotransposon was not sexually dimorphic in the caudate nucleus of healthy individuals from the BrainSeq dataset (Extended Data Fig. 5-2). In general, SVA-F plus strand and minus strand expression showed slight or no significant difference between most age groups for the caudate nucleus, DLPFC and hippocampus (Fig. 5g–i). No significant differences were seen for any comparison of SVA-F minus strand expression. In the caudate nucleus and DLPFC, SVA-F plus

strand expression was slightly but significantly decreased in 31- to 50-year-old individuals, compared with 0- to 15-year-old samples (FDR < 0.01). In DLPFC, decreased SVA-F plus strand expression was also significantly decreased in 51+ (FDR < 0.01) samples, compared with 0- to 15-year-old group. In summary, we found abundant expression of SVA-F in neurotypical brains, slightly decreased SVA-F expression in caudate nucleus and DLPFC after 30 years of life and no differential expression of SVA-F between sexes.

Transcripts from the XDP-specific SVA and the hexanucleotide repeat could drive neurodegeneration in XDP. To evaluate the potential for disease-associated SVA-derived transcripts, we analyzed SVA expression in XDP female carrier iPSCs. We identified abundant SVA-derived transcripts (Fig. 5j). However, because of the multicopy nature of SVA-F, it was unclear whether these transcripts arose from XDP-specific SVA or another SVA-F. Most transcripts aligned to the SINE-R and Alu-like regions in both sense and antisense orientation, and aligned equally well to the XDP-specific SVA-F and ~4–10 other SVA-F regions in the genome. In addition, transcripts were not significantly differentially expressed between wild-type-equivalent or mutant-equivalent XDP carrier iPSCs (Fig. 5j).

To detect XDP-specific SVA transcripts, we created a modified reference genome by introducing the XDP-specific SVA sequence at its corresponding position in chromosome X in hg38. We aligned the RNA-Seq reads to this modified reference genome, and assessed sequencing reads that aligned to the XDP-specific SVA. We found weak evidence for XDP-specific SVA-derived transcripts. In one mutant iPSC sample, one sequencing read spanned the Alu-like region to the hexamer repeat, and preferentially aligned to the XDP-specific SVA, because of a 1 bp difference unique to the XDP SVA compared with all other SVA-Fs. No transcripts or junctions were detected that spliced into the SVA from the *TAF1* region. We attribute the weak evidence for XDP-specific SVA-derived transcripts, to the scarcity of transcripts and the redundancy of the sequence. In summary, we found abundant expression of SVA-F in iPSC lines regardless of XDP status, and weak evidence of XDP-specific SVA transcripts in iPSC cells.

### Discussion

Our results demonstrate significant sex-specific, age-specific, and tissue-specific differences in XDP-associated transcript expression in female iPSCs and in human

postmortem brain. *TAF1* was moderately decreased in most female cells and tissues compared with males, including wild-type iPSCs, carrier XDP iPSCs, and in several regions of the female postmortem brain; *TAF1* was expressed equally in the male and female striatum. *TAF1* expression uniquely decreased after 15 years of age in the caudate nucleus but not in the hippocampus or DLPFC. SVA-F plus strand expression slightly decreased in caudate nucleus and DLPFC after 30 years of life in neurotypical brain, but we find weak evidence for XDP-specific SVA derived transcripts in female iPSC cells. These findings indicate that regional-specific, age-specific, and sex-specific mechanisms regulate *TAF1* and SVA-F expression, further highlighting the importance of disease-relevant cellular models and postmortem tissue analysis.

### Analysis of XCI status is crucial for *in vitro* disease modeling

As a valuable cell resource for disease modeling, we extensively characterized eight iPSC lines from XDP female carrier individuals, and identified 5 clonal lines that are expressing the XDP haplotype and three lines that are expressing the wild-type X chromosome. From one individual, we identify isogenic lines where one clonal iPSC line expressed the wild-type X, and the two other clonal iPSC lines expressed the XDP haplotype. In our study, we provided strong evidence that monoallelic expression and dosage compensation occur in all XDP female carrier-derived iPSCs here examined, independently of *XIST* RNA levels or *XIST/XACT* nuclear location patterns. This finding is consistent with previous analyses conducted on large cohorts of female iPSC lines and demonstrating monoallelic X-linked expression and dosage compensation for the majority of X-linked genes (Shen et al., 2008; Silva et al., 2008; Mekhoubad et al., 2012; Vallot et al., 2013; Bar et al., 2019; Patrat et al., 2020).

The variability of XCI status in primed iPSCs remains a significant challenge for female stem cell models. Dosage compensation state for primed pluripotency comprises: XCI with *XIST* coating a single silenced X, heterochromatinization of the silent X, and monoallelic expression for most X linked genes (Shen et al., 2008; Silva et al., 2008; Mekhoubad et al., 2012). The XCI state often erodes in primed pluripotent human ES cells, such that portions of the previously inactive X chromosome become reactivated and are not dosage compensated (Shen et al., 2008; Silva et al., 2008; Mekhoubad et al., 2012). The naive state is a pluripotent state that captures an earlier developmental state and can reset the XCI state of primed human stem cells (Sahakyan et al., 2017). However, the naive state is characterized by increased genomic instability and altered imprinting (Theunissen et al., 2016), and also require additional culture optimization to provide more suitable cells for disease modeling.

Differentiating the set of isogenic XDP female carrier-derived iPSC into specific neural lineages will help elucidate the molecular mechanism causing XDP, as these models may capture early stages of disease no longer present in end-state postmortem analysis. It

has consistently been demonstrated that the differentiation of iPSC lines maintains the dosage compensation status during differentiation (Silva et al., 2008; Mekhoubad et al., 2012). We anticipate that the XDP female carrier iPSC lines described in this study should maintain XCI during neural differentiation. Importantly, we demonstrate absence of *XIST* expression in the majority of female carrier primed iPSCs, which indicates that the inactive X chromosome may further erode in some cultures. Therefore, careful analysis of the XCI status of future iPSC-derived neuronal models will be essential.

### XDP in female carriers and isogenic carrier-derived iPSCs

In this study, we characterized eight iPSC lines obtained from three XDP heterozygous female carriers (Ito et al., 2016; Aneichyk et al., 2018). Our analyses demonstrated that clonal carrier-derived iPSC lines containing the silent wild-type, and the active mutant allele, retain molecular features of XDP mutant lines, such as increased expression of the XDP-associated aberrant transcript TAF1-32i. In contrast, female lines containing the active wild-type allele and silenced mutant allele had low levels of TAF1-32i equivalent to control and CRISPR-deleted SVA iPSC samples. A recent study demonstrated that TAF1-32i expression was elevated in XDP patients and female carriers compared with control samples but TAF1-32i expression levels did not distinguish carriers versus affected XDP patients (Al Ali et al., 2021). Our data suggest that the XDP haplotype is fully silenced by XCI and that TAF1-32i expression is a biomarker for XDP haplotype expression in female iPSCs. Future investigations will be needed to determine whether TAF1-32i expression is a biomarker for XDP haplotype expression in carrier samples and differentiated cells.

Carrier-derived cell lines and tissues are a precious resource to understand XDP pathology. For example, female heterozygous XDP tissues that are mosaic for functionally wild-type or mutant cells can reveal cell intrinsic versus cell extrinsic mechanisms. If neurodegeneration is limited to functionally mutant neurons, then a cell intrinsic mechanism is necessary for degeneration. In contrast if mutant and wild-type neurons degenerate similarly but at a reduced rate compared with male XDP, then cell extrinsic mechanisms, such as network activity properties, degree of inflammation and/or secreted toxic proteins, are likely more important for degeneration than cell intrinsic mechanisms.

The limited number of females with severe XDP examined suggests that mosaic XCI might be a protective factor. An extremely skewed (98%:2%) XCI was noted in only a single case with full XDP syndrome (Domingo et al., 2014), and X chromosome monosomy (45,X/46,XX) was found in a subset of cells of another female patient with XDP (Westenberger et al., 2013). A recent study examining the peripheral blood of carrier females is also consistent with a mosaic XCI status being protective (Al Ali et al., 2021). Decreased full-

length *TAF1* expression, as measured by the ratio of *TAF1* 3'/5', is at an intermediate level for asymptomatic carriers such that XDP patients demonstrated the most reduction and asymptomatic female carriers demonstrated intermediate reduction compared with healthy controls.

Here, for the first time, we reported a sex-differential expression of the *TAF1* gene in stem cell models. We found that XDP female carrier iPSCs as well as a large cohort of female wild-type iPSCs displayed significantly reduced levels of *TAF1* gene, in comparison to male control iPSCs. In our present study, we did not investigate whether there is a correlation between *TAF1* changes at mRNA and protein levels in female cells. However, a decreased protein expression, in accordance with reduced *TAF1* mRNA expression, was detected in both XDP iPSCs and NSCs, compared with control samples, although the decrease was not statistically significant (Aneichyk et al., 2018; Rakovic et al., 2018).

Decreased expression of X-linked genes is not usually observed in females (Disteche, 2012). Additional studies are needed to determine the downstream effects of sex-linked variation of *TAF1* gene expression in iPSCs, particularly in light of its crucial involvement in transcription initiation, and in the regulation of cellular processes, such cell growth (Sekiguchi et al., 1991) or progression through G<sub>1</sub> phase of cell cycle (O'Brien and Tjian, 1998).

### ***TAF1* expression variation in human postmortem brains and XDP**

The specific role of *TAF1* in the striatum and how it relates to regional differences in pathology has not yet been defined (Bragg et al., 2019). *TAF1* is required for postnatal maintenance of striatal cholinergic interneurons in mice (Cirnaru et al., 2021). In humans, rare loss of function mutations in *TAF1* cause intellectual disability (O'Rawe et al., 2015; Hurst et al., 2018). Importantly, XDP patients samples have a moderate decrease of *TAF1* (~29% decrease from controls; Al Ali et al., 2021), which is less severe than the loss of function mutations.

A number of reports have demonstrated decreased *TAF1* exon expression downstream of the SVA insertion, both in patient samples and in *in vitro* cellular models (Makino et al., 2007; Domingo et al., 2016; Ito et al., 2016; Aneichyk et al., 2018; Al Ali et al., 2021). However, there are inconsistent reports about which *TAF1* transcripts are mainly altered in XDP: the most prominent *TAF1* isoform, c*TAF1*, or the neuron-specific n*TAF1* which varies from c*TAF1* by only an additional 6 bp derived from the alternative exon 34' (Makino et al., 2007; Ito et al., 2016; Capponi et al., 2021).

In this study, we have interrogated *TAF1* expression variations in both stem cell models and human postmortem brains. Gene-level quantification, which sums the quantification of all *TAF1* transcripts and does not distinguish transcripts or isoforms, was used for all RNAseq analysis. For analysis of GTEx and BrainSeq consortium datasets of postmortem brain, we further investigated differential expression at the individual exon level and found

that all exons throughout *TAF1* with sufficient expression for rigorous statistical analysis ( $n = 25$ ) followed the same trend of expression as the gene level quantification. Lower sequencing depth of the iPS samples and 13 exons in the postmortem samples prohibited exon level analysis.

Our data also provided strong evidence that *TAF1* expression varies significantly: between sexes, among different brain regions, and across the life span in humans. The decrease in females is generally modest and of unknown significance, but it is an important consideration when comparing male and female subjects. Of particular relevance to XDP, we found that *TAF1* expression decreased after age 15, specifically in the caudate nucleus. It is important to note that all transcriptome analyses were performed on publicly available datasets generated from bulk tissues, composed of heterogeneous cell types, which do not allow to investigate any differential expression at cellular level.

To our knowledge, there are no other reports of differential age-associated *TAF1* and SVA expression in human postmortem brain tissues. Previously, expression of *TAF1* throughout development and aging was only investigated in mouse brains using quantitative RT-PCR analysis (Jambaldorj et al., 2012). In that study, the authors reported the highest levels of *Taf1* at the embryonic stage, with a gradual reduction till three postnatal weeks, followed by a stable expression of *Taf1* transcript till forty postnatal weeks. While other brain regions including the cerebellum are affected in XDP, the analysis here are restricted to the publicly available human brain lifespan postmortem datasets of GTEx and BrainSeq, which do not include early life and adolescent periods in other regions. Therefore, future work will be required to evaluate whether *TAF1* expression decreases during adolescence in other vulnerable tissue.

### **SVA-derived transcripts and XDP**

In addition, abundant evidence shows that SVA-driven pathologic events lead to XDP dysregulation, although the precise mechanisms underlying the pathophysiology remains elusive. It is proposed that SVA alters local chromatin structure and forms G-quadruplex DNA (G4; Hancks and Kazazian, 2010), leading to transcriptional interference impeding the progression of RNA polymerase II in XDP samples (Bragg et al., 2017). Also, the SVA insertion includes a hexameric (CCCTCT)<sub>n</sub> repeat of variable length, which is inversely correlated with age of disease onset in individuals with XDP (Bragg et al., 2017; Westenberger et al., 2019). These findings directly correlate genetic variation in individuals with XDP to differences in their manifestation of clinical disease, thereby supporting a causal role for the SVA insertion in disease pathogenesis.

Therefore, we hypothesized that the regulation of SVA-F expression in the brain may also influence the XDP manifestations. Our data showed slight but significantly decreased SVA expression after 30 years of life in the DLPC and caudate nucleus. No significant difference between sexes or other age groups were observed.

Approximately 2700 SVA retrotransposon copies are present in the human genome (Hancks and Kazazian, 2010), creating inherited genetic diversity by actively mutagenizing the genome, and contribute the vast majority of GC-rich human-specific tandem repeat expansions in the human genome (Sulovari et al., 2019). While neurotypical brains do not have the patient specific SVA-F insertion in TAF1 intron 32, other SVA-F copies with identical sequences are present in all humans. These identical SVA sequences are regulated by cellular host factors in trans by members of the large family of protein coding Kruppel-associated box zinc finger proteins (KZFP) bind SVA elements to dock KRAB-associated protein 1 (KAP1/TRIM28; Rowe et al., 2010). KZFP and KAP/TRIM28 are generally lowly expressed in postnatal striatum, for example the caudate, basal ganglia and putamen are among the tissue with the lowest expression of TRIM28, ZNF587, ZNF91/93 in GTEx (GTEx Consortium, 2020). The SVA region harbors extensive DNA methylation in XDP patient blood and the level of methylation is reduced in the basal ganglia of the single patient tissue examined thus far (Lüth et al., 2022). Our findings suggest dynamic regulation of SVA elements in the adult caudate nucleus. It is uncertain which factors cause the decreased SVA-F expression after 30 years, but this regulation may modify XDP phenotype.

Currently, the hypothesized causal path is that SVA alters TAF1 and the altered TAF1 leads to XDP phenotype (SVA → TAF1 → pathology). It is possible that XDP specific SVA-derived transcripts directly affect TAF1 and/or that SVA derived transcripts directly affect pathology (SVA → TAF1). Because of the multicopy nature of SVA-F, no reports have conclusively detected XDP-specific SVA-derived transcripts. Here, we specifically used stranded total RNA sequencing to detect any XDP specific transcripts. We find weak evidence for the presence of any XDP specific transcripts. Therefore, we conclude SVA-derived transcripts are either not expressed or below the limits of detection.

In conclusion, we unveiled sex-specific, tissue-specific, and age-specific regulation of *TAF1* in human brain. We found that, uniquely in the caudate nucleus, where *TAF1* expression is not sexually dimorphic, *TAF1* expression significantly decreases after adolescence. Although these findings were obtained analyzing brain tissues from neurotypical individuals, we hypothesize that the decrease of *TAF1* after 15 years of age may have implications for the XDP adult neostriatum. In our proposed model, the decrease in *TAF1* after adolescence in the caudate nucleus may synergize with the XDP-specific partial loss of *TAF1* function, causing an even further loss of *TAF1*, and thereby triggering degeneration in adult life. This model is consistent with the late onset of XDP neurodegeneration, with patients appearing neurologically unaffected until 35–40 years of age, and with the female carriers only showing attenuated symptoms because of their mosaic neostriatum.

To better frame the role of *TAF1* in the XDP neurodegeneration, and to inform about the causal mechanisms leading to XDP pathology, it is crucial to perform extensive transcriptomic analysis on XDP postmortem brain

tissues. To capture novel cellular phenotypes and mechanisms underlying the striatal degeneration, further studies are needed to differentiate the isogenic XDP female carrier iPSCs reported here, together with patient-derived iPSCs into specific neuronal models. Such studies will help us better understand the impact of *TAF1* loss in iPSC-derived striatal models, as well as the direct or indirect contribution of SVA-driven pathologic mechanisms.

## References

- Al Ali J, et al. (2021) TAF1 transcripts and neurofilament light chain as biomarkers for X-linked dystonia-parkinsonism. *Mov Disord* 36:206–215.
- Aneichyk T, et al. (2018) Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* 172:897–909.e21.
- Arasaratnam CJ, Singh-Bains MK, Waldvogel HJ, Faull RLM (2021) Neuroimaging and neuropathology studies of X-linked dystonia parkinsonism. *Neurobiol Dis* 148:105186.
- Bar S, Seaton LR, Weissbein U, Eldar-Geva T, Benvenisty N (2019) Global characterization of X chromosome inactivation in human pluripotent stem cells. *Cell Rep* 27:20–29.e3.
- Benjamin KJ, Feltrin AS, Barbosa AR, Jaffe AE, Stolz J, Collado-Torres L, Huuki-Myers LA, Burke E, Shin JH, Ulrich WS, Deep-Soboslay A, Tao R, Hyde TM, Kleinman JE, Erwin JA, Weinberger DR, Paquola AC (2020) Genetic and environmental regulation of caudate nucleus transcriptome: insight into schizophrenia risk and the dopamine system. medRxiv.
- Blood AJ, Waugh JL, Münte TF, Heldmann M, Domingo A, Klein C, Breiter HC, Lee LV, Rosales RL, Brüggemann N (2018) Increased insula-putamen connectivity in X-linked dystonia-parkinsonism. *Neuroimage Clin* 17:835–846.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- 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.
- Bragg DC, Sharma N, Ozelius LJ (2019) X-linked dystonia-parkinsonism: recent advances. *Curr Opin Neurol* 32:604–609.
- Brüggemann N, Heldmann M, Klein C, Domingo A, Rasche D, Tronnier V, Rosales RL, Jamora RDG, Lee LV, Münte TF (2016) Neuroanatomical changes extend beyond striatal atrophy in X-linked dystonia parkinsonism. *Parkinsonism Relat Disord* 31:91–97.
- Cantone I, Fisher AG (2017) Human X chromosome inactivation and reactivation: implications for cell reprogramming and disease. *Philos Trans R Soc Lond B Biol Sci* 372:20160358.
- Capponi S, Stöffler N, Penney EB, Grütz K, Nizamuddin S, Vermunt MW, Castelijn B, Fernandez-Cerado C, Legarda GP, Velasco-Andrada MS, Muñoz EL, Ang MA, Diesta CCE, Creyghton MP, Klein C, Bragg DC, De Rijk P, Timmers HTM (2021) Dissection of TAF1 neuronal splicing and implications for neurodegeneration in X-linked dystonia-parkinsonism. *Brain Commun* 3:fcab253.
- Cirnar M-D, 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.
- Collado-Torres L, Burke EE, Peterson A, Shin J, Straub RE, Rajpurohit A, Semick SA, Ulrich WS; BrainSeq Consortium, Price AJ, Valencia C, Tao R, Deep-Soboslay A, Hyde TM, Kleinman JE, Weinberger DR, Jaffe AE (2019) Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron* 103:203–216.e8.
- Disteche CM (2012) Dosage compensation of the sex chromosomes. *Annu Rev Genet* 46:537–560.

- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- 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.
- Domingo A, Westenberger A, Lee LV, Brænne I, Liu T, Vater I, Rosales R, Jamora RD, Pasco PM, Cutiongco-Dela Paz EM, Freimann K, Schmidt TG, Dressler D, Kaiser FJ, Bertram L, Erdmann J, Lohmann K, Klein C (2015) New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3). *Eur J Hum Genet* 23:1334–1340.
- Domingo A, Amar D, Grütz K, Lee LV, Rosales R, Brüggemann N, Jamora RD, Cutiongco-Dela Paz E, Rolfs A, Dressler D, Walter U, Krainc D, Lohmann K, Shamir R, Klein C, Westenberger A (2016) Evidence of TAF1 dysfunction in peripheral models of X-linked dystonia-parkinsonism. *Cell Mol Life Sci* 73:3205–3215.
- Duan JE, Flock K, Jue N, Zhang M, Jones A, Seesi SA, Mandoiu I, Pillai S, Hoffman M, O'Neill R, Zinn S, Govoni K, Reed S, Jiang H, Jiang ZC, Tian XC (2019) Dosage compensation and gene expression of the X chromosome in sheep. *G3 (Bethesda)* 9:305–314.
- Erwin JA, Lee JT (2010) Characterization of X-chromosome inactivation status in human pluripotent stem cells. *Curr Protoc Stem Cell Biol* Chapter 1:Unit 1B.6.
- 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.
- Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048.
- Fromer M, et al. (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* 19:1442–1453.
- Goto S, Lee LV, Munoz EL, Tooyama I, Tamiya G, Makino S, Ando S, Dantes MB, Yamada K, Matsumoto S, Shimazu H, Kuratsu J, Hirano A, Kaji R (2005) Functional anatomy of the basal ganglia in X-linked recessive dystonia-parkinsonism. *Ann Neurol* 58:7–17.
- 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.
- GTEX Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369:1318–1330.
- GTEx Consortium, et al. (2017) Genetic effects on gene expression across human tissues. *Nature* 550:204–213.
- Hancks DC, Kazazian HH (2010) SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol* 20:234–245.
- Hanssen H, Heldmann M, Prasuhn J, Tronnier V, Rasche D, Diesta CC, Domingo A, Rosales RL, Jamora RD, Klein C, Münte TF, Brüggemann N (2018) Basal ganglia and cerebellar pathology in X-linked dystonia-parkinsonism. *Brain* 141:2995–3008.
- Hurst SE, et al. (2018) A novel variant in *TAF1* affects gene expression and is associated with X-linked TAF1 intellectual disability syndrome. *Neuronal Signal* 2:NS20180141.
- Ito N, Hendriks WT, Dhakal J, Vaine CA, Liu C, Shin D, Shin K, Wakabayashi-Ito N, Dy M, Multhaupt-Buell T, Sharma N, Breakefield XO, Bragg DC (2016) Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells. *Dis Model Mech* 9:451–462.
- 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.
- Jue NK, Murphy MB, Kasowitz SD, Qureshi SM, Oberfell CJ, Elsis S, Foley RJ, O'Neill RJ, O'Neill MJ (2013) Determination of dosage compensation of the mammalian X chromosome by RNA-seq is dependent on analytical approach. *BMC Genomics* 14:150.
- Kilpinen H, et al. (2017) Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 546:370–375.
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–360.
- Law CW, Chen Y, Shi W, Smyth GK (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15:R29.
- Lee LV, Rivera C, Teleg RA, Dantes MB, Pasco PM, Jamora RDG, Arancillo J, Villareal-Jordan RF, Rosales RL, Demaisip C, Maranon E, Peralta O, Borres R, Tolentino C, Monding MJ, Sarcia S (2011) The unique phenomenology of sex-linked dystonia parkinsonism (XDP, DYT3, “Lubag”). *Int J Neurosci* 121 [Suppl 1]:3–11.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28:882–883.
- Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.
- Lipska BK, Deep-Soboslay A, Weickert CS, Hyde TM, Martin CE, Herman MM, Kleinman JE (2006) Critical factors in gene expression in postmortem human brain: focus on studies in schizophrenia. *Biol Psychiatry* 60:650–658.
- Lüth T, La $\beta$  J, Schaake S, Wohlers I, Pozojevic J, Jamora RDG, Rosales RL, Brüggemann N, Saranza G, Diesta CCE, Schlüter K, Tse R, Reyes CJ, Brand M, Busch H, Klein C, Westenberger A, Trinh J (2022) Elucidating hexanucleotide repeat number and methylation within the X-linked dystonia-parkinsonism (XDP)-related SVA retrotransposon in *TAF1* with nanopore sequencing. *Genes (Basel)* 13:126.
- Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, Matsumoto S, Tabuena MD, Maranon E, Dantes M, Lee LV, Ogasawara K, Tooyama I, Akatsu H, Nishimura M, Tamiya G (2007) Reduced neuron-specific expression of the *TAF1* gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* 80:393–406.
- McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40:4288–4297.
- Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K (2012) Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* 10:595–609.
- Migeon BR (2010) X-linked diseases: susceptible females. *Genet Med* 22:1156–1174.
- 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 U S A* 100:10347–10352.
- O'Brien T, Tjian R (1998) Functional analysis of the human *TAFII250* N-terminal kinase domain. *Mol Cell* 1:905–911.
- O'Rawe JA, et al. (2015) *TAF1* variants are associated with dysmorphic features, intellectual disability, and neurological manifestations. *Am J Hum Genet* 97:922–932.
- Oliva M, et al. (2020) The impact of sex on gene expression across human tissues. *Science* 369:eaba3066.
- Patel S, Bonora G, Sahakyan A, Kim R, Chronis C, Langerman J, Fitz-Gibbon S, Rubbi L, Skelton RJP, Ardehali R, Pellegrini M, Lowry WE, Clark AT, Plath K (2017) Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep* 18:54–67.
- Patrat C, Ouimette JF, Rougeulle C (2020) X chromosome inactivation in human development. *Development* 147:dev183095.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419.
- Rakovic A, Domingo A, Grütz K, Kulikovskaja L, Capetian P, Cowley SA, Lenz I, Brüggemann N, Rosales R, Jamora D, Rolfs A, Seibler P, Westenberger A, König I, Klein C (2018) Genome editing in induced pluripotent stem cells rescues *TAF1* levels in X-linked dystonia-parkinsonism. *Mov Disord* 33:1108–1118.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res* 43:e47.

- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463:237–240.
- Sahakyan A, Kim R, Chronis C, Sabri S, Bonora G, Theunissen TW, Kuoy E, Langerman J, Clark AT, Jaenisch R, Plath K (2017) Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* 20:87–101.
- Savage AL, Bubb VJ, Breen G, Quinn JP (2013) Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* 13:101.
- Sekiguchi T, Nohiro Y, Nakamura Y, Hisamoto N, Nishimoto T (1991) The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. *Mol Cell Biol* 11:3317–3325.
- Shen Y, Matsuno Y, Fouse SD, Rao N, Root S, Xu R, Pellegrini M, Riggs AD, Fan G (2008) X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc Natl Acad Sci U S A* 105:4709–4714.
- Silva SS, Rowntree RK, Mekhoubad S, Lee JT (2008) X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc Natl Acad Sci U S A* 105:4820–4825.
- Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3.
- Stavropoulos N, Lu N, Lee JT (2001) A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proc Natl Acad Sci U S A* 98:10232–10237.
- Sulovari A, Li R, Audano PA, Porubsky D, Vollger MR, Logsdon GA; Human Genome Structural Variation Consortium, Warren WC, Pollen AA, Chaisson MJP, Eichler EE (2019) Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc Natl Acad Sci U S A* 116:23243–23253.
- Tange O (2018) Gnu parallel 2018. Zenodo. doi: 10.5281/zenodo.1146014.
- Theunissen TW, et al. (2016) Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* 19:502–515.
- Vallot C, Huret C, Lesecque Y, Resch A, Oudrhiri N, Bennaceur-Griscelli A, Duret L, Rougeulle C (2013) XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat Genet* 45:239–241.
- Vallot C, Patrat C, Collier AJ, Huret C, Casanova M, Liyakat Ali TM, Tosolini M, Frydman N, Heard E, Rugg-Gunn PJ, Rougeulle C (2017) XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* 20:102–111.
- Walter U, Rosales R, Rocco A, Westenberger A, Domingo A, Go CL, Brüggemann N, Klein C, Lee LV, Dressler D (2017) Sonographic alteration of substantia nigra is related to parkinsonism-predominant course of X-linked dystonia-parkinsonism. *Parkinsonism Relat Disord* 37:43–49.
- Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, Luo Y (2012) RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* 14:22–29.
- Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28:2184–2185.
- Westenberger A, Rosales RL, Heinitz S, Freimann K, Lee LV, Jamora RD, Ng AR, Domingo A, Lohmann K, Walter U, Gölnitz U, Rolfs A, Nagel I, Gillesen-Kaesbach G, Siebert R, Dressler D, Klein C (2013) X-linked dystonia-parkinsonism manifesting in a female patient due to atypical turner syndrome. *Mov Disord* 28:675–678.
- Westenberger A, et al. (2019) A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. *Ann Neurol* 85:812–822.