From the Institute of Neurogenetics
at the University of Lübeck
Director: *Prof. Dr. med. Christine Klein*

Molecular Genetic Mechanisms Surrounding X-Linked Dystonia-Parkinsonism

Dissertation

in fulfilment of the requirements
for a Doctoral Degree
at the University of Lübeck

Department of Natural Sciences

Submitted by
Aloysius Domingo, M.D.
from Manila, Philippines

Lübeck, 25 October 2016

First referee: Dr. Ana Westenberger
Second referee: Prof. Dr. Stefan Anemueller

Date of oral examination: 18 April 2017

Approved for printing. Lübeck, 28 April 2017

Table of contents

Molecular genetic mechanisms surrounding X-linked dystonia-parkinsonism

1 Introduction

1.1 X-linked dystonia-parkinsonism

Dystonia is a neurologic disorder characterized by intermittent or sustained muscle contractions, leading to abnormalities in muscle tone or posture. There is a high genetic burden in its etiology as evidenced by the multitude of familial cases, and the increasing number of genes that are being found to cause dystonic phenotypes [1]. Among the monogenic forms, i.e. dystonic disorders associated with disease-causing mutations in single genes, the syndrome of X-linked dystonia-parkinsonism (XDP) is unique in its onset in adulthood, male predominance and sex-linked inheritance, transition from generalized dystonia to predominant parkinsonism, and observation of neurodegeneration in autopsied brains [2–4].

XDP typically manifests in adult males originating from Panay island in the Philippines (Figure 1). Patients initially suffer from focal or segmental dystonia that quickly becomes debilitating as symptoms generalize. Within ten to fifteen years, the dystonic symptoms wane and become less prominent; symptoms of parkinsonism then dominate the clinical picture. Affected individuals die early from aspiration pneumonia, starvation, or suicide, and the disability that affects probands prior to death has wide-reaching family and societal consequences [5].

XDP was initially described as primary torsion dystonia, due to the observation of twisting dystonic postures in the first set of cases reported in literature [6,7]. In fact, the term “lubag”, which means “to twist” in the local language, was applied in most early accounts of the disease. The concentration of the disease on Panay island was

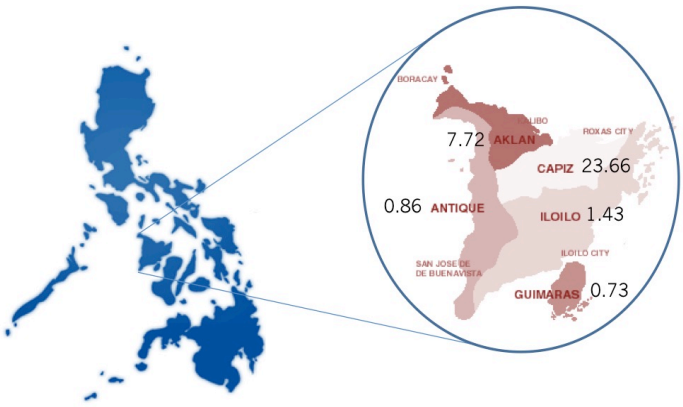

Figure 1 The Philippines and Panay Island. Numbers show the prevalence of XDP per 100,000 population in the different provinces of Panay (2011 data [2]). XDP is rare worldwide; in the Philippines, the prevalence is 0.31, and on Panay island, 5.74. Thus, XDP is endemic to the country and in particular, on this island, illustrating a genetic founder effect.

first reported by Lee et al., who investigated 28 males with the disorder, 23 of whom came directly from the island [6]. The resounding clinical phenotype observed was of generalized or axial dystonia, although focal or regional involvement of the limbs were noted to be common initial presentations. As the cohort was expanded to 42 cases, it became clear that symptoms of parkinsonism form part of the clinical phenotype [8]. The natural history of XDP, i.e. transition from focal/segmental dystonia to generalized disease, and later, predominance of parkinsonism (Figure 2), was realized as more probands were collected and as the original families were followed through time [9].

Figure 2 Natural history of XDP. The first phase (*leftmost*) is characterized by focal or segmental dystonia (here, blepharospasm and truncal dystonia are shown) followed by generalized dystonia overlapping with signs of parkinsonism (*middle*). After 10-20 years, the dystonic symptoms become less prominent and signs of parkinsonism (e.g. paucity of movement, coarse hand tremor) evolve as the predominant manifestation (*rightmost*).

To date, there are 717 cases of XDP registered, only a third of whom are alive (Lee, personal communication). The overwhelming majority of patients live on Panay island, while about 50 cases are found elsewhere in the Philippines or abroad; these individuals have either emigrated from the island, or have ancestry from there [10,11].

Due to X-linked recessive inheritance, males are most commonly affected. However, fifteen females with XDP have been described in the literature, manifesting XDP due to a variety of genetic mechanisms, e.g. homozygosity, extremely skewed X-chromosome inactivation in a heterozygous female, or mosaic X-chromosome monosomy (also in a heterozygous female) [12–15]. Interestingly, female patients express a more parkinsonian phenotype, and also have a later age at onset compared to affected males, so much so that XDP has been suggested to contribute to the epidemiology of parkinsonism in elderly females in the Philippines [15].

Neuroimaging and neuropathologic studies in XDP point to an involvement of the striatum; they also reflect the different phases of the phenotype [16]. On magnetic resonance tomography, a hyperintense putaminal rim is observed frequently regardless of duration of illness, while caudate and putaminal (i.e. striatal) atrophy is often seen in patients imaged during the later stages of their disease. Recent investigations using diffusion-weighted imaging also implicate white matter tracts in addition to striatal atrophy [17]. Nevertheless, it appears that the primary anatomical structure involved is still the striatum, given that studies in autopsied brains report differential involvement of basal ganglia structures in accordance with the phenotypic phases, with neuronal loss occuring primarily in the striosomes in the dystonic phase, and involvement of the entire striatal architecture, i.e. striosomes and matrix, in the phase of predominant parkinsonism [4,18,19]. Importantly, XDP is the only syndrome among the primary dystonias that is accompanied by overt neurodegeneration.

1.2 Genetics of XDP

It was clear even in early literature that XDP was hereditary and that the inheritance pattern is X-linked recessive; multiple affected brothers and nontransmission between successive generations of males was observed in all 21 families included in the first study describing XDP [6,7]. Further analysis of pedigrees confirmed the mode of inheritance, and estimated penetrance to be complete by the age of 50 years based on family analysis [20]. A review which included 505 XDP patients collected during a 35-year period likewise identified positive family history in 465 probands [2].

Because of the strikingly high prevalence of the disease on Panay island and the observation that individuals with XDP identified elsewhere always had lineage from this geographical isolate, a founder effect could be easily conjectured. Genetic homogeneity was used as the model to identify the XDP disease locus, i.e. all XDP patients were assumed to carry the same disease-causing mutation [21]. This postulation allowed methods of allelic association to determine the critical interval where the XDP mutation could be found [20,22]. Thus, a series of genetic linkage studies and analyses of haplotypes subsequently identified and assigned the XDP/*DYT3* locus to a 1.8-Mb region on the proximal long arm of the X chromosome (Xq13.1). Further mapping using new short tandem repeat polymorphic markers (STRPs) refined the interval to a region spanning 350 kb [23].

After the human reference sequence was revealed through the Human Genome Project, genes located within the XDP critical interval were each sequenced to search for segregating mutations. Sequence analyses in XDP patients did not reveal mutations in the *NONO, ACRC, OGT, and CXCR3* genes [23–26], however, sequencing

of overlapping PCR products covering the XDP critical interval in patients identified five single-nucleotide changes and one 48-bp deletion in introns of *TAF1*, or in intergenic regions (Figure 3) [26]. The genetic alterations were not found in 178 Filipino and 180 non-Filipino controls, hence they were termed disease-specific single-nucleotide changes (DSC1, DSC2, DSC3, DSC10, and DSC12).

Neither the DSCs nor the 48-bp deletion are found within coding sequences of known genes. Rather, they are located in spans of repetitive DNA. Among them, however, DSC3 is contained in sequences transcribed in human fetal brain and caudate [26]. The transcripts form a so-called multiple transcript system (MTS), with some variants originating from – or spliced to – exons of *TAF1*, and others being transcribed independently (Figure 3). Furthermore, the different transcripts utilize alternative exons of *TAF1*, and also form its own exons, called exons d1 to d5. DSC3 is located in exon d4 of this MTS [26,27].

With the assumption that the DSC3 C>T transition represents a missense mutation in a transcribed exon, the MTS was thus labeled as the *DYT3* gene. Importantly, no polypeptide product has been demonstrated to be coded by *DYT3*/MTS [27]. Instead, in profiling experiments performed in overexpression models, transcripts that express DSC3 have been shown to affect overall gene expression [28]. Dysregulated genes included those involved in vesicular transport in the brain and dopamine metabolism – pathways that are important in the molecular pathogenesis of dystonic and parkinsonian syndromes [29–32].

Using shotgun sequencing of cloned contigs spanning the XDP critical region, a second molecular genetics study identified a 2,627-bp insertion as yet another disease-specific genetic alteration in XDP [33]. The large insertion, categorized as an SVA (short interspersed nuclear element, variable number of tandem repeats, and Alu composite) retrotransposonal element, is found within deep intronic regions of *TAF1*, in intron 32, the largest intron of the gene (Figure 3). Subsequent genetic analyses confirmed that the change is harbored only by affected individuals and heterozygous carriers, and not by ethnicity matched controls [34].

In summary, despite extensive linkage and sequence analyses, the disease-causing mutation in XDP is still under debate [35,36]. Seven different genetic alterations (5 DSCs, one 48-bp deletion, one SVA retrotransposon insertion) segregate with the phenotype and are disease-specific, however, none are found in coding regions. Instead, all lie either in introns of *TAF1* or in the *DYT3/MTS*, a complex alternative splicing system that also involves *TAF1* exons (Figure 3).

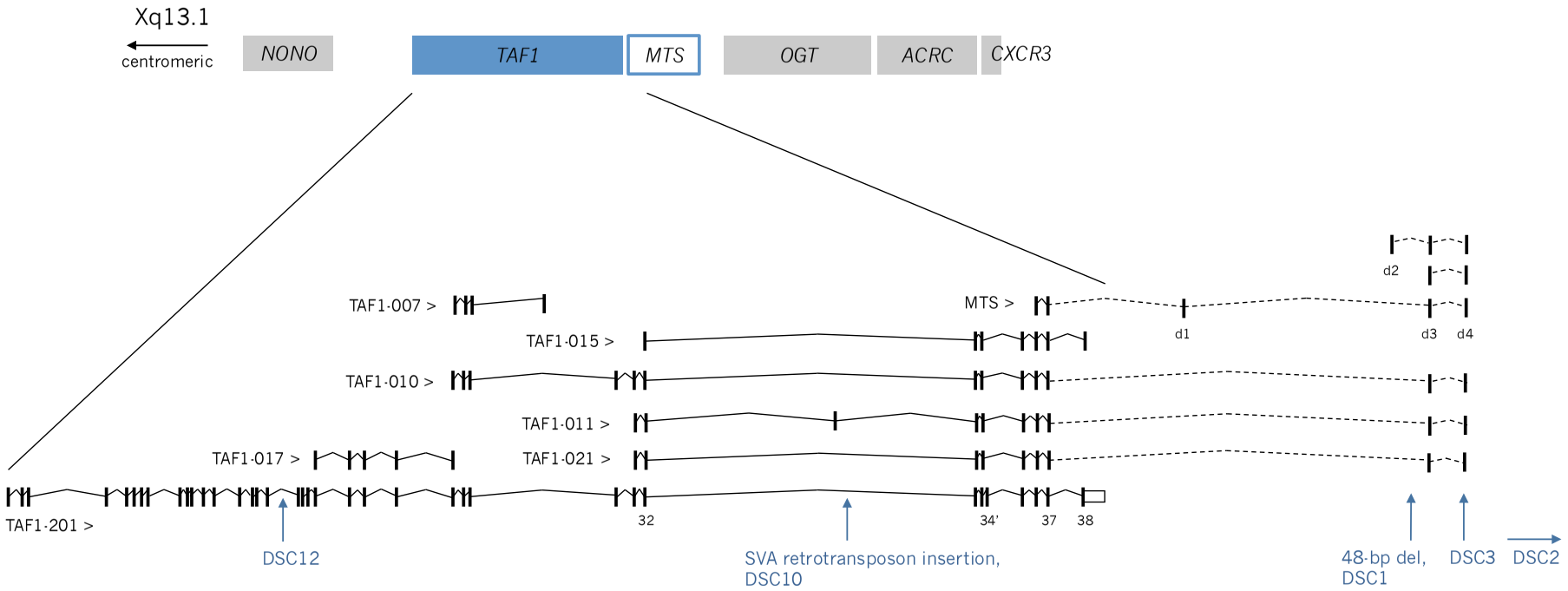

Figure 3 Transcripts of TAF1, the MTS, and XDP-specific genetic alterations. Shown are some of the known transcripts of *TAF1*, which has many different isoforms due to extensive alternative splicing events. Exon 34' for example, is not expressed in all cells. The neuron-specific isoform (*nTAF1*) includes this exon. A multiple transcript system (MTS) is distal to *TAF1*, has 4 exons (d1-d4), and may or may not be spliced to canonical *TAF1* exons (broken lines). Exons d2-d4 may also be transcribed independently. The genetic alterations associated with XDP lie either in *TAF1* introns (i.e. DSC12 in intron 18, DSC10 and the SVA retrotransposon insertion in intron 32), or in the MTS (i.e. 48-bp deletion and DSC1 in intron 2 of the MTS). DSC3 is the only alteration that is in a putative exon (d4). DSC2 is in the intergenic region between the *TAF1*/MTS and the next gene, *OGT*.

1.3 *TAF1* and its isoforms

The *TAF1* gene codes for TATA box-binding protein-associated factor 1 (also called TAFII250), the largest component of the transcription factor complex (TFIID) that initiates RNA polymerase II-based transcription of most cellular genes [37–39]. In addition to participating in this transcription machinery, the TAF1 protein also plays a regulatory role in directly activating genes central to cell growth and proliferation [40], [41]. Mutations in the DNA-binding domain of *TAF1* have thus been shown to arrest the cell cycle *in vivo* [42]; in humans, mutations in *TAF1* cause a syndrome of intellectual disability, dysmorphic facial features, and in some patients, dystonia and parkinsonism [43].

The *TAF1* gene is alternatively spliced, creating a multitude of isoforms (Figure 3), particularly in the brain [27,33]. Among the different splice variants that have been detected, one species, TA14-391, utilizes an alternative exon of 6 bp (called exon 34'). This transcript is preferentially expressed in the brain, and is second to the major form of *TAF1* in terms of abundance in neuronal tissue [33,44]. In contrast to *TAF1*, the expression of which is highest in embryonic stages in the mouse brain, almost no expression of the neuronal transcript is seen 10.5 days post-conception; expression then becomes stable from 17.5 days to 40 weeks, arguing for a role in mature neurons [44].

Interestingly, significantly reduced expression of the TA14-391 isoform (which is also referred in literature as neuron-specific *TAF1* or *nTAF1*) was demonstrated in the caudate of one patient with XDP [33]. Reduced expression in comparison to controls was also identified in the cortex and nucleus accumbens, i.e. areas where neuronal loss was not seen, implying that the loss of *nTAF1* was the cause rather than the result of the neurodegeneration seen in the caudate [33]. Significantly reduced expression of *nTAF1* (among other *TAF1* species) was also recently documented in neural progenitor cells derived from induced pluripotent stem (iPS) cells obtained from XDP patients [45]. Further, immunohistochemistry studies in rat brain using an antibody specific to nTAF1 protein demonstrate a localization pattern that is similar to the striatal neuropathology seen in XDP, i.e. nTAF1-positive nuclei are enriched in the striatum vs. other brain regions, and distribute to the striosomes preferentially [46].

Thus, genetic and expression studies converge on *TAF1* as the putative disease-causing gene in XDP, classifying the disease as a transcriptional dysregulation disorder along with Huntington's disease, a movement disorder that is also characterized by prominent striatal pathology [47]. The disease-specific genetic alterations found in XDP patients lie in or in the proximity of *TAF1*, while expression

studies show loss of a neuron-specific isoform of the gene in the population of neurons that degenerate in XDP. Here, the SVA retrotransposon insertion is said to play a critical role by altering the expression of *TAF1* and its isoforms, as has been shown for other retrotransposons that alter the expression of nearby genes and cause disease [48,49].

1.4 New genetic models in the study of neurologic disorders

Studying the molecular mechanisms of neurologic disorders – such as of dystonia and XDP – could be especially challenging due to the inaccessibility of the central nervous system. Although peripheral tissue, e.g. blood and fibroblasts, have proven to be useful surrogates in some molecular investigations [50–52], genetic models that recapitulate the phenotype in neurons likely have even greater value and relevance.

iPS cells reprogrammed from human dermal fibroblasts retain the full genetic background of the individual and can themselves be used to discover disease mechanisms [53,54]. Furthermore, they can be differentiated into nervous tissue, and importantly, into different neuronal subtypes (e.g. dopaminergic, cortical, medium spiny striatal), allowing the study of neuron-specific processes in an endogenous in-vitro model [55,56]. iPS cell-derived neurons have been used to model neurodegenerative disorders, including genetic forms of Parkinson's disease and of Huntington's disease, where they have proven essential in clarifying neuron-specific phenotypes associated with mutations, while also recapitulating the situation *in vivo* [57–59].

A second important recently developed technology is one that can be used to modify genetic models, i.e. the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated proteins (Cas) system [60,61]. The CRISPR/Cas complex makes use of bacterial endonucleases to generate double stranded breaks (DSB) in DNA when it encounters a proto-spacer-associated motif (PAM) [62]. Thus, the system has been used for targeted genome editing in cells, and in particular, in iPS cells [63], which, when further differentiated into target tissue, generates isogenic controls, i.e. cell lines that harbor the entire genomic architecture of the parent, except for the mutation targeted by the system for editing [64].

Using guide RNAs to direct the Cas9 protein to the target site, repair of CRISPR/Cas-mediated DSBs ensue through non-homologous end-joining, which induces small insertions or deletions (indels) that abolish the expression of the target gene (and is thus useful to generate isogenic knock-outs). Meanwhile, precise point mutations and indels can be introduced with the use of a donor template, permitting DSB repair via homologous recombination. Lastly, dual guide RNA systems, i.e. the use of two guide

RNAs – one at each end of the target genomic region, have recently been used to produce larger deletions or other genomic rearrangements [61,65,66]. Given its flexibility, CRISPR/Cas9-mediated genome editing has been applied successfully to various hematologic, oncologic, genetic, and infectious diseases, but is also recently being applied to model diseases of the nervous system [67–69].

Being an orphan disease, little is known about the molecular mechanisms causing XDP. Despite extensive linkage and genetic studies, a convincing, protein-altering or functionally significant mutation has yet to be identified. Furthermore, a tangible link between several genetic alterations found within the region of the *TAF1* gene to the reduced expression of the gene or its isoforms has not yet been demonstrated. As such, complementary genetic, genomic, and transcriptomic methods, coupled with new genetic modeling techniques through the use of iPS cells and CRISPR/Cas9-mediated genome editing – technologies that have not previously been used to study XDP – present a viable multimodal approach to disentangle the genetic and molecular mechanisms surrounding this enigmatic disease.

1.5 Hypotheses and objectives

Hypothesis 1: There is only one truly disease-specific (and thus disease-causing) mutation among the genetic alterations described in XDP.

Strategy: The disease-specificity of the other variants can be ruled out by performing a combination of genetic analyses in large groups of XDP patients and appropriate ethnically matched healthy controls.

Objective 1.1: To perform haplotype analysis in XDP patients using STRP markers within the XDP critical region, and in doing so, detect possible recombination events that would narrow the disease locus.

Objective 1.2: To perform targeted sequencing and PCR-based detection of all previously described genetic alterations in XDP patients, in order to identify affected individuals that carry only the truly disease-specific XDP change, and not the rest.

Objective 1.3: To perform targeted sequencing and PCR-based detection of all previously described genetic alterations in healthy controls – specifically from the population where the XDP genetic founder likely originated from – in order to identify individuals carrying single occurrences of any of the XDP-associated changes, thus ruling these out as disease-specific.

Hypothesis 2: The disease-causing mutation in XDP may be a previously undetected protein-coding sequence change or copy number variation (CNV).

Strategy: Use of unbiased next-generation sequencing and microarray-based genotyping can reveal new XDP-specific genetic alterations not identified by older methods.

Objective 2.1: To use whole genome sequences obtained from XDP patients in order to identify a previously undetected sequence change (single nucleotide variant, SNV) within or outside the XDP-linked region.

Objective 2.2: To use genotyping data obtained from XDP patients in order to detect CNVs (within and outside the X-chromosome) that are specific to XDP patients.

Hypothesis 3: The expression of *TAF1* is dysregulated in peripheral models of XDP.

Strategy: Consistent with the function of the gene as a transcriptional activator, global gene expression changes identified via expression profiling in peripheral tissue would support *TAF1* dysregulation in XDP.

Objective 3.1: To perform quantitative expression analyses of coding genes and the MTS in easily accessible tissue from XDP patients, i.e. blood and fibroblasts, and to identify whether *TAF1* expression is reduced in these peripheral models.

Objective 3.2: To use microarray-based expression profiling data and to compare global gene expression in XDP patients vs. controls. To perform validation and replication experiments using quantitative PCRs.

Objective 3.3: To perform bioinformatic analyses of expression profiling data in order to list differentially expressed genes in XDP. To use bioinformatic tools to identify enrichments in this list that are relevant to the molecular basis of XDP, e.g. enriched networks, transcription factors, gene sets, and gene ontology enrichments.

Hypothesis 4: The pattern of *TAF1* and *nTAF1* expression in endogenous models of XDP, i.e. iPS cells and iPS cell-derived neurons, are similar to their patterns of expression in the brain, validating these as useful models for molecular genetic studies in XDP.

Strategy: Targeted expression analyses of *TAF1* isoforms, in particular, measurement of *nTAF1* transcripts using different methods, will determine whether cellular models recapitulate the situation *in vivo*.

Objective 4.1: To quantify the abundance of *TAF1* and *nTAF1* transcripts in iPS cells and iPS cell-derived cortical and striatal neurons, and to compare to amounts described in the brain.

Objective 4.2: To design and use a highly specific quantitative approach that will determine the *nTAF1* and non-*nTAF1* expression in iPS cell-derived cortical and striatal neurons.

Hypothesis 5: The presence of the SVA retrotransposon insertion is intimately linked to the dysregulation of *TAF1* transcripts in XDP.

Strategy: Removal of the SVA retrotransposon insertion via genome editing will thus result in the 'rescue' of *TAF1* underexpression in cellular models of XDP.

Objective 5.1: To design and perform CRISPR/Cas9-mediated genome editing in iPS cells in order to excise the SVA retrotransposon insertion in lines obtained from XDP patients (or the corresponding genomic region in control lines).

Objective 5.2: To determine the changes in *TAF1* expression that result from excision of the SVA insertion or the corresponding genomic region.

2 Methods

2.1 Molecular genetics and DNA-based analyses

The molecular genetic methods encompass experiments in **Objectives 1.1** to **2.2**.

2.1.1 Patients and samples

Through collaboration with the XDP Study Group of the Philippines, blood samples from XDP patients were obtained from Panay Island and shipped to the Institute of Neurogenetics. Included in the genetic analyses were 163 adult males (average age: 49 ± 9.6 years, range: 31-73 years) from the Philippines clinically diagnosed with XDP, representing already a fifth of all cases ever documented in the registry, and the largest collection of XDP samples for genetic analysis to date. The control cohort consisted of 452 neurologically normal individuals from Panay island, six sons of XDP males, and 15 healthy Filipinos (who do not come from Panay island). Thus, in total, 473 controls representing 695 X-chromosomes (251 male, 222 female, average age: 37 ± 12.5 years in males and 38 ± 13.5 years in females) were included. All subjects gave written informed consent and institutional review boards from Germany and from the Philippines approved the studies. Genomic DNA was extracted from all blood samples using standard methods.

2.1.2 Haplotype and STRP analysis

STRP markers within and surrounding the previously described XDP locus on Xq13.1 (DXS10015, DXS10016, DXS10017, DXS10018, and DXS559) were analyzed using M13-tailed, infrared dye-based labeling in multiplex polymerase chain reactions (PCRs) [70,71]. In this method, the forward primer is synthesized with an M13 tail at its 5' end (5'-CACGACGTTGTAAAACGA-3'), and a reverse primer labeled with infrared dye is added to the reaction. For each reaction, 10 ng of genomic DNA was mixed with 0.5 µl 10x buffer, 1 µl 1 mM dNTPs, 0.6 pmol each of STR-specific forward and reverse primers, 0.25 U Taq polymerase, and 0.6 pmol M13-infrared dye reverse primer (5'-GGATAACAATTTCACACAGG-3'); water was added to a final reaction volume of 5 µl. A multiplex PCR reaction was then performed with 35 cycles of denaturation (95°C), annealing (55°C or 60°C), and extension (72°C) on a thermocycler. Products were visualized on polyacrylamide gels after mixing with 5 µl of formamide-based loading buffer, using a Li-Cor® 4200 Genetic Analyzer 4200. Gene ImagIR™ software was used to analyze the obtained images. Primer sequences and conditions are described in the Appendix.

To analyze allele lengths and compare between XDP patients, exemplar products were first sequenced using an ABI3500XL Genetic Analyzer® (Applied Biosystems), i.e. via Sanger sequencing. After which, STRPs were assigned values by lining up with exemplar products of known lengths.

2.1.3 Sanger sequencing and PCR-based insertion detection

Sanger sequencing of each of the DSCs and the 48-bp deletion was performed. First, a PCR was performed by mixing 40 ng genomic DNA with 1 µl 10x $MgCl_2$-containing buffer, 2 µl 1 mM dNTPs, 4 pmol each of sequence-specific forward and reverse primers, and 0.35 U Taq polymerase; water was added to a final reaction volume of 10 µl. A PCR reaction was then performed with 35 cycles of denaturation, annealing, and extension (see Appendix for primer sequences and PCR conditions). PCR products were checked and visualized via agarose gel electrophoresis after addition of loading buffer, and then subjected to ExoI-FastAP (alkaline phosphatase)-based cleanup to remove unincorporated primers and nucleotides. For each reaction, 10 U Exonuclease I and 1 U FastAP (ThermoFisher Scientific) were used, followed by incubation at 37°C for 15m and deactivation at 85°C for 15m. After which, the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used to incorporate flourescent-labeled ddNTPs into the cleaned PCR fragments. For this, 2 µl of each ExoI-FastAP product was added to 1.5 µl of 5x buffer, 5 pmol of sequencing primer, and 1 µl of Terminator mix; water was added to a final reaction volume of 10 µl. The resulting mix was cycled 25x (conditions: 95°C for 10s, 60°C for 1m5s) using a thermocycler. Products were then sequenced on a capillary electrophoresis-based sequencer, i.e. ABI3500XL Genetic Analyzer® (Applied Biosystems). Electropherograms were visualized using Sequencher® (Gene Codes Corporation, Ann Arbor, MI) or Mutation Surveyor® (Softgenetics, State College, PA) software.

To detect the SVA insertion, a step-down protocol was adapted from Kawarai, et al. [34]. The PCR reaction was mixed using 10-20 ng of genomic DNA, 15 pmol each of forward and reverse primers, 25 µl 10x buffer, 5 µl 2 mM dNTPs, and 0.7 U high-fidelity KOD polymerase (Merck Millipore); water was added to a final reaction volume of 50 µl. Step-down cycling was performed (see Appendix for primer sequences and step-down cycling conditions). Products were visualized on a 1% agarose gel, where the length of the SVA insertion-containing allele after gel electrophoresis was 3,226 bp and the wildtype allele, 599 bp.

2.1.4 Genome sequencing and analysis

From the patient cohort, five males were selected for genome sequencing (three 'XDP haplotype' carriers: L-5709, L-5712, L-7153, and two non-carriers: L-5708, L-7677, see below). In addition, one female (mother of one patient, L-7159, homozygous for all the XDP genetic alterations) was included. Aside from the mother-son pair, none of the patients selected for genome sequencing were obviously related.

Genome sequencing was outsourced to two companies, i.e. one genome (L-5712) was sequenced by Complete Genomics (Mountain View, CA), and others were sequenced by Knome, Inc. (Boston, MA) using an Illumina® platform. Resulting bam files were obtained and submitted to a collaborating bioinformatician for annotation (Ingrid Braenne, Institute for Integrative and Experimental Genomics). Annotation was performed using Annovar [72]. From the list of annotated variants, the following filters were applied: i) only rare variants (with a minor allele frequency <0.01 based on estimations from the 1000 Genomes Project, dbSNP, or the Exome Variant Server), ii) not within regions of known segmental duplications, and iii) not found in three in-house genomes from patients without XDP. Variants of interest within the locus defined by STRP markers above, as well as chromosome-wide (X-chromosome), and genome-wide were then compared between all six genomes sequenced.

Novel SNVs revealed by genome sequencing, i.e., not previously detected in other studies, were verified via Sanger sequencing in the rest of the XDP samples, and in a select number of controls, using procedures described above. Primer sequences and PCR conditions are described in the Appendix.

2.1.5 Microarray-based SNP genotyping and copy number analysis

Analyses of copy number were performed using microarray-based SNP genotyping data. 120 patient samples were submitted to an external service provider for genotyping using an Illumina® OmniExpress-24 Beadchip array (Illumina, San Diego, CA). CNV calls on the X-chromosome and genome-wide were generated by a collaborating bioinformatician (Lars Bertram, Lübeck Platform for Genome Analytics), using the PennCNV software [73]. Of note, to reduce spurious CNV calls, only CNVs longer than 10 kb and spanning at least 5 SNPs were included; CNV segments with >50% overlap with centromeres, telomeres, and pseudo-autosomal regions were also not included.

2.2 Expression studies and transcriptomic analyses

The methods described for expression analysis cover **Objectives 3.1** to **3.3**.

2.2.1 Patients and samples

As part of multimodal clinical studies, 16 XDP patients were invited to Lübeck, from whom blood and skin biopsies were taken for expression studies. All were in the combined dystonic-parkinsonian stage of disease at the time of sample collection. Age-, sex-, and ethnicity-matched healthy controls were likewise recruited from accompanying relatives, or from Filipinos living in Germany. Blood samples from nine patients and 13 controls were collected in PAXGene® tubes (Qiagen). From these, RNA was extracted using manufacturer's protocols. Skin biopsies were obtained from 14 patients and from eight controls. Genomic DNA from blood and/or fibroblast cultures were also extracted, and the XDP-specific genetic alterations were verified in each sample using PCR detection and Sanger sequencing, as described above. All subjects gave written informed consent prior to sample collection.

Primary fibroblast cultures were grown from skin biopsies on Dulbecco's modified Eagle medium (Gibco) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. Cultures were passaged when confluent, and at passages seven to nine, were harvested via Accutase® (Life Technologies) treatment. Total RNA was isolated from a defined amount of cells using the RNeasy® Mini Kit (Qiagen), and after DNaseI treatment (Qiagen).

All RNAs were subjected to column-based RNA clean-up (RNeasy® Mini Kit), and all were determined to have RIN values >8.0 on the Bioanalyzer 2100 (Agilent Technologies) and Nanodrop 2000 (Thermo Scientific).

2.2.2 Expression analysis using quantitative PCR

Quantitative PCR (qPCR) was used to test the expression profile of protein-coding genes in the XDP locus and of *DYT3*/MTS, and to validate the results of the microarray-based expression profiling (below) in discovery and replication sets.

RNA was first reverse transcribed using the Maxima® First Strand cDNA Synthesis Kit (Thermo Scientific), by mixing 500 ng RNA, 4 µl 5x buffer, and 2 µl Maxima® reverse transcriptase solution; water was added to a final reaction volume of 20 µl. Incubation was performed on a thermocycler for 10m at 25°C, followed by inactivation at 50°C for 15m. Real-time PCR was then performed on a Roche LC480 LightCycler® System using SYBR Green I chemistry. SYBR Green I dye is a fluorescent DNA-binding dye, which when detected and measured in real-time, is able to determine the amount of template (DNA or cDNA) present in the solution. To set the qPCR reaction, a diluted

amount of cDNA (1:6 dilution from the reverse transcription reaction) was mixed with 3 pmol each of forward and reverse primers, and 6 µL 2x SYBR™ Green Master Mix (Thermo Fisher Scientific), which contains buffer, dNTPs, thermostable hot-start DNA polymerase and SYBR Green I dye. Primer sequences were obtained from a publicly available database (http://primerdepot.nci.nih.gov). This database returns primers that span the largest exon of target genes, thus avoiding amplifying genomic DNA. Furthermore, cycling was limited to 40 cycles, with only 10 seconds each of denaturation (95°C), annealing (57 to 60°C), and extension (72°C), in order to again rule-out DNA contamination. No-template and no-reverse-transcriptase blanks were included in each experiment; all measurements were done in duplicate for every qPCR experiment, and experiments were done in accordance with MIQE guidelines (Minimum Information for publication of Quantitative real-time PCR Experiments) [74]. Full primer sequences and corresponding PCR conditions can be found in the Appendix.

Expression levels were computed using the Advanced Relative Quantification method (Roche LightCycler® software, release 1.5.0). This automated method approximates the well-accepted delta-delta-CT method [75], but corrects for differences in efficiency between target and reference genes. Expression levels of target genes were either normalized against the expression of Expressed Alu Repeats (EARs) [76,77], or using two reference genes. First, the Genevestigator platform (https://genevestigator.com) was used to determine the usual expression profiles of the target genes, *TAF1* being the main target. The RefGenes [78] tool was then used to determine the most stable reference genes in fibroblasts (*FLAD1* and *ZNF672*). Furthermore, a literature search was done to determine appropriate, stable reference genes in blood (*UBC* and *FPGS*) [79].

Absolute fold-change (FC) of expression was determined by obtaining the ratio of the average expression value in the XDP group vs. the average in controls. The inverse was obtained when this value was below 1 (denoting underexpression in patients). The p-value was obtained by performing a two-tailed Student's t-test between expression values in the XDP group vs. controls. Significance level was set at p-value <0.05.

2.2.3 Microarray-based expression profiling and validation

Microarray-based expression profiling was performed by OakLabs GmbH (Henningsdorf, Germany). 10/15 blood-/fibroblast-derived RNA samples from patients, and 12/10 from controls were subjected to microarray expression profiling, i.e. one to two technical replicates were included in each group. An Agilent SurePrint™ G3 Human Gene Expression 8x60k v2 oligonucleotide microarray (Agilent

Technologies) was used. All samples belonging to a tissue group (blood, fibroblasts) were hybridized onto the array in a single batch, thereby eliminating the possibility of batch effects, which are common confounders in gene expression experiments [80].

Quality control was performed by a collaborating bioinformatician (David Amar, Edmond J. Safra Center for Bioinformatics, Tel Aviv University), via quantile normalization and exclusion of probes with intensity less than the median value in all samples. Expression values in probes that mapped to the same Entrez genes were also averaged. Differential expression analyses (XDP vs. controls, blood and fibroblast-derived data treated independently) was then performed in R, using i) Wilcoxon rank-sum tests, and ii) the limma statistic with SAM (Significance Analysis of Microarrays) correction, both at 0.1 FDR-corrected cut-offs, i.e. q-value <0.10 [81,82]. The list of differentially expressed genes (DEGs) was then ranked according to SAM q-value and Wilcoxon score.

From this list of DEGs, genes were selected for qPCR validation by hand, based on putative function, relevance to XDP pathophysiology, and expression in the brain (Table 1). qPCR experiments were performed using SYBR Green I chemistry as described above, on i) a "discovery set" of RNAs, i.e. the exact same RNAs used in the expression profiling, ii) a "technical replication set", i.e. RNAs derived from fibroblasts regrown in culture but from the same patients included in the microarray/validation experiments, and iii) an "independent replication set", i.e. a completely new set of RNAs from a separate set of XDP patients and controls (Figure 4). Sample-level analyses were performed for *TAF1* by determining the relative expression in each individual sample (including discovery and all replication samples).

2.2.4 Network and enrichment analyses

From the entire list of DEGs in fibroblasts, enrichment analysis was performed in EXPANDER (Expression Analysis and Displayer), a bioinformatic tool for gene expression analysis, and especially, enrichment analysis of gene lists (http://acgt.cs.tau.ac.il/expander/) [83,84], developed by collaborators at the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. In particular, enrichments of Gene Ontology (GO) terms and networks were searched using the TANGO procedure [85], pathway enrichment was probed using the KEGG and Wikipathways features, miRNA enrichment using FAME [86], and transcription factor motif enrichment and putative transcriptional regulators using PRIMA [87]. Additional GO/network analysis was also performed (outside of the EXPANDER software environment) using GSEA (Gene Set Enrichment Analysis) [88] and GeneMania (http://www.genemania.org) [89].

Table 1 Genes validated in qPCR experiments.

Gene	Function
TAF1	Codes for the largest subunit of TFIID; putative disease gene in XDP
OGT	Adds N-acetylglucosamine via O-glycosidic bonds to serine or threonine
ACRC	Predicted acidic nuclear protein with putative N-glycosylation sites
CXCR3	Expressed by T helper-1 (Th1) cells and recruited to inflamed tissue
SYTL2	Involved in RAB27A-mediated vesicle transport and secretion
SLC5A3	Codes for a Na^+/myoinositol cotransporter present in the brain
EFNB1	Ligand of an ephrin-related kinase; located in XDP linkage region, Xq13.1
MRPS6	Mitochondrial ribosomal protein
ZADH2	Codes for a zinc alcohol dehydrogenase
ATP6V0E2	Subunit of a lysosomal H^+-ATPase pump
BHLHE40	Basic-helix-loop-helix protein that binds DNA to transduce external signals
ZC3H12A	Codes for a zinc finger protein with RNAse activity
GZF1	Regulator that represses transcription when bound to response element
GRIN2D	Codes for an NMDA receptor, which may induce brain neurotrophic factor
KCND2	Voltage-gated K^+ channel acting in repolarization phase of action potential
DBN1	Actin-binding protein important for dendrite growth in neurons
ETV3	Oncogene, represses NF-kappaB signalling
DUSP1	Codes for a dual-specificity phosphatase

TAF1, OGT, ACRC, and *CXCR3* are the protein-coding genes within the XDP-linked region on Xq13.1. Full primer sequences can be found in the Appendix.

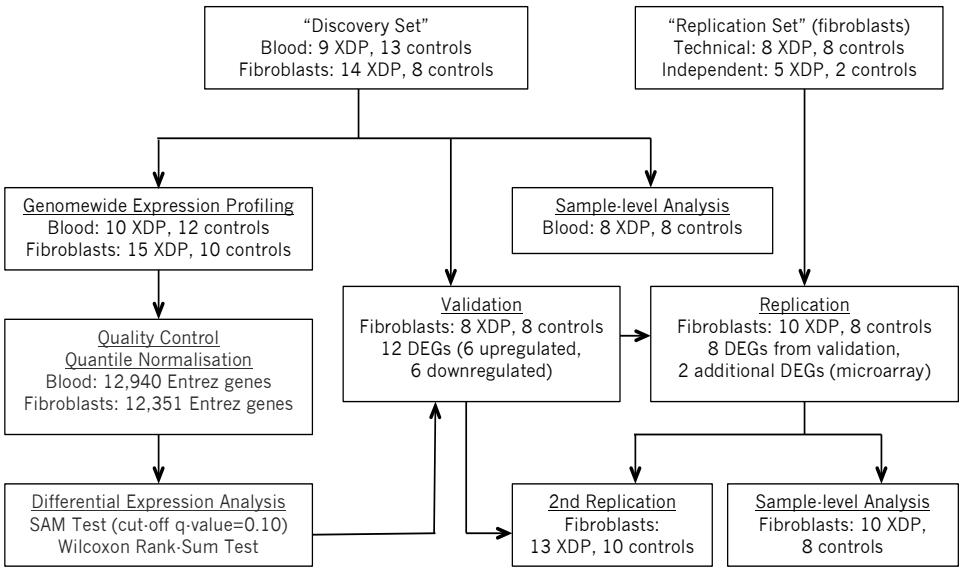

Figure 4 Flow of samples used for expression analyses. Numbers indicate samples from XDP patients and controls used for microarray-based profiling and the qPCR experiments (validation, replication, sample-based analyses). In the profiling, validation, and in the replication experiments, groupwise comparison was performed (average expression level in XDP group vs. average in controls). In the sample-level analyses, each sample from an XDP patient was compared against the average expression level in controls.

2.3 Isoform analyses in cellular models

The experiments in this section cover **Objectives 4.1** and **4.2**.

2.3.1 Samples

iPS cells and iPS cell-derived cortical and striatal neurons from XDP patients and controls were obtained from the iPS cell facility at the Institute of Neurogenetics. These iPS cells have previously been reprogrammed from human adult skin fibroblasts using the Sendai method, and differentiation into dopaminergic, cortical and striatal neurons was performed by others using published protocols (Karen Grütz, Institute of Neurogenetics) [55,56,90]. Prior to use in the following experiments, these neuronal precursors and mature neurons were characterized for expression of key neuronal markers using qPCR and immunofluorescence staining.

2.3.2 Molecular cloning analysis

cDNA fragments containing *TAF1* exon 34' were obtained by PCR amplification of reverse-transcribed RNA (as above) obtained from three control dopaminergic neuronal lines, and in a separate mix, three control striatal neuronal lines, and using primers placed in exon 33 and in the 5' UTR of *TAF1* (see Appendix for primer sequences and PCR conditions). Purified products were cloned into a pCR™4-TOPO®-TA cloning vector (Thermo Fisher) by incubation at room temperature for 5m with 1 μl vector, 2 μl PCR product, and 1 μl salt solution; water was added to achieve a final volume of 6 μl. This TA cloning® method is based on the ligation of PCR products with 3' –A overhangs generated by Taq polymerase amplification, and uses DNA Topoisomerase I. The ligated construct was first diluted with 18 μl of water, then 1.5 μl of this solution was mixed with 50 μl of electrocompetent *E. coli* (Invitrogen) and electroporated. After which, incubation with 5 ml of LB (Luria Broth) bacterial broth for 1h at 37°C was done to promote bacterial amplification and gene transcription. 25 μl of the resulting bacterial culture was then plated on antibiotic (ampicillin)-containing LB agar plates and incubated at 37°C overnight. Colonies were picked the next day (60 colonies from cDNA obtained from dopaminergic neurons, and 70 from striatal neuron-cDNA), dissolved in 100 μl of water in 96-well plates, and DNA released via heat disruption (5m at 95°C). Each well was then analyzed for the presence of exon 34' using Sanger sequencing analysis (protocol as described above), using 5 μl of DNA-containing solution as template. Universal M13 forward and reverse primers were used for sequencing.

2.3.3 Optimization of transcript-specific qPCR

To determine the abundance of exon 34'-containing transcripts in different cell models, a transcript-specific PCR reaction had to be optimized. Transcript-specific forward primers were designed by using the last 16 bases of exon 34 (preceding exon 34'), a mismatched base at the third-to-the-last position of the primer, and a transcript-specific base at the last position, i.e. the first base of exon 34' (for the *nTAF1*-specific primer) or the first base of exon 36 (representing non-*nTAF1*, or canonical *TAF1*, hereafter referred to as *cTAF1*) (Figure 5). Similar thermodynamic efficiencies were thus ensured by the fact that the transcript-specific bases were only either a 'C' or a 'G', and by using the same reverse primer for all qPCR reactions (placed at the distal end of exon 36).

Transcript-specificity was tested and optimized by i) performing a qPCR reaction using the *nTAF1*-specific primer on cDNA from tissues where *nTAF1* expression is supposed to be null, e.g. fibroblast and blood, ii) performing extended qPCR reactions using the *nTAF1*-specific and *cTAF1*-specific primers on plasmid DNA bearing *cTAF1* and *nTAF1* fragments, respectively, and iii) evaluating the linearity of a standard curve with different proportions of *nTAF1*- and *cTAF1*-containing plasmid DNA, i.e. 100:0, 80:20, 60:40, 40:60, etc.

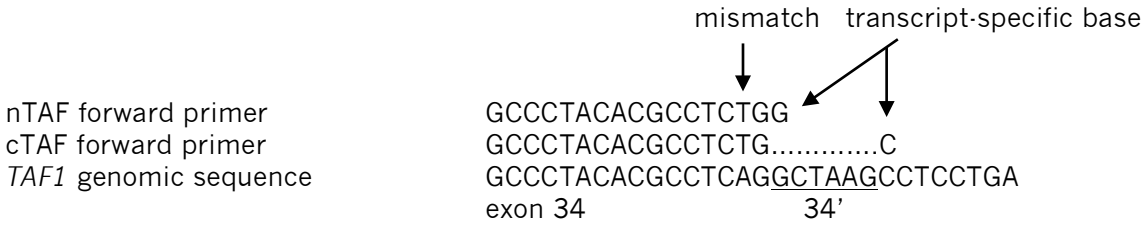

Figure 5 Design of transcript-specific primers. The last bases of *TAF1* exon 34 were used, followed by a transcript-specific base, i.e. 'G' for *nTAF1*, representing the first base of exon 34' and 'C' for *cTAF1*, representing the first base of the next exon (exon 36). A mismatched base was introduced at the third-to-the-last position to ensure that primer annealing depended exclusively on the last/transcript-specific base.

Parts ii) and iii) were performed using plasmid DNA expressing single copies of either *nTAF1* or *cTAF1* fragments. These were obtained by cloning full-length *TAF1* cDNA, obtained by amplifying neuronal cDNA using primers placed at the start and stop codons of *TAF1*, into an expression vector. Due to the large size of the insert, cloning was achieved only by using 'ultracompetent' bacteria. First, full-length *TAF1* cDNA was cut with the restriction enzyme NheI. 5 µl of the product was then ligated with 3 µl of pNK-NheI vector, and 1 µl each of T4 ligase, buffer, and water (pNK-NheI vector provided by Aleksandar Rakovic, Institute of Neurogenetics). Ligation was performed

for 1h at room temperature, followed by inactivation at 65°C for 5m. Transformation protocol followed the instructions in the Agilent XL10-Gold® Kit (Agilent). Briefly, 3 μl of the ligation product was mixed with 1 aliquot of ultracompetent cells, incubated for 30m on ice, then heat pulsed in a water bath 42°C for 55s, followed by another 2m on ice. 900 μl of pre-heated LB bacterial broth was then incorporated into the transformation product, and incubated at 37°C for 1h (at 200 rpm rotation) to promote amplification. 100 μl of bacterial culture was then plated into antibiotic (ampicillin)-containing LB agar plates, and incubated overnight at 37°C. Recombinant colonies was tested by first extracting DNA using the QIAprep® Plasmid MiniPrep Kit (Qiagen), then subjecting to restriction analysis (using NheI) and gel electrophoresis, and subsequent Sanger sequencing of insert-containing plasmids. Plasmids expressing *nTAF1* and *cTAF1* sequences were then selected and used to optimize the *cTAF1*-specific and *nTAF1*-specific primers, respectively. Serial dilution curves were constructed to approximate *TAF1* expression in iPS cells and in neuronal tissue. Sequences of primers used to verify the full sequence of *TAF1* can be found in the Appendix.

2.3.4 Transcript-specific PCRs

Three iPS cell-derived neuronal (cortical) lines and 3 striatal neuronal lines were subjected to qPCR analysis (SYBR Green I chemistry, as in Expression Analysis, Section 3.2.2 above). Optimized transcript-specific primers (above) were used to detect the targets; *FLAD1* was used as a reference gene. This reference gene was chosen as it is a stable housekeeping gene that is expressed at approximately the same bandwidth as the target (i.e. *TAF1*) in neuronal cells based on microarray data [78]. The proportion of *nTAF1* transcripts was defined as the relative expression of *nTAF1*, divided by total *TAF1* expression, i.e., *nTAF1* + *cTAF1*.

To compare the relative expression of *cTAF1* and *nTAF1* in XDP patients vs. controls in striatal neurons, six samples from patients and six from controls were used. The six samples in each group came from three different iPS cell lines, from three different patients/controls (representing biological replicates). From these iPS cell lines, two neuronal differentiations each were performed by others at the Institute of Neurogenetics iPS cell facility (representing differentiation replicates). Furthermore, in the qPCR experiments, the expression of the target and reference genes were measured twice for each sample (representing technical replicates). Two-tailed Student's t-tests were performed, with significance set at p-value <0.05.

2.4 Genome editing using the CRISPR/Cas9 system

The methods in this part cover **Objectives 5.1** and **5.2**.

2.4.1 CRISPR/Cas9 system design

A dual-guide RNA strategy was used to excise the SVA insertion in XDP iPS cell lines [66]. For this, two 17-bp gRNAs were designed, each targeting sequences in the flanking regions of the SVA insertion (GenBank: AB191243.1). gRNA1 targeted 5'-TTTAGTTTTACAAGACA-3'; the PAM sequence immediately after (CGG) was 10 bp proximal to the first base of the SVA insertion. gRNA2 targeted a genomic segment distal to the insertion, 5'-GTACTGGCATTGTAATA-3', with the PAM sequence (AGG) 41-bp distal to the last base of the SVA insertion. Each gRNA was cloned into the pLKO1-Puro vector (Addgene plasmid #8453) using the AgeI and EcoR1 restriction sites via electroporation (cloning procedure as above), and validated using restriction analysis and Sanger sequencing, using the universal CMV forward and M13 reverese sequencing primers (https://www.addgene.org/mol-bio-reference/sequencing-primers/). Plasmids expressing Cas9 protein and the resistance gene to the antibiotic blastocidin (pCas9-Blast) were obtained from the genome editing facility at the Institute of Neurogenetics (c/o Aleksandar Rakovic). Off-target sites were predicted using the online tool Cas-OFFfinder (http://www.rgenome.net/cas-offinder/) [91].

2.4.2 Transfection and cell culture

As with the experiments performed in Section 2.3 (isoform analysis in cellular models), iPS cells reprogrammed from dermal fibroblasts of XDP patients and controls were from the iPS cell facility at the Institute of Neurogenetics. Three iPS cell lines obtained from XDP patients (L-5748, L-7149, L-7994) and one control iPS cell line (L-2131) were thawed from stock and maintained on Matrigel®-coated (Corning) 6-well plates with mTESR™1 feeder-free iPSC medium (Stem Cell) prior to transfection. iPS cells from each line were harvested using Accutase® treatment (Life Technologies), passed through a 40 μm nylon cell strainer (Falcon) to achieve single-cell status, and 1 x 10^6 cells were then transfected with 5 μg pCas9-Blast and with 5 μg of each vector expressing the two gRNAs (i.e. pLKO1-gRNA1-Puro and pLKO1-gRNA2-Puro), using 100 μl Nucleofector™ Kit 1 solution (Lonza) and an Amaxa® Nucleofection II device (program A-023). Transfected cells were then plated on multiple Matrigel®-coated wells with mTESR™1 medium supplemented with ROCK inhibitor (Santa Cruz Biotech).

After 24h, medium was replaced with mTESR™1 + ROCK inhibitor + selection antibiotics (3.5 μg/mL blastocidin and 0.5 μg/mL puromicin), which was maintained

for 48h. Selected cells were then grown subsequently on feeder-free medium only until colony formation was apparent. Individual colonies were transferred to Matrigel®-coated 12-well plates for amplification of cell number; 12 colonies were selected at random from each line for screening.

2.4.3 Genomic DNA extraction and screening

DNA was extracted from each colony-forming unit of each transfected line using the QiaPrep® Plasmid MiniPrep Kit (Qiagen). The presence of the SVA insertion was determined using a step-down PCR protocol as described above, and visualized on 1% agarose gel. Sequences of the expected junction regions were determined via Sanger sequencing using the SVA retrotransposon insertion forward sequencing primer, i.e. 5'-GTTCCATTGTGTGGTTGTACCAGCGTTTGTTC-3'.

2.4.4 *TAF1* expression analysis

RNA was extracted from successfully edited, unedited, and from parent iPS cell lines using a standard kit (Qiagen RNeasy® Mini Kit) and after DNAseI treatment. Reverse transcription was performed as described above. *TAF1* expression analyses were performed as in previous expression experiments (Sections 3.2.2 and 3.2.3 above), using the same primers, and using *ACTB* and *FPGS* as reference genes.

First, group analysis of *TAF1* expression was performed between all parent, untransfected lines vs. all successfully edited colonies vs. all unedited colonies, after normalization against the average expression in unedited controls (L-2131). Second, each XDP and control line was analyzed individually. The relative expression of *TAF1* in each edited or unedited colony from the same line was first determined by normalizing against either its parent untransfected line (L-7149), or against one unedited colony, i.e. the unedited colony from the same line with the lowest absolute expression value (for L-5748, L-7994, L-2131), given that in these lines, the parent had to be re-thawed from stock, thus not representing a "true" parent line. The average value among successfully edited and unedited colonies was then determined, and these were compared against each other to determine the absolute FC.

2.5 Statistical analyses

Relevant statistical tests were performed either using Microsoft Excel®, GraphPad Prism™, or R. For high-throughput experiments, FDR-corrected q-values were used, at significance level $q<0.10$. For comparison of two groups, p-values were determined using Student's t-tests, and the level of significance was set at $p<0.05$. Boxplots were rendered using an R-based tool, BoxplotR (http://boxplot.tyerslab.com).

3 Results

3.1 Sequencing, haplotype, and genomic analyses

The results in this section cover **Objectives 1.1** to **2.2.**

3.1.1 Targeted sequencing and STRP marker analysis

Targeted sequencing (DSC1, DSC2, DSC3, DSC10, DSC12, 48-bp deletion) and PCR-based detection (SVA retrotransposon insertion) of previously described genetic alterations in 163 XDP patients revealed i) occurrence of all genetic changes in 158, ii) wild-type alleles in all loci in 5, iii) in all cases, co-occurrence of each genetic alteration with all the others, i.e. none of the changes were observed in isolation in any patient sample (Table 2).

Genetic analyses in 473 controls revealed only one 35-year old male who carried all genetic alterations. In accordance with the findings in patients, singly occurring changes were not observed in any control sample.

Analysis of STRP markers revealed shared alleles among most XDP patients at DXS10015, DXS10016, DXS10017, DXS10018, and DXS559. Small differences in allele length can be explained by 'slippage mutations' (Table 2 and Figure 6).

The above data support a non-random association of all seven genetic alterations specific for XDP, i.e. the changes seem to be in complete linkage disequilibrium. Furthermore, STRP marker analysis revealed a common 'XDP haplotype' carried by patients in a 427-kB region subtended by DXS10015 (centromeric) and DXS559 (Figure 6). Thus, none of the previously identified genetic alterations were ruled-out via targeted sequencing in patients and controls as not disease-specific.

3.1.2 Genome sequencing analysis

The entire genomes of three carriers (XDP-hap+: L-5709, L-5712, L-7153) and two non-carriers (XDP-hap0: L-5708, L-7677) of the common XDP haplotype were sequenced. SNV search was performed first on genome data from all three XDP-hap+, and on the genome sequence of the female homozygous carrier of the haplotype (L-7159), i.e. the mother of L-7153. This revealed 13 SNVs between DXS10015 and DXS559, i.e. proximal and distal ends of the linked region, that were common among the four genomes searched; all SNVs were intronic/intergenic, and five of these were the previously described DSCs (Table 3).

Outside of the linked region, there were no exonic SNVs shared by all three XDP-hap+, and none with either of the two XDP-hap0 genomes. Thus, no novel protein-coding mutations were detected by the genome sequencing analysis performed.

Table 2 **Genotypes at STRPs, DSCs, and the loci of the SVA insertion and 48-bp deletion.**

n	DXS 10015	DXS 10016	DXS 10017	DSC 12	SVA ins	DSC 10	48bp del	DSC 1	DSC 3	DSC 2	DXS 10018	DXS 559
130	137	127	293	G	+	T	del	A	T	A	139	238
4	137	129	293	G	+	T	del	A	T	A	139	238
1	137	127	297	G	+	T	del	A	T	A	139	238
4	141	127	293	G	+	T	del	A	T	A	139	238
1	137	127	293	G	+	T	del	A	T	A	143	238
6	137	127	293	G	+	T	del	A	T	A	139	236
7	137	127	293	G	+	T	del	A	T	A	139	240
5	137	127	293	G	+	T	del	A	T	A	139	242
1	133	135	291	T	-	C	wt	T	C	G	141	238
1	144	127	291	T	-	C	wt	T	C	G	145	232
1	133	135	291	T	-	C	wt	T	C	G	143	238
1	144	127	291	T	-	C	wt	T	C	G	139	232
1	133	135	291	T	-	C	wt	T	C	G	139	238

n indicates number of XDP patients carrying a specific haplotype (total 163).
'Slippage mutations' in some STRPs were observed (blue), but allele lengths were otherwise shared among 158 patients with the haplotype.
Last five rows contain alleles for patients who were haplotype non-carriers (XDP-hap0).

Table 3 **Variants discovered in the linked region via genome sequencing.**

GRCh38 position	dbSNP ID	Ref	Obs	Notes
ChrX:71262608	rs41377154	G	A	first SNV distal to DXS10015
ChrX:71268784	rs41339045	A	T	
ChrX:71301439	novel	GG	G	
ChrX:71339837	rs41532445	G	A	
ChrX:71344555	rs41416246	C	T	
ChrX:71391195		T	G	DSC12
ChrX:71441489		C	T	DSC10
ChrX:71483870	rs4844149	C	A	as part of a repeat polymorphism
ChrX:71513660		T	A	DSC1
ChrX:71529785		C	T	DSC3
ChrX:71531433		G	A	DSC2
ChrX:71633571	rs41438158	C	T	
ChrX:71653235	novel	C	T	last SNV proximal to DXS559

Blue – five SNVs between linked STRP markers and the DSCs that were then genotyped in additional patients and controls.

3.1.3 SNV validation and narrowing of the XDP haplotype

Among the eight SNVs that were discovered by genome sequencing (beyond the DSCs), five were in regions between linked STRP markers and the DSCs (Table 3). These were then sequenced in all other XDP patient samples and in controls to determine the exact boundaries of the XDP haplotype. Targeted sequencing revealed non-segregating alleles in at least five patients at rs41438158 – which is downstream

to DSC2 but centromeric to DXS559 – thus establishing the distal recombination event at this position. In two patients, alternative alleles at three loci between DXS10015 and DSC12 establish the proximal end of the haplotype at rs41532445. Thus, validation of discovered SNVs in the entire patient cohort further narrowed the linked region to a 294.7kb region that includes four genes, i.e. *TAF1, OGT, ACRC,* and *CXCR3* (Figure 6).

3.1.4 CNV analysis

CNV events in the X chromosome were detected in 14/105 samples (eight distinct CNVs, of which seven were deletions), however, none were novel nor identified in more than a small subset of the samples subjected to genome-wide SNP genotyping. Furthermore, no CNV loci were observed within the XDP haplotype, and no disease-specific CNVs were detected genome-wide.

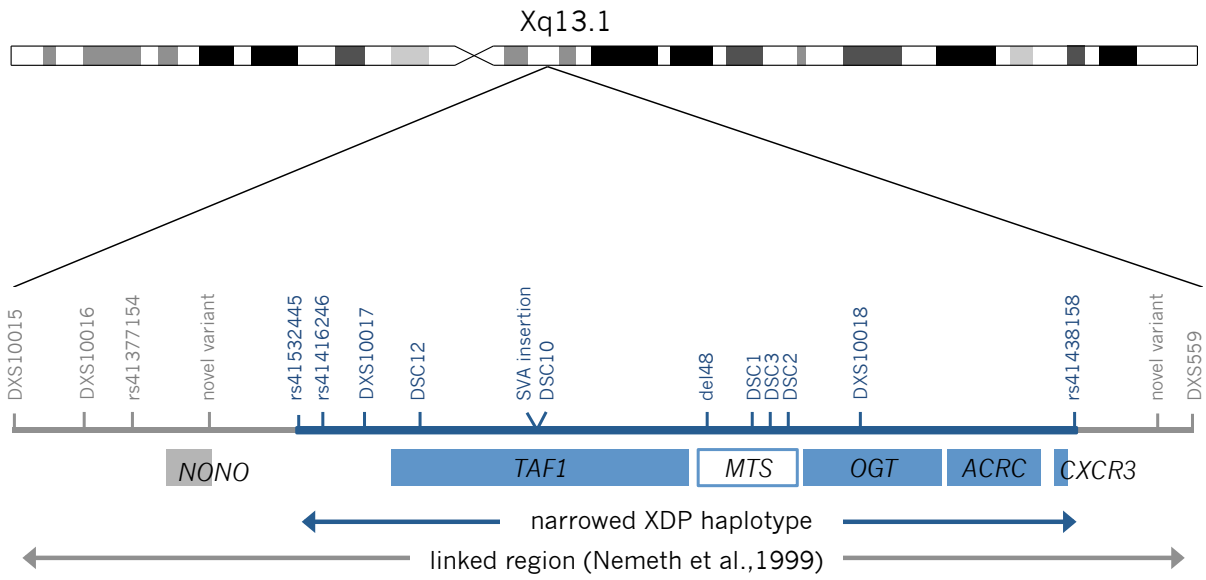

Figure 6 XDP haplotype. The combination of next-generation sequencing and Sanger-based validation of discovered SNVs enabled narrowing of the XDP haplotype from a 427kb region (reported as 350kb by Ref. [23]) between DXS10015 and DXS559, to a 294.7kb segment that contains four known coding genes: *TAF1, OGT, ACRC,* and *CXCR3*. The genetic alterations found in XDP patients are all found either in introns of *TAF1* or in the noncoding MTS distal to it. Figure adapted from Domingo et al., *Eur. J. Hum. Genet.,* 2015.

3.2 Expression and transcriptomic analyses

The results in this section cover **Objectives 3.1** to **3.3.**

3.2.1 Analysis of *TAF1* expression in blood and fibroblasts

Using qPCR, a statistically significant reduction of *TAF1* expression was seen in both blood- and fibroblast-derived RNA in the XDP group (p-value in blood <0.01, in fibroblasts <0.05, FC in both tissues <1.5) (Figure 7). Transcripts of *nTAF1*, or of *DYT3*/MTS, were extremely low in expression level. The error encountered during late detection of the transcripts during real-time qPCR thus did not allow for robust calculation of differences in expression. The expression of *OGT* and *CXCR3*, i.e. two other genes inside the XDP-linked region, were not changed between XDP and control samples.

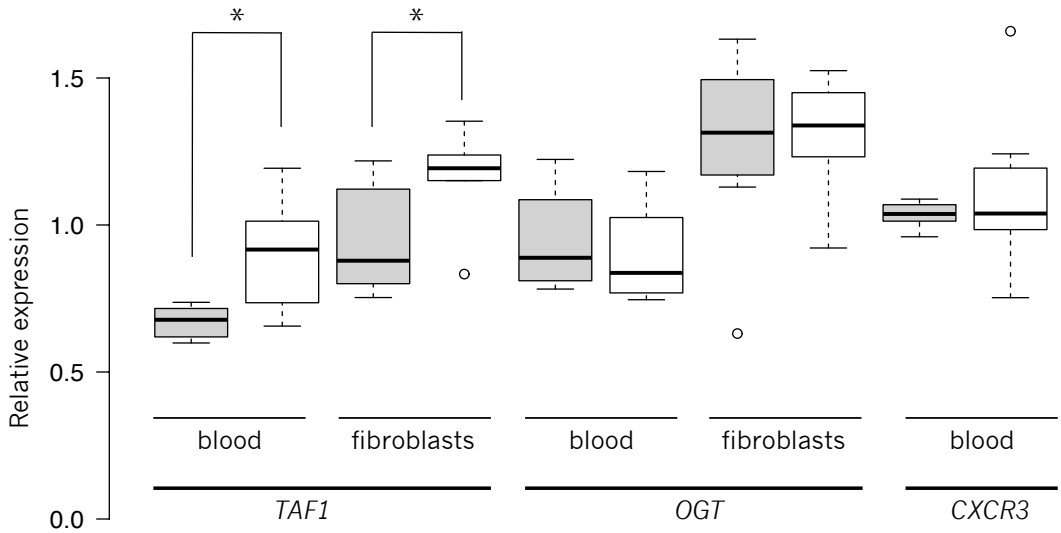

Figure 7 Relative expressions of genes in the XDP-linked region (XDP – gray boxes, controls – white boxes). A 1.3-fold reduction in the relative expression of *TAF1* was observed in the XDP group, in experiments performed in blood- and fibroblast-derived RNA. No significant differences were observed in the expression of *OGT* and *CXCR3*. Boxplot whiskers extend to minimum/maximum data points that are within 1.5x the interquartal ranges, and outliers are indicated by open circles. Relative expression was normalized against expression of EARs. Figure adapted from Domingo et al., *Cell. Mol. Life Sci.,* 2016.

3.2.2 Microarray expression profiling

In the microarray-based expression profiling experiment performed on fibroblast-derived RNA i) 307 differentially expressed genes (DEGs) were identified, representing 2.4% of the total number of transcripts analyzed, ii) 250 transcripts were upregulated and 57 were downregulated in the XDP group, iii) with the exception of a handful, all

DEGs discovered had modest fold-changes, i.e. between 1.1 and 2.0 (Figure 8). In contrast, no DEGs were identified in the blood microarray experiment, even when the q-value cut-off was relaxed to 0.20.

TAF1 was one of the downregulated DEGs, consistent with the results of the qPCR experiments. Two microarray probes targeting *TAF1* transcripts (A_23_P11237, A_32_P192615) showed lower expression in the XDP group compared to controls (FC=1.2, averaged q-value=0.03).

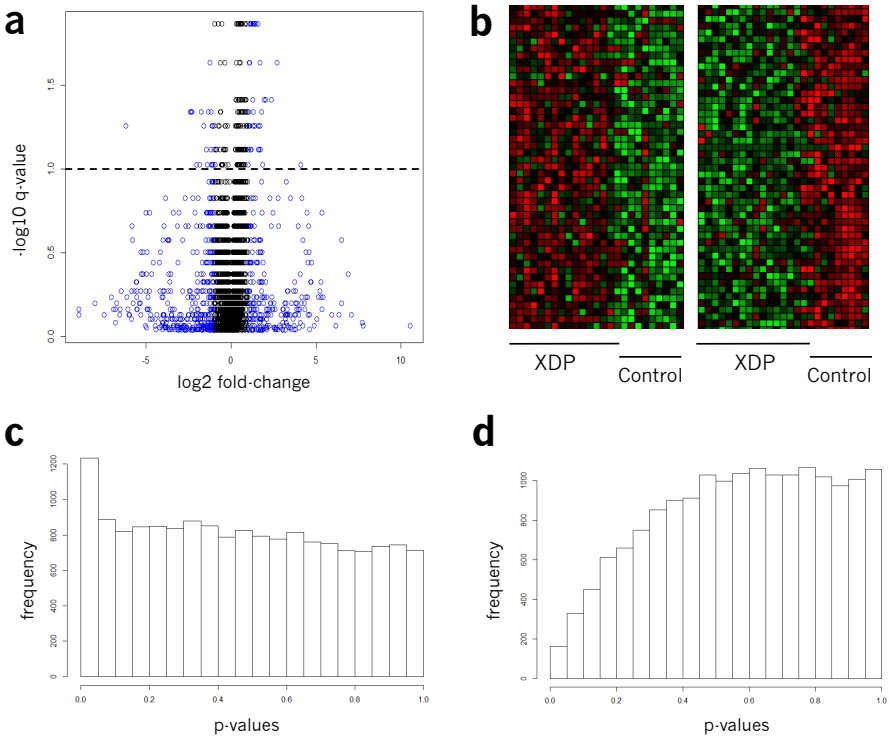

Figure 8 Features of the microarray-based genome-wide expression profiling experiments. *a-c* are analyses performed on the experiment in fibroblasts. Most DEGs in the XDP group (blue circles) had modest fold-changes, as shown in the volcano plot (*a*). Clear differences in between the XDP and control groups can be seen in the heat map of top DEGs (*b*). A high frequency of genes have p-values <0.10 in the histogram (*c*) – indicating significant differential expression, in contrast to the experiment performed in blood, where the p-values are skewed towards 1.0 (*d*).

3.2.3 Validation experiments

Validation analysis was performed first on the discovery set of RNAs, on 12 selected DEGs (six downregulated and six upregulated genes, Table 1). Significant differential expression was confirmed via qPCR in 7/12 DEGs in fibroblast-derived RNA, i.e. p-value was <0.10 in *SYTL2*, *EFNB1*, *SLC5A3*, *MRPS6*, *ATP6V0E2*, *BHLHE40*, and *KCND2* (Figure 9).

DEGs that were validated in the preceding experiment, two additional *DEGs* (*ETV3* and *DUSP1*, both upregulated in the microarray experiment), and *TAF1* were then further investigated using technical and independent replicates. Consistently lower expression of *TAF1* in the XDP group was observed in experiments with 18 samples (10 XDP/8 controls) (p-value=0.03, FC=1.3) as with experiments with 23 samples (13 XDP/10 controls) (p-value=2.1 x 10^{-6}, FC=1.3). Consistent downregulation of *SYTL2* (p-value=0.01), and *SLC5A3* (p-value=0.08-0.09) in the XDP group were also seen in the replication experiments (Figure 9).

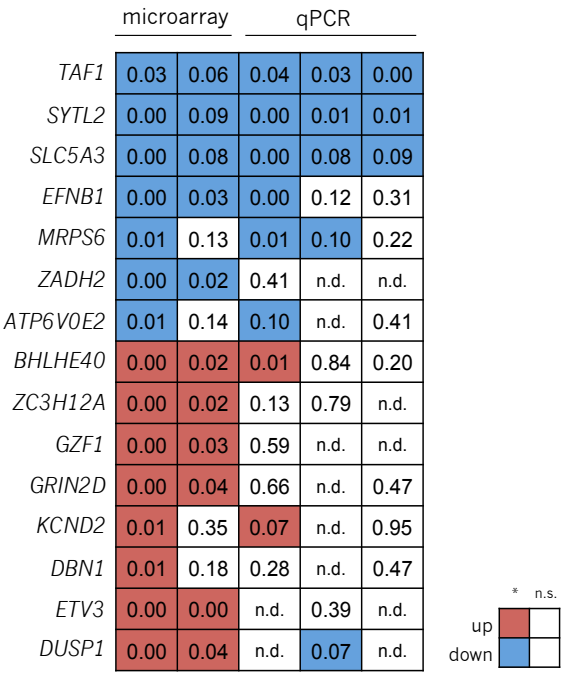

Figure 9 p-values in microarray and qPCR experiments. Shown are p-values in the series of experiments (discovery, replication, 2nd replication) performed in fibroblasts-derived RNA. *TAF1* was consistently downregulated in all experiments, as was *SYTL2* and *SLC5A3* (q- and p-values<0.10). Figure adapted from Domingo et al., *Cell. Mol. Life Sci.*, 2016.

n.d. – not done n.s. – not significant

3.2.4 Sample-level analysis

Lower expression of *TAF1* was observed in all but one sample obtained from an affected individual, when compared to average expression in controls in fibroblast-derived RNA (log2fc range: 0.07 to -0.72), and in all samples from blood (log2fc range: -0.86 to -2.80) (Figure 10). In the analysis performed in blood, greatest downregulation was observed in two samples that came from affected individuals with longer disease duration and/or more severe parkinsonism.

Thus, high-throughput (microarray-based) and gene-level (qPCR) experiments in both group- and sample-level analyses show reduced *TAF1* expression in peripheral tissue obtained from XDP patients, i.e. blood and fibroblasts.

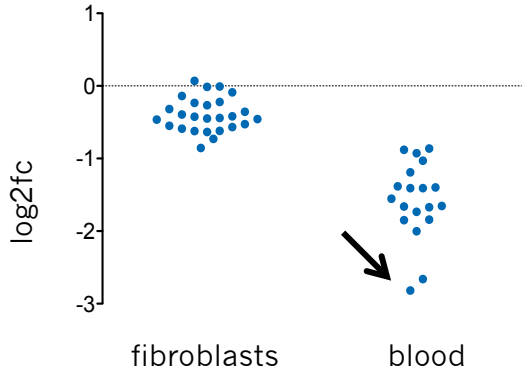

Figure 10 Sample level expression analyses. In both blood and fibroblast-based experiments, all but one sample had reduced *TAF1* expression in comparison to controls. Two samples with downregulation of >2-fold came from patients with long disease duration (arrow). Each data point represents one sample from an XDP patient (discovery and replication sets combined).

3.2.5 Enrichment analysis

Significantly enriched GO terms seen in upregulated DEGs in the fibroblast microarray experiment included i) "vasculature development" (q-value=6.0 x 10^{-4}) ii) "regulation of cell differentiation" (q-value=5.2 x 10^{-3}), and iii) "regulation of transcription from RNA polymerase II promoter" (q-value=1.2 x 10^{-2}). There were no enrichments seen in downregulated DEGs upon GO and transcription factor analysis. There were no significantly enriched KEGG pathways/Wikipathways or miRNA enrichments seen. Gene set enrichment analysis (GSEA) revealed one enriched gene set at 0.2 FDR: "cell cycle" (q-value=3.0 x 10^{-2}).

Network analysis using GeneMania revealed enriched GO terms in upregulated DEGs that were similar to the ones discovered using EXPANDER GO analysis (Table 4 and Figure 11). Redundant terms included "DNA-dependent transcription" (q-value=3.5 x 10^{-2}) and "negative regulation of sequence-specific DNA-binding transcription factors" (q-value=4.5 x 10^{-2}, "response to oxidative stress" (q-value=3.5 x 10^{-3}) and "response to reactive oxygen species" (q-value=4.5 x 10^{-2}), and "actin binding" (q-value=1.4 x 10^{-2}) and "actin cytoskeleton" (p-value=4.7 x 10^{-2}), which were related to the tissue derivation (cultured fibroblasts).

Table 4 Significantly enriched Gene Ontology (GO) terms among DEGs.

q-value	GO Annotation/s
1.5x10^{-7}	Adherens junction, anchoring junction, focal adhesions
3.5x10^{-3}	Response to oxidative stress
8.6x10^{-3}	Blood vessel development, blood vessel morphogenesis
1.4x10^{-2}	Actin binding, actin cytoskeleton organization
3.5x10^{-2}	DNA-dependent transcription, initiation
4.5x10^{-2}	Negative regulation of sequence-specific DNA binding transcription factors
4.5x10^{-2}	Response to reactive oxygen species

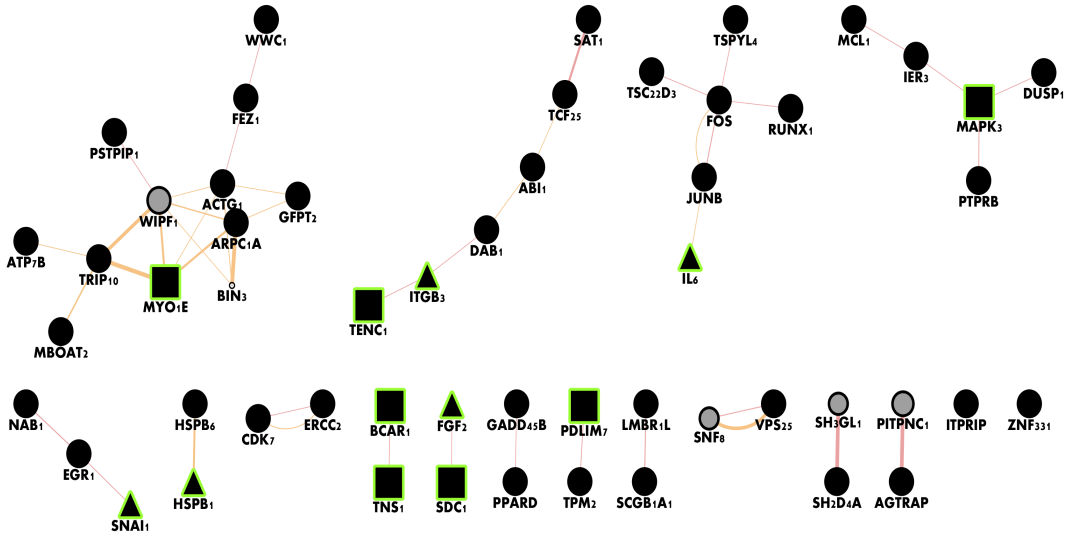

Figure 11 Network analysis using GeneMania. Overexpressed genes in the microarray as input revealed active connected components. Black nodes denote differentially expressed genes and lines denote physical interactions. Grey nodes: added by the GeneMania analysis. Rectangular nodes are related to adherens junction (reflecting tissue derivation) and triangular nodes are related to positive regulation of cell development, likely related to *TAF1* function in cells. Figure adapted from Domingo et al., *Cell. Mol. Life Sci.,* 2016.

3.3 Analyses of *TAF1* isoforms

The results in this section cover **Objective 4.1** and **4.2**.

3.3.1 Molecular cloning

Out of 60 recombinant colonies from cloning experiments performed using dopaminergic neuron-derived cDNA, 14 harbored exon 34' (23.33%). For striatal neuron-derived cDNA, 7/70 expressed exon 34' (10%) (Figure 12).

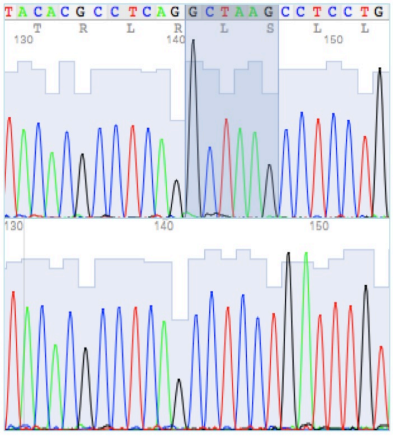

Figure 12 Identification of nTAF1 transcripts. Sanger sequencing identifies exon 34' (GCTAAG) in recombinant colonies (*top*). Bottom figure shows sequence in non-exon 34' expressing colonies.

3.3.2 Optimized qPCR assay

The assay designed to detect *nTAF1* and non-*nTAF1* expression was judged to be highly specific in that: i) no amplification was seen in blood-, fibroblast-, and iPS cell-derived cDNA when *nTAF1*-specific primers were used, ii) no amplification was seen when using *nTAF1* primers with *cTAF1* plasmid DNA (and vice versa, *cTAF1* primers with *nTAF1* plasmid DNA), both at concentrations approximating *nTAF1*/*cTAF1* expression in iPS cell and neuronal

models, as well as at significantly higher (100-, 1000-fold) concentrations, iii) a linear curve could be established using different proportions of *nTAF1* and *cTAF1* plasmid DNA, with the assay correctly quantifying the amount of *nTAF1* concentration.

3.3.3 Isoform analysis in endogenous models

The highly specific qPCR assay was then used to determine *nTAF1* expression in cortical and striatal neurons, in relation to total *TAF1* expression. In cortical neurons, *nTAF1* represented 26.80% of total *TAF1* transcripts. In striatal neurons, *nTAF1* expression was 34.96% of total *TAF1* expression. Both molecular cloning and qPCR analyses thus confirm the relative abundance of the *nTAF1* isoform in neuronal models derived from iPS cells.

Furthermore, in striatal neurons, consistent with experiments performed in blood and fibroblasts, *cTAF1* expression was reduced in the XDP group in comparison controls (FC=1.15). *nTAF1* was also reduced (FC=1.30) (Figure 13).

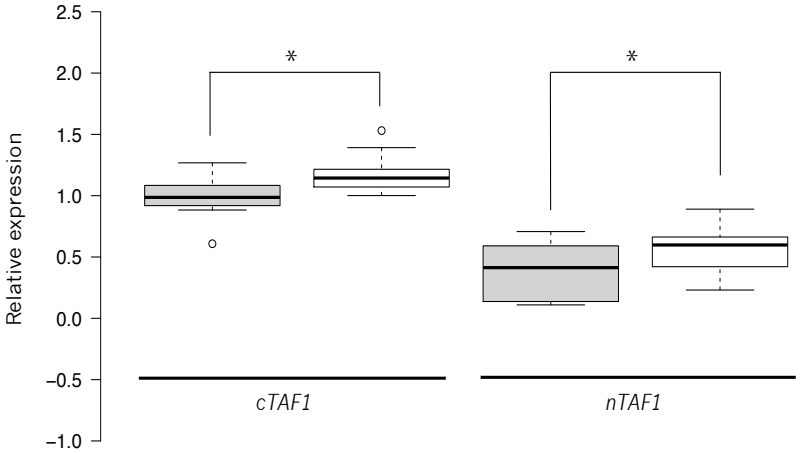

Figure 13 Relative expression of *cTAF1* and *nTAF1* in XDP (gray boxes) vs. controls (white boxes). Experiment performed on striatal neurons**.** Reduced expression of both common and neuronal forms of *TAF1* were seen in the XDP group. Boxplot whiskers extend to minimum/maximum data points that are within 1.5x the interquartal ranges. Outliers are indicated by open circles and are not included in the calculation of the median (bar).

3.4 Genome editing

The results in this section cover **Objective 5.1** and **5.2.**

3.4.1 CRISPR/Cas9 genome editing

PCR-based and sequencing analysis revealed successful excision of the SVA retrotransposon insertion in three iPS cell lines from XDP patients, in about 10-30% of colonies screened (Figure 14). In successfully edited XDP lines, the entire genomic region encompassing the insertion was edited out (Figure 15).

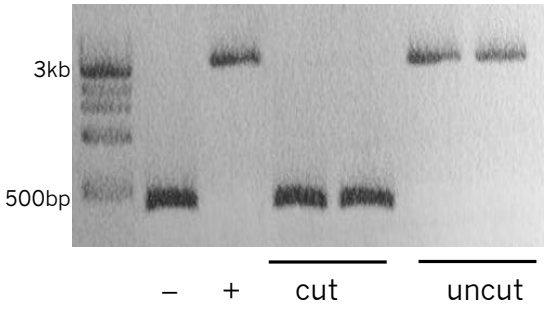

Figure 14 PCR-based analyses of the SVA insertion after CRISPR/Cas9-based genome editing. Shown are results for selected successfully cut and uncut colonies in comparison to negative and positive controls. *Leftmost* is 1kb-ladder.

Notably, a 165/166 bp genomic segment immediately proximal/distal to the locus of the insertion was excised additionally in the control and in all XDP lines (i.e., 120 bp proximal to the first base of the SVA insertion, 45/46 bp after the last base), representing a non-homologous end-joining-mediated indel/'scar'.

Figure 15 Sequencing analyses after CRISRP/Cas9-mediated genome editing. Sequences in (*a*) unedited control line (*b*) unedited XDP line (*c*) successfully edited XDP line. Sequence in *a* and *b* are similar until the start site of the SVA insertion. The cut site does not change the sequence. In the edited colony (*c*), 120 bp are excised proximal to the start of the SVA insertion. Following the cut site is sequence that is already 46 bp distal to the SVA.

Cas-OFFinder revealed two possible off-target sites. However, both were targeted only by gRNA1, had the sequence 5'-TTTAGTTTTACAAGACAAGG-3', and were located at Chr5:131093433 and Chr22:31765714. No DNA segments were targeted by both gRNAs genome-wide, indicating specificity of the designed CRISPR/Cas9 system.

3.4.2 Analysis of *TAF1* expression

In lines derived from XDP patients, relative and normalized *TAF1* expression was i) not significantly different between parent lines and unedited colonies, but ii) was significantly increased in colonies where the SVA retrotransposon insertion was successfully excised (Figure 16). Interestingly, increase in *TAF1* expression was also observed in edited colonies in the control line.

Consistent with the results of previous expression analyses in blood, fibroblasts, and in iPS cell-derived striatal neurons, lower *TAF1* expression was observed in XDP lines in general. *TAF1* expression after CRISPR/Cas9-mediated excision of the SVA retrotransposon insertion was not significantly different from the expression level in unedited colonies from the control line (p-value>0.05), i.e. the 'rescue' was able to approximate *TAF1* expression in controls (Figure 16).

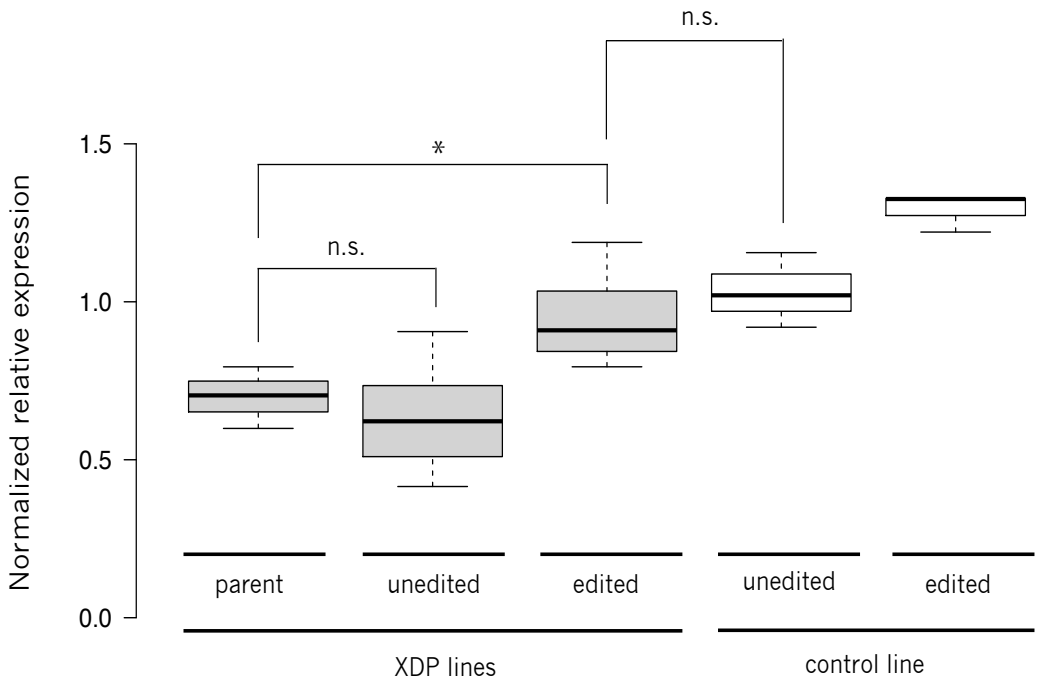

Figure 16 Group-level analysis of TAF1 expression after CRISPR/Cas9-mediated genome editing. Shown are parent, unedited, and edited colonies from XDP (grey boxes) and control lines (white boxes). Significant increase in *TAF1* expression was seen in colonies where the SVA insertion was successfully excised, when compared to parent lines, and to unedited colonies. The increase resulted in *TAF1* expression ('rescue') approximating that in unedited colonies from the control line.

The magnitude of difference in *TAF1* expression (between edited and unedited colonies) in each line was determined after normalization with the parent or with the expression in one unedited colony. Significant difference was seen in all XDP lines (FC: 1.3 to 1.8, $p<0.05$), but not in the control line (FC=1.1, $p>0.05$) (Figure 17), indicating specificity of the increase in CRISPR-mediated *TAF1* expression to lines obtained from XDP patients.

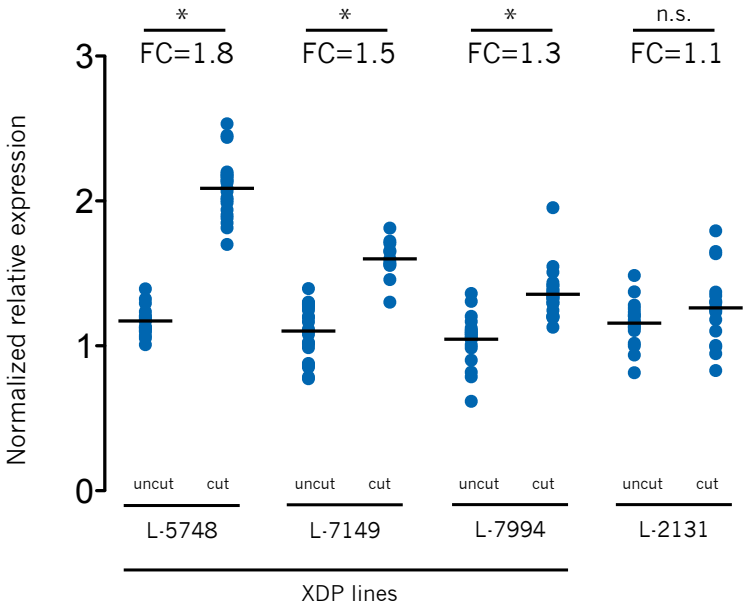

Figure 17 Analysis of relative TAF1 expression in individual lines. Significant difference in relative *TAF1* expression between unedited/uncut and edited/cut colonies was seen only in XDP lines, indicating that the 'rescue' of *TAF1* underexpression is specific to the XDP group. Each circle represents one experimental determination of *TAF1* expression (thus, up to 3 per colony, per line).

4 Discussion

XDP is a unique movement disorder that is thus far found only in individuals of Filipino descent. Although previous molecular genetic studies have identified genetic alterations that have been confirmed to be specific to affected individuals, the genetic basis of this orphan disease is still incompletely clarified. None of the DNA alterations occur in protein-coding regions based on current annotations of the human genome. Furthermore, their causative link to the reduced expression of *TAF1* isoforms and to the striatal pathology that has been observed in the brains of XDP patients has not been established.

In this work, the genetic and molecular mechanisms surrounding XDP were investigated at both genomic (DNA and genome sequencing, haplotype analysis) and transcriptomic (expression profiling, quantitative gene expression analysis) levels. Biomaterials from patients and controls (blood, fibroblasts) as well as endogenous cell models (iPS cells, iPS cell-derived neurons) were used. A range of techniques were employed to achieve the objectives: ranging from high-throughput methods (next-generation sequencing, SNP array, microarray-based expression profiling) to gene-level assays (targeted sequencing, qPCR, molecular cloning), complemented throughout by bioinformatic approaches (annotation, enrichment and network analysis). Lastly, state-of-the-art CRISPR/Cas9-mediated genome editing was performed to identify the role of the SVA retrotransposon insertion to *TAF1* dysregulation. Importantly, these molecular genetic studies were performed using the largest gathering of XDP-related samples and biomaterials in the world to date, a collection made possible only through international collaboration, but also via PhD candidate's personal effort in multiple field work activities in the Philippines.

4.1 The genetic alterations in XDP form a single haplotype

The first hypothesis posited was that there was only one truly pathogenic mutation amongst the genetic alterations associated with XDP, and that the rest were merely 'passengers' in the haplotype, denoting the locus of the disease-causing change but functionally benign. The rationale was the improbability of the situation that seven independent genetic variants arose simultaneously in the first XDP case (i.e. the XDP genetic founder); rather, the more plausible scenario imagined was of the genetic founder belonging to a population in which six of seven XDP-associated variants were already present. A subsequent, seventh (disease-causing) mutation could have then arose in him, producing the first XDP case.

Thus, genotyping of all currently known genetic variants associated with the disease in a large group of patients and in a carefully chosen cohort of geographically and ethnically matched controls was done to identify this pre-XDP population. In parallel, next-generation (genome) sequencing and microarray-based CNV analysis was performed to identify protein-coding changes that have not been detected by previous technologies used to elucidate the genetic basis of XDP.

However, no individual carrying any of the original six genetic variants was identified in the genetic screening of controls from Panay island. One explanation could be that the haplotype was brought onto the island from an (unknown) geographical region, and that the de-novo disease-causing and seventh mutation occurred locally later. Supporting this notion is one well-known Filipino legend that tells the story of ten royals from the island of Borneo (part of present-day Malaysia) who established themselves on Panay island during the early seafaring era. However, these orally transmitted histories have not been confirmed from an anthropological point of view, and public genome databases that include Malay genomes also exclude the existence of the XDP-associated genetic variants in other populations, including the Malays [92], [93]. One other explanation could be that the pre-XDP population was obliterated by a genetic bottleneck (i.e. a natural calamity) in the past. No such catastrophic event is known to have occurred on the island, although it might have happened in pre-history.

The combination of genome sequencing and CNV detection via SNP genotyping also did not identify a novel genetic alteration that could be equivocally considered as the pathogenic mutation in XDP. Of note, genetic alterations that escape these NGS methods are not completely ruled out, i.e. large insertions or deletions, translocations, chromosomal inversions, and mosaicism [94]. One study for example describes an 8.5-Mb inversion in Xq13.1 that spans the entirety of the XDP haplotype in one individual, including 57 genes [95]. Also, Xq13.1 harbors breakpoints for chromosomal inversions associated with X-linked mental retardation [96]. To date, no formal experiments have been done to detect inversions of and within the XDP haplotype, although previous Southern blotting in the region around the *ACRC* gene has not revealed rearrangements [25].

Interestingly, a disease-specific 'XDP haplotype' was observed in almost all XDP patients investigated, consisting of genetic variants occurring together in complete linkage disequilibrium. Five symptomatic individuals that were non-carriers of XDP haplotype (XDP-hap0) were identified, possibly representing phenocopies. Other adult-onset genetic movement disorders that have been described in Filipinos include ataxia due to mutations in *KCNC3*, previously described in a Filipino pedigree from Panay

island [97], and early-onset Parkinson's disease caused by a Filipino founder mutation in *PINK1* [98,99]. Although the strikingly high prevalence of XDP on Panay island supports a founder mutation and genetic homogeneity, in other forms of dystonia, homogeneous communities have been found to harbor a variety of pathogenic mutations, supporting heterogeneity of causes. For example, in Amish-Mennonites with primary dystonia and in whom the founder mutation in *THAP1* was not identifiable, other variants in the same gene as well as mutations in other genes (*GNAL* and *TOR1A*) were later discovered to be causative for the phenotype [100,101]. Given that there is as yet no truly convincing protein-coding alteration harbored by all XDP patients, one should keep in mind the possibility of genetic heterogeneity in this disease, misleading linkage studies.

Importantly, the identification of new SNVs using genome sequencing coupled with their use in segregation and haplotype analysis enabled narrowing of the XDP haplotype to a 294.7-kb genomic region. This facilitates further focusing of molecular genetic studies to the involvement of the *TAF1*, given that this gene is included in this narrowed genomic region, and because the genetic alterations associated with the disease converge around this gene. In particular, the SVA retrotransposon insertion is located in intron 32 and the DSCs are located around the MTS. Though not protein-coding themselves, these variants could theoretically alter the function of regulators of the expression of protein-coding genes within the disease locus, in particular, and as hypothesized, of *TAF1*.

4.2 *TAF1* expression is dysregulated in peripheral models of XDP

TAF1 codes for the largest component of the TFIID transcription factor complex, a general transcription factor that facilitates RNA polymerase II-dependent transcription of many protein-coding genes [38]. Alteration of *TAF1* function in XDP occurring at the expression level was therefore hypothesized to also affect global gene expression, given the critical function of *TAF1* in transcriptional processes. Furthermore, dysfunction of ubiquitously expressed *TAF1* does not necessarily have to be manifest only in the primary tissue where the phenotype is obvious (i.e. brain), but may also be present in more easily accessible peripheral tissue [52,102,103]. Thus, dysregulated *TAF1* expression in peripheral models of XDP, i.e. i) blood, and ii) fibroblasts, the latter an endogenous cell model with the same embryonic origin as neurons, was studied.

There was consistent downregulation of *TAF1* transcripts in the XDP group in both high-throughput (microarray-based) and gene-level (qPCR) experiments. Although the magnitude of downregulation was small (i.e. FC<1.5), the signal was reliably seen in

consistently significant p- and q-values, statistics that incorporate variability. While fold-change reflects effect size, it inadequately accounts for variance between comparisons, and arbitrary cutoffs (such as >2.0, as in most expression studies) limit the analysis to DEGs that vary considerably between groups without consideration of other factors such as biological relevance or consistency [104,105].

TAF1 expression was reduced in the XDP group in both blood and fibroblasts in qPCR experiments and in the fibroblast microarray experiment; *TAF1* was also the only gene within the XDP haplotype showing expression difference. Thus, despite the small fold-change, the hypothesis of *TAF1* dysregulation was followed-up. In the qPCR experiments performed on iPS cell-derived striatal neurons later, for example, *TAF1* underexpression was also evident in the XDP group. Due to heterogeneity of global expression profiles, statistical analysis did not show significant *TAF1* dysregulation in the microarray profiling experiment performed in blood. Nevertheless, taking all experiments as a whole, there was robust evidence of *TAF1* dysfunction occurring at the expression level in XDP.

Notably, most other DEGs in the fibroblast microarray experiment displayed small fold-changes as well, an effect which may be due to the fact that the expression profiling experiment was carried out in an endogenous cell model and not in an overexpressed system. The only DEGs with FC>2.0 were a handful of genes with low expression at baseline – predisposing to large, even exponential fold-changes of differential expression. Globally reduced variation in the expression of housekeeping and cell cycle genes in cultured cells is common, due to controlled cell culture conditions and absence of noise [106]. Alternatively, the signal obtained from the transcriptome in the expression profiling experiment can be inferred to be primarily due to the mutation status – as there are no environmental influences – thus providing a link between the DNA changes in XDP and the molecular phenotype of dysregulated *TAF1* expression.

Despite the low magnitude of downregulation in *TAF1*, there was an extensive signal of global gene expression changes, reflected in 307 genes being differentially expressed in the XDP group in the microarray experiment performed on fibroblasts. This represents a significant downstream effect that supports dysregulation of a critical component of the cell's transcriptional machinery, in this case, *TAF1* activation of gene expression. Furthermore, in network and GO analyses, there was a clear and consistent enrichment of genes involved in RNA polymerase II-based transcriptional processes, where *TAF1* is involved.

XDP is frequently cited as a transcriptional dysregulation syndrome, as is Huntington's disease, with which XDP shares similarities in brain pathology [16], [107]. *TAF7*, which also codes for a transcriptional coactivator that interacts closely with *TAF1* as part of TFIID complex, is dysregulated in Huntington's disease [47], [108]. Thus, the enrichment of genes involved in transcriptional processes in the experiments here could denote a direct downstream effect of *TAF1* dysregulation, or of genome-wide alterations in transcription factor activity, corollary to *TAF1* dysfunction. One could speculate altered interaction between the activator (TAF1 protein) and the transcription factor (TFIID), as has been shown for transgenic spinocerebellar ataxia 17 mice with mutations in the TATA-box binding protein [109]. The resulting dysregulation of target genes has been implicated in the occurrence of neurodegeneration in this model [110].

Aside from *TAF1* dysregulation, consistent signatures were also seen in *SYTL2* and *SLC5A3*. Synaptotagmin-like 2 is a RAB27A interactor that participates in membrane trafficking [111,112]. Interestingly, expression profiling in neuroblastomas overexpressing MTS transcripts also identified altered vesicle trafficking and synaptogenesis as a significant signature using pathway analysis [28]. Furthermore, disruption of endosomal membrane trafficking is disrupted in many neurodegenerative diseases, including genetic forms of Parkinson's disease [29,30]. *SLC5A3* meanwhile codes for a sodium/myoinositol transporter that regulates brain osmoregulation [113]. Dysfunction of inositol metabolism in Huntington's disease is observed as loss of the inositol polyphosphate multikinase enzyme in the striatum in humans, and in cellular and murine models [114].

Other interesting DEGs identified include *EFNB1*, mutations in which cause a neurodevelopmental phenotype (craniofrontonasal syndrome, OMIM #304110) [115], which could be related to the facial dysmorphism phenotype observed due to exonic and splicing mutations in *TAF1* [43]. Furthermore, *EFNB1* is located (as is *TAF1*) on Xq13.1. *GRIN2D*, which codes for a glutamate receptor that is also differentially expressed in Parkinson's disease [116], is also overexpressed in the microarray experiment. *BHLHE40* encodes a transcription factor and is also upregulated in the microarray and validation experiments, interacts with the TFIID complex to promote differentiation [117], while also mediating p53-dependent premature senescence [118]. The contribution of these genes to the molecular mechanisms and the neurodegenerative phenotype of XDP should be investigated in further studies.

Although the experimental setup was not designed to identify biomarkers in the peripheral tissues studied, more striking downregulation of *TAF1* was identified in two

affected individuals with longer disease duration (and more predominant parkinsonian symptoms). XDP is a progressive disease, and studies in brain (imaging, pathology) report greater neuronal loss in patients in the parkinsonian phase of the disease [4], [18]. Thus, the increasing downregulation of *TAF1* with longer disease duration, and hypothetically, greater neuronal loss, is intriguing. In relation to this, network analysis revealed enrichment of genes related to the cellular response to oxidative stress, a commonly implicated process in age-related neurodegenerative diseases [119,120]. Thus, a plausible hypothesis is progressive increase in *TAF1* dysfunction in mutation carriers, accompanied by increasing oxidative damage with longer disease duration. These temporal derangements may be related to the adult onset of the disease, to the transition of the phenotype from dystonia to degenerative parkinsonism, and to the progressive imaging and brain findings.

Taken together, the expression studies performed here on peripheral tissue, using both high-throughput and gene-level methods, and analyzed both quantitatively and *in silico* using network and enrichment analyses, provide robust evidence that *TAF1* transcripts are dysregulated in XDP.

4.3 *nTAF1* transcripts are reduced in neuronal models of XDP

TAF1 is characterized by the existence of many polymorphic exons and extensive alternative splicing [27]. Because variant transcripts created by alternative splicing events exert different effects on molecular processes and may localize to different cell populations [121], the contribution of different common and alternative isoforms of *TAF1* to XDP disease causation are important to elucidate. Furthermore, while peripheral biomaterials could be useful surrogates to study disease mechanisms, experiments performed in relevant tissue and in validated models (i.e. iPS cells and iPS cell-derived neurons) elucidate cell type-specific changes and processes [122].

nTAF1, a neuronal isoform of *TAF1* that is preferentially found in neuronal cells has been shown to be reduced in the caudate of one individual with XDP. It is thus implicated as the primary isoform of *TAF1* that is responsible for the molecular pathogenesis of XDP [4,33,46]. Importantly, the study which detected this alteration additionally reported that overall *TAF1* expression was also slightly reduced in the XDP brain, but that extensive astrogliosis secondary to cell death in a postmortem specimen may have accounted for a secondarily increased expression of *TAF1* – negating the underexpression secondary to the XDP genotype [33]. In the in-vitro fibroblast model used above (in microarray-based profiling and qPCR experiments), where gliosis does not take place, the overall expression of *TAF1* was downregulated in comparison to controls. This was also the finding in another study which used

fibroblasts to identify dysregulated transcripts in XDP. Interestingly, although many different *TAF1* transcripts were found to be dysregulated, with some being overexpressed and some underexpressed in XDP compared to controls, transcripts derived from exons 32 through 36, i.e. the region of i) the SVA retrotransposon insertion (intron 32), and ii) the additional 6-bp that comprises exon 34' (*nTAF1*), were always significantly reduced in XDP fibroblasts [45].

nTAF1 is the second most abundant *TAF1* isoform in the brain, and its expression is also specific to neuronal tissue [33,44]. This was confirmed in the qPCR experiments performed here, where sufficient amounts of *nTAF1* were not detected in fibroblast and blood-derived RNA, but were quantifiable in RNA extracted from iPS cell-derived neurons. Using qPCR on human brain samples, one study quantified the abundance of *nTAF1* transcripts to be 21% of the major form of *TAF1* in the caudate, and 12.5% in the cortex [33]. The combination of molecular cloning and qPCR performed here identifies *nTAF1* expression in iPS cell-derived neurons to be 10-35% of total *TAF1* expression. Furthermore, in immunohistochemistry studies in rat brain, nTAF1+ cells localized preferentially to the striosome component of the basal ganglia [46]. There was also a higher expression of *nTAF1* in iPS cell-derived neurons differentiated to a striatal fate in the experiments performed here.

Overall, the isoform-specific quantification analyses performed do not only confirm the suitability of iPS cell-derived neurons as a model that approximates the in-vivo situation, they also imply that *nTAF1* is sufficiently abundant to exert a biological effect, i.e. the proportion of *nTAF1* in total *TAF1* transcripts is to a level that can be said to be relevant. Furthermore, the assay reveals reduced expression of both *nTAF1* and *cTAF1* transcripts in iPS cell-derived striatal neurons, consistent with the reductions seen in total *TAF1* expression in peripheral tissue (blood and fibroblasts). Even the magnitude of reduction is similar, i.e. between fold-change 1.15 to 1.30, in the independently performed experiments in different tissues. Importantly, *nTAF1* was a well-expressed isoform not just in striatal neurons but also in cortical and dopaminergic neurons, reflecting recent neuroimaging findings that fiber tracts outside the striatum are also impacted in XDP [17]. It would be interesting to determine in future experiments whether *nTAF1* and *cTAF1* reduction also occurs in other cell populations, and to hypothesize which group (of neurons or other cell types) suffers the primary insult and which are only secondarily affected by degeneration.

4.4 Insights from genome editing

An SVA retrotransposon insertion in intron 32 of *TAF1* is among the genetic alterations implicated in XDP. SVAs are mobile genetic elements that could insert in the genome, usually in introns; despite this, they remain active and have been shown to contribute to disease causation, via exonization or by way of altering the expression of the genes in which they insert [48,49,123]. Introns usually harbor cis-regulatory enhancer or repressor elements that modify the expression of nearby or even distant genes [124], [125]. Recent precision genome editing technology via CRISPR/Cas systems have confirmed that noncoding/intronic variants could indeed harbor sites of transcription factor binding, and that excision of these loci could reverse local gene expression changes [126,127].

The hypothesis here was thus, that the SVA retrotransposon insertion in intron 32 contains a cis-regulatory repressor of *TAF1* expression (thus the reduced expression of *TAF1* in XDP patients), and that excising the genetic element from XDP cells would result in a 'rescue' of the cellular phenotype, i.e. increased *TAF1* expression in CRISPR-modified XDP lines. Notably, although any among the XDP-specific alterations could theoretically be the cause of the observed reduction *TAF1* expression, database search and in-silico prediction does not show evidence of regulatory elements being harbored by any of the SNVs in the XDP haplotype, including the DSCs [36]. Instead, the SVA retrotransposon insertion is postulated to exert an effect on gene expression due its high GC content and hypermethylated CpG sites, characteristics that influence enhancer/repressor function on an epigenetic level [33,128]. Lastly, previous studies indicate that *TAF1* transcripts that are dysregulated in XDP cellular models are derived from exons that surround the locus of the SVA retrotransposon insertion [45].

CRISPR/Cas9-mediated genome editing was performed and the SVA retrotransposon insertion was successfully excised in XDP iPS cell lines. Notably, in addition to the 2,627 bps comprising the insertion, a total of 165 bp (or 166 bp in some colonies) were additionally excised, reflecting imperfect double-stranded repair mediated by non-homologous end-joining of a break in double-stranded DNA. In the control line, only this 165/166-bp 'editing scar' was seen after CRISPR modification. Current technologies are being developed for CRISPR-mediated 'scarless editing' [129]. Nevertheless, the experiment here is proof-of-principle that excision of the locus of the SVA retrotransposon insertion results in changes in *TAF1* expression.

Increase in *TAF1* expression was seen in the three XDP lines where the SVA insertion was successfully excised, using different comparators: i) against colonies transfected with the complete CRISPR/Cas9 system but where the SVA insertion was not

successfully cut, ii) against the parent lines, i.e. untransfected colonies, iii) against a control line (and here, also three subcomparators: parent, transfected and successfully edited, transfected and unedited lines and colonies). Interestingly, an increase in *TAF1* expression was also seen between successfully edited and unedited control colonies. However, after normalization, it was evident that the increase in *TAF1* expression was only significant in lines where the SVA retrotransposon insertion was excised, i.e. XDP lines. The setup emphasizes the importance of normalizing using multiple levels of isogenic, parent, and disease controls, and the use of different parent lines in modeling using iPS cells [130].

Thus, genome editing and subsequent analyses of resulting changes in *TAF1* expression suggest that the SVA retrotransposon insertion is causal to this molecular phenotype in XDP. The experiments further imply that common/overall *TAF1* transcripts participate in the mechanism of the disease. It cannot be ruled out that the neuron-specific *nTAF1* transcript plays a more central role, especially given the neurologic phenotype observed in humans; this will be determined by differentiation of the iPS cells used here into neuronal lines and further investigation of CRISPR/Cas9-mediated genome editing-induced changes in *TAF1* and *nTAF1* expression.

5 Conclusions

This dissertation clarifies several important aspects about the molecular genetic mechanisms surrounding XDP:

i) The disease-specific genetic variants associated with the disease co-segregate in complete linkage disequilibrium, in one 'XDP haplotype'. This haplotype subtends a narrowed region of 294.7 kb and *TAF1* is located in this haplotype, with the genetic variants found either in introns of the gene or in the extragenic region immediately distal to it.

ii) The expression of *TAF1* is dysregulated in peripheral models of XDP (blood, fibroblasts). *TAF1* is the only gene within the haplotype displaying changed expression in XDP patients and the reduction of *TAF1* expression in a fibroblast model led to consequent changes in global gene expression, network analysis of which reveals enrichment of transcriptional processes. In blood, increasing *TAF1* underexpression may follow disease duration.

iii) The expression of *nTAF1* in iPS cell-derived neurons parallels that in studies performed in human and rat brain, indicating the validity of neurons differentiated from iPS cell lines as a model for XDP. Reduced *TAF1* expression was evident in iPS cell lines obtained from XDP patients and in differentiated striatal neurons.

iv) Excision of the SVA retrotransposon insertion in intron 32 of *TAF1*, one of the genetic alterations within the XDP haplotype, resulted in increased *TAF1* expression to a level approximating controls. This suggests that this specific DNA variant is causal to the reduced *TAF1* expression observed in XDP.

Overall, the experiments propose that alterations in *TAF1* and dysregulation of its expression represent a consistent molecular phenotype in XDP. The studies here lay important groundwork for further investigation of the consequences of *TAF1* dysfunction and its relevance to the pathogenesis of XDP.

Summary

X-linked dystonia-parkinsonism (XDP) is a monogenic, neurodegenerative disease manifesting as combined dystonia and parkinsonism. This unique disorder has only been described in individuals of Filipino descent, suggesting that its etiology is derived from a founder haplotype; however, the condition represents an exemplar to gain insights into disease mechanisms in other forms of dystonia, parkinsonism, and other disorders of the basal ganglia. Despite previous efforts using conventional genetic approaches, the causal mutation in XDP remains unclear. Seven different alterations that lie on the X-chromosome are found only in patients and are said to be disease-specific. These are found in introns of, or in the extragenic region proximate to the *TAF1* gene. Given that they all occur in noncoding regions, their contribution to the causation of XDP is yet undetermined.

The first objective of this work was to identify which among the previously described genetic alterations represents a truly disease-specific DNA change. To this end, sequencing and haplotype analyses in more than 600 affected individuals and controls were performed. In parallel, whole-genome sequencing and copy number analysis were undertaken to identify a novel, possibly protein-coding variant undetected by previous methods. An XDP haplotype, composed of all previously described genetic variants occurring in complete linkage equilibrium, was found to be disease-specific. However, no novel and unambiguous protein-coding sequence change nor copy number variation was observed, and none among the known DNA changes was ruled out as unspecific. The definitive mutation in XDP is thus still not elucidated, its discovery hindered by genetic homogeneity in patients and the absence of partial haplotypes in controls. Single nucleotide variants discovered via genome sequencing however narrows the region where this mutation could be found, i.e. in a 294.7-kb DNA segment on Xq13.1, where the *TAF1* gene is located.[1]

Because the linked genetic variants converge in the region of the *TAF1* gene, a known transcription activator, expression studies were undertaken next, in the second part of the dissertation. The hypothesis was reduced expression of *TAF1* in peripheral models of XDP, given the ubiquitous function and expression of this gene. *TAF1* was underexpressed in patients vs. controls, in blood and in fibroblast-derived RNA. Gene expression profiling also identified a reduced signal from the gene. Although the magnitude of *TAF1* underexpression was modest, this was accompanied by the dysregulation of 307 genes genome-wide in fibroblasts, implying a downstream effect that is consistent with the disruption of the function of a transcriptional activator.

[1] The work included in this objective is published in Domingo et al., *Eur. J. Hum. Genet.*, 2015.

Network analysis of differentially expressed genes revealed enrichment of genes involved in transcriptional processes. Thus, the expression studies in peripheral tissue suggest that TAF1 underexpression and transcriptional dysregulation are central to the pathogenesis of XDP.[2]

TAF1 is extensively alternatively spliced, and a neuron-specific isoform, *nTAF1,* was previously shown to be underexpressed in the brain of an XDP patient. Brain tissue is however scarce and often inaccessible, and one alternative is the use of endogenous cellular models, i.e. induced pluripotent stem (iPS) cells and differentiated neurons, to study molecular alterations in the appropriate and relevant tissue context. *TAF1* and *nTAF1* expression analyses were performed in iPS cells and iPS cell-derived neurons in the third part of this thesis, to verify the abundance of *nTAF1* in these models and also to confirm the underexpression observed in XDP brain tissue. Consistent with the expression studies performed in blood and fibroblasts, reduced *TAF1* expression was seen in iPS cells and in striatal neurons, where *nTAF1* was also found to be abundantly expressed.

One genetic alteration within the XDP haplotype, a large, 2,627-kb SVA retrotransposon insertion in intron 32 of *TAF1*, is postulated to exert an effect on gene expression. To verify its contribution to the hitherto observed underexpression of *TAF1*, targeted genome editing was performed in the fourth part of this dissertation, followed by expression analysis. The hypothesis was that experimental removal of this genetic alteration using CRISPR/Cas9-mediated genome editing in iPS cells would result in increased *TAF1* expression. The insertion was successfully excised, and *TAF1* expression was indeed 'rescued' in edited cells. This effect was observed in three different cell lines from patients, and was not seen in one control line. This work therefore provides initial molecular evidence suggesting that the SVA retrotransposon insertion is causal to *TAF1* underexpression in XDP, and that when this putative repressor is removed, the expression of the gene can be restored.

The results of the genetic, transcriptomic, bioinformatic, cell modeling, and genome editing experiments here converge on *TAF1* alterations as the consistent molecular phenotype of XDP. Multiple lines of evidence show that the variants in the gene as well as its reduced expression in various tissues and models are disease-specific. Lastly, the experiments suggest that an SVA retrotransposon insertion is causal to the reduced expression of *TAF1* in this disease. The different experiments and resulting conclusions here advance current knowledge on the molecular genetic mechanisms surrounding this orphan disease. The lack of a definitive causal mutation and tractable

[2] The work included in this objective is published in Domingo et al., *Cell. Mol. Life Sci.*, 2016.

molecular phenotype in XDP has been a frustrating barrier to rational therapeutics. This work provides initial progress in answering questions related to the pathogenesis of XDP, and in laying the groundwork necessary for an ultimate solution to a long-standing problem in neurogenetic research.

Zusammenfassung

Das X-chromosomale Dystonie-Parkinson-Syndrom (*X-linked dystonia-parkinsonism*, XDP) ist eine monogene, neurodegenerative Erkrankung, die sich als Dystonie in Kombination mit Parkinsonismus manifestiert. Diese besondere Form der Dystonie betrifft wegen eines genetischen Gründereffekts wahrscheinlich nur Menschen philippinischer Abstammung. Obwohl XDP sehr selten auftritt, stellt es dennoch eine Modellerkrankung dar, um die Krankheitsmechanismen von Dystonie und Parkinsonismus im Speziellen und Erkrankungen der Basalganglien im Allgemeinen zu untersuchen und damit besser zu verstehen. Trotz verschiedener konventioneller genetischer Studien ist die kausale Mutation für XDP noch nicht genau bekannt. Bisher wurden sieben verschiedene, krankheitsspezifische, X-chromosomale Varianten (*disease-specific changes*, DSCs) gefunden, die alle nur bei Patienten vorkommen. Diese DSCs liegen in intronischen oder extragenischen Regionen des Genoms im Bereich des *TAF1*-Gens. Da sie in nicht kodierenden Regionen des Genoms liegen, ist ihre kausale Rolle bei der Krankheitsentstehung umstritten.

Das erste Ziel dieser Doktorarbeit war es herauszufinden, welcher DSC bzw. welche andere Sequenzveränderung ursächlich für XDP ist. Dazu wurde das Auftreten dieser DSCs bei einer großen Gruppe von 600 Patienten und Kontrollen untersucht, um Rekombinationsereignisse aufzudecken, d. h. den krankheits-assoziierten Haplotyp einzugrenzen, indem Patienten identifiziert werden sollten, die nicht alle sieben DSCs tragen bzw. Kontrollen zu finden, die mindestens einen der DSCs aufweisen. Parallel wurde mittels Genom-Sequenzierung und Kopienzahlbestimmung (*copy number*-Analyse) nach einer bisher unbekannten, möglicherweise Protein-verändernden Mutation gesucht. Diese konnte allerdings nicht gefunden werden, stattdessen wurde ein Haplotyp, bestehend aus allen sieben DSCs identifiziert, die sich im Kopplungsgleichgewicht (*linkage equilibrium*) befinden. Die entscheidende XDP-Mutation ist damit jedoch noch immer nicht genau geklärt, weil alle Patienten eine genetische Homogenität aufweisen und kein Teil des Haplotyps bei Kontrollen gefunden werden konnte. Durch die Entdeckung einzelner Polymorphismen mittels Genom-Sequenzierung konnte allerdings die gekoppelte Region (XDP-Haplotyp) auf eine 294.7 kb große Region auf Chromosom Xq13.1 eingegrenzt werden, in der auch das *TAF1*-Gen liegt.[1]

Da die genetischen Varianten alle im Bereich des *TAF1*-Gens, welches für einen Transkriptionsfaktor kodiert, aufeinandertreffen, wurden im zweiten Teil dieser Arbeit zunächst Expressions-Analysen von *TAF1* durchgeführt. Die Hypothese dabei war, dass

[1] Dieser Teil der Arbeit wurde bereits in Domingo et al., *Eur. J. Hum. Genet.*, 2015 veröffentlicht.

bei Patienten eine verringerte Expression des *TAF1*-Gens auch in die Peripherie zu finden ist, da die Funktion von *TAF1* ubiquitär ist. Mittels quantitativer PCR konnte gezeigt werden, dass die Expression von *TAF1* auf mRNA-Ebene bei Patienten im Vergleich zu Kontrollen sowohl im Blut als auch in Fibroblasten tatsächlich herunterreguliert ist. Die Analyse des genomweiten Expressionsprofils (*gene expression profiling*) bestätigte die verringerte Expression von *TAF1*. Obwohl der Effekt auf *TAF1* relativ klein war, konnte eine veränderte Expression von 307 Genen in Fibroblastenkulturen von Patienten gezeigt werden, was eine Dysfunktion eines Transkriptionsfaktors impliziert. Eine Netzwerk-Analyse dieser Befunde belegte eine Anreichung von Genen, die an der Transkription beteiligt sind. Die Expressions-Studien in peripheren Geweben legen nahe, dass die verringerte Expression von *TAF1* und eine Fehlregulation der Transkription eine zentrale Rolle in der Pathogenese von XDP spielen.[2]

Das *TAF1*-Gen hat verschiedene Isoformen. Besonders interessant ist dabei eine Neuronen-spezifische Form (*nTAF1*), für die eine reduzierte Expression im Gehirn eines XDP-Patienten gezeigt worden war. Neuronales Gewebe steht bei lebenden Probanden nicht zur Verfügung. Die Nutzung von endogene Zellmodellen, wie induzierte pluripotente Stammzellen (iPS) und daraus differenzierten Neuronen stellt eine geeignete Alternative dar, um Krankheitsmechanismen im relevantem Gewebe zu untersuchen und wurde im dritten Teil der vorliegenden Arbeit angewendet. Die Expression von *TAF1* und vor allem auch von *nTAF1* wurde in iPS und in aus iPS-Zellen differenzierten Neuronen analysiert. Dabei wurde gezeigt, dass die *TAF1*-Expression in iPS-Zellen und daraus differenzierten striatalen Neuronen reduziert und *nTAF1* relativ stark exprimiert war.

Eine der krankheitsspezifischen Varianten, eine 2,6 Mb große SVA-Retrotransposon-Insertion die im Intron 32 des *TAF1*-Gens liegt, könnte für die Expressionsveränderungen verantwortlich sein. Um dies zu bestätigen, wurde im vierten Teil dieser Arbeit ein maßgeschneidertes *Genome-Editing*-Verfahren mit anschließender Expressions-Analyse durchgeführt. Die Annahme war, dass die experimentelle Entfernung der Insertion in iPS-Zellen durch die CRISPR/Cas9-Methode zu einer Erhöhung der *TAF1*-Expression führen würde. Diese Annahme konnte bei drei verschiedenen XDP-Linien experimentell bestätigt werden. Dies legt nahe, dass die SVA-Retrotransposon-Insertion kausal für die reduzierte Expression von *TAF1* ist und dass eine Entfernung der Insertion die normale TAF1-Expression wiederherstellen könnte.

[2] Dieser Teil der Arbeit wurde bereits in Domingo et al., *Cell. Mol. Life Sci.*, 2015 veröffentlicht.

Die vielschichtigen Ergebnisse durch Experimente auf unterschiedlichen Ebenen wie Gen-, Chromosom-, Genom- und Transkriptom-Untersuchungen in unterschiedlichen Zellmodellen mittels genetischer, bioinformatischer und statistischer Methoden konvergieren und beweisen, dass die Veränderung von *TAF1* mit XDP im Zusammenhang steht. Diese Arbeit zeigt, dass mehrere Varianten im *TAF1*-Gen und die damit verbundene reduzierte Expression in verschiedenen Geweben und Modellen krankheitsspezifisch sind und möglicherweise die SVA-Insertion kausal für die reduzierte Expression ist. Die verschiedenen Experimente und davon abgeleiteten Schlussfolgerungen haben zu einem besseren Verständnis der molekulargenetischen Mechanismen dieser seltenen Erkrankung geführt. Das fehlende Wissen über die kausale Mutation und den damit verbundenen molekularen Phänotyp von XDP stellte bisher in der Entwicklung einer gezielten Therapie ein unüberwindbares Hindernis dar. Im Rahmen dieser Arbeit sind wichtige Grundlagenerkenntnisse erzielt worden, die zur Lösung vieler Fragen in Bezug auf XDP beitragen und neue therapeutische Ansatzpunkte bieten.

References

[1] A. Domingo, R. Erro, and K. Lohmann, "Novel Dystonia Genes: Clues on Disease Mechanisms and the Complexities of High-Throughput Sequencing," *Mov. Disord.*, vol. 31, no. 4, pp. 471–477, 2016.

[2] L. V. Lee, C. Rivera, R. Teleg, M. Dantes, P. M. Pasco, R. D. G. Jamora, J. Arancillo, R. Villareal-Jordan, R. L. Rosales, C. Demaisip, E. Maranon, O. Peralta, R. Borres, C. Tolentino, M. Monding, and S. Sarcia, "The unique phenomenology of X-linked dystonia-parkinsonism (XDP, DYT3, 'Lubag')," *Int. J. Neurosci.*, vol. 121 Suppl, pp. 3–11, 2011.

[3] V. G. H. Evidente, J. Advincula, R. Esteban, P. Pasco, J. A. Alfon, F. F. Natividad, J. Cuanang, A. San Luis, K. Gwinn-Hardy, J. Hardy, D. Hernandez, and A. Singleton, "Phenomenology of 'Lubag' or X-linked dystonia-parkinsonism," *Mov. Disord.*, vol. 17, no. 6, pp. 1271–1277, 2002.

[4] R. Kaji, S. Goto, G. Tamiya, S. Ando, S. Makino, and L. Lee, "Molecular dissection and anatomical basis of dystonia: X-linked recessive dystonia-parkinsonism (DYT3)," *J. Med. Investig.*, vol. 52, no. Supplemental, pp. 280–283, 2005.

[5] R. L. Rosales, "X-linked dystonia parkinsonism: clinical phenotype, genetics and therapeutics.," *J. Mov. Disord.*, vol. 3, no. 2, pp. 32–8, 2010.

[6] L. V Lee, F. M. Pascasio, F. D. Fuentes, and G. H. Viterbo, "Torsion dystonia in Panay, Philippines.," *Adv. Neurol.*, vol. 14, pp. 137–151, 1976.

[7] K. G. Kupke, L. V Lee, G. H. Viterbo, J. Arancillo, T. Donlon, and U. Muller, "X-linked recessive torsion dystonia in the Philippines.," *Am. J. Med. Genet.*, vol. 36, no. 2, pp. 237–242, Jun. 1990.

[8] L. V Lee, K. G. Kupke, F. Caballar-Gonzaga, M. Hebron-Ortiz, and U. Müller, "The phenotype of the X-linked dystonia-parkinsonism syndrome. An assessment of 42 cases in the Philippines.," *Medicine*, vol. 70, no. 3. pp. 179–187, 1991.

[9] L. V Lee, E. Maranon, C. Demaisip, O. Peralta, R. Borres-Icasiano, J. Arancillo, C. Rivera, E. Munoz, K. Tan, and M. T. Reyes, "The natural history of sex-linked recessive dystonia parkinsonism of Panay, Philippines (XDP).," *Parkinsonism Relat. Disord.*, vol. 9, no. 1, pp. 29–38, Oct. 2002.

[10] H. Deng, W.-D. Le, and J. Jankovic, "Genetic study of an American family with DYT3 dystonia (lubag).," *Neurosci. Lett.*, vol. 448, no. 2, pp. 180–183, Dec. 2008.

[11] C. Plummer, J. Bradfield, A. B. Singleton, D. Hernandez, A. A. Singleton, and J. O'sullivan, "First case report of X linked dystonia parkinsonism (XDP) or 'lubag' in Australia.," *J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas.*, vol. 12, no. 8, pp. 945–946, Nov. 2005.

[12] C. H. Waters, H. Takahashi, K. C. Wilhelmsen, R. Shubin, B. J. Snow, T. G. Nygaard, C. B. Moskowitz, S. Fahn, and D. B. Calne, "Phenotypic expression of X-linked dystonia-parkinsonism (lubag) in two women.," *Neurology*, vol. 43, no. 8, pp. 1555–1558, Aug. 1993.

[13] V. G. H. Evidente, D. Nolte, S. Niemann, J. Advincula, M. C. Mayo, F. F. Natividad, and U. Muller, "Phenotypic and molecular analyses of X-linked dystonia-parkinsonism ('lubag') in women.," *Arch. Neurol.*, vol. 61, no. 12, pp. 1956–1959, Dec. 2004.

[14] A. Westenberger, R. L. Rosales, S. Heinitz, K. Freimann, L. V Lee, R. D. Jamora, A. R. Ng, A. Domingo, K. Lohmann, U. Walter, U. Golnitz, A. Rolfs, I. Nagel, G. Gillessen-Kaesbach, R. Siebert, D. Dressler, and C. Klein, "X-linked Dystonia-Parkinsonism manifesting in a female patient due to atypical turner syndrome.," *Mov. Disord.*, vol. 28, no. 5, pp. 675–678, May 2013.

[15] A. Domingo, L. V. Lee, N. Brüggemann, K. Freimann, F. J. Kaiser, R. D. G.

Jamora, R. L. Rosales, C. Klein, and A. Westenberger, “Woman With X-Linked Recessive Dystonia-Parkinsonism,” *JAMA Neurol.*, vol. 71, no. 9, p. 1177, 2014.

[16] P. M. D. Pasco, C. V Ison, E. L. Munoz, N. S. Magpusao, A. E. Cheng, K. T. Tan, R. W. Lo, R. A. Teleg, M. B. Dantes, R. Borres, E. Maranon, C. Demaisip, M. V. T. Reyes, and L. V Lee, “Understanding XDP through imaging, pathology, and genetics.,” *Int. J. Neurosci.*, vol. 121 Suppl, pp. 12–17, 2011.

[17] N. Brüggemann, M. Heldmann, C. Klein, A. Domingo, D. Rasche, V. Tronnier, R. L. Rosales, R. D. G. Jamora, L. V Lee, and T. F. Münte, “Neuroanatomical changes extend beyond striatal atrophy in X-linked dystonia parkinsonism,” *Parkinsonism Relat. Disord.*

[18] S. Goto, L. V. Lee, E. L. Munoz, I. Tooyama, G. Tamiya, S. Makino, S. Ando, M. B. Dantes, K. Yamada, S. Matsumoto, H. Shimazu, J. I. Kuratsu, A. Hirano, and R. Kaji, “Functional anatomy of the basal ganglia in X-linked recessive dystonia-parkinsonism,” *Ann. Neurol.*, vol. 58, no. 1, pp. 7–17, 2005.

[19] C. H. Waters, P. L. Faust, J. Powers, H. Vinters, C. Moskowitz, T. Nygaard, A. L. Hunt, and S. Fahn, “Neuropathology of lubag (x-linked dystonia parkinsonism).,” *Mov. Disord.*, vol. 8, no. 3, pp. 387–390, Jul. 1993.

[20] K. G. Kupke, M. B. Graeber, and U. Müller, “Dystonia-parkinsonism syndrome (XDP) locus: flanking markers in Xq12-q21.1.,” *Am. J. Hum. Genet.*, vol. 50, no. 4, pp. 808–815, 1992.

[21] M. B. Graeber and U. Muller, “The X-linked dystonia-parkinsonism syndrome (XDP): clinical and molecular genetic analysis.,” *Brain Pathol.*, vol. 2, no. 4, pp. 287–295, Oct. 1992.

[22] G. Haberhausen, I. Schmitt, A. Köhler, U. Peters, S. Rider, J. Chelly, J. D. Terwilliger, A. P. Monaco, and U. Müller, “Assignment of the dystonia-parkinsonism syndrome locus, DYT3, to a small region within a 1.8-Mb YAC contig of Xq13.1.,” *Am. J. Hum. Genet.*, vol. 57, no. 3, pp. 644–50, 1995.

[23] a H. Németh, D. Nolte, E. Dunne, S. Niemann, M. Kostrzewa, U. Peters, E. Fraser, E. Bochukova, R. Butler, J. Brown, R. D. Cox, E. R. Levy, H. H. Ropers, a P. Monaco, and U. Müller, “Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3).,” *Genomics*, vol. 60, no. 3, pp. 320–329, 1999.

[24] U. Peters, G. Haberhausen, M. Kostrzewa, D. Nolte, and U. Müller, “AFX1 and p54(nrb): Fine mapping, genomic structure, and exclusion as candidate genes of X-linked dystonia parkinsonism,” *Hum. Genet.*, vol. 100, no. 5–6, pp. 569–572, 1997.

[25] D. Nolte, J. Ramser, S. Niemann, H. Lehrach, R. Sudbrak, and U. M??ller, “ACRC codes for a novel nuclear protein with unusual acidic repeat tract and maps to DYT3 (dystonia parkinsonism) critical interval in Xq13.1,” *Neurogenetics*, vol. 3, no. 4, pp. 207–213, 2001.

[26] D. Nolte, S. Niemann, and U. Müller, “Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 18, pp. 10347–10352, 2003.

[27] T. Herzfeld, D. Nolte, and U. Müller, “Structural and functional analysis of the human TAF1/DYT3 multiple transcript system,” *Mamm. Genome*, vol. 18, no. 11, pp. 787–795, 2007.

[28] T. Herzfeld, D. Nolte, M. Grznarova, A. Hofmann, J. L. Schultze, and U. Müller, “X-linked dystonia parkinsonism syndrome (XDP, lubag): Disease-specific sequence change DSC3 in TAF1/DYT3 affects genes in vesicular transport and dopamine metabolism,” *Hum. Mol. Genet.*, vol. 22, no. 5, pp. 941–951, 2013.

[29] A. M. A. Schreij, E. A. Fon, and P. S. McPherson, “Endocytic membrane trafficking and neurodegenerative disease.,” *Cell. Mol. Life Sci.*, Dec. 2015.

[30] R. M. Perrett, Z. Alexopoulou, and G. K. Tofaris, “The endosomal pathway in Parkinson’s disease.,” *Molecular and cellular neurosciences*, vol. 66, no. Pt A. pp.

21–28, May-2015.
[31] G. Sciamanna, A. Tassone, G. Martella, G. Mandolesi, F. Puglisi, D. Cuomo, G. Madeo, G. Ponterio, D. G. Standaert, P. Bonsi, and A. Pisani, “Developmental profile of the aberrant dopamine D2 receptor response in striatal cholinergic interneurons in DYT1 dystonia.,” *PLoS One*, vol. 6, no. 9, p. e24261, 2011.
[32] R. E. Goodchild, K. Grundmann, and A. Pisani, “New genetic insights highlight ‘old’ ideas on motor dysfunction in dystonia.,” *Trends Neurosci.*, vol. 36, no. 12, pp. 717–725, Dec. 2013.
[33] S. Makino, R. Kaji, S. Ando, M. Tomizawa, K. Yasuno, S. Goto, S. Matsumoto, M. D. Tabuena, E. Maranon, M. Dantes, L. V Lee, K. Ogasawara, I. Tooyama, H. Akatsu, M. Nishimura, and G. Tamiya, “Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism.,” *Am. J. Hum. Genet.*, vol. 80, no. 3, pp. 393–406, 2007.
[34] T. Kawarai, P. M. D. Pasco, R. A. Teleg, M. Kamada, W. Sakai, K. Shimozono, M. Mizuguchi, D. Tabuena, A. Orlacchio, Y. Izumi, S. Goto, L. V. Lee, and R. Kaji, “Application of long-range polymerase chain reaction in the diagnosis of X-linked dystonia-parkinsonism,” *Neurogenetics*, vol. 14, no. 2, pp. 167–169, 2013.
[35] U. Müller, T. Herzfeld, D. Nolte, G. Tamiya, S. Makino, and R. Kaji, “Letters to the editor,” *Am. J. Hum. Genet.*, vol. 81, no. August, pp. 414–421, 2007.
[36] A. Domingo, A. Westenberger, L. V Lee, I. Brænne, T. Liu, I. Vater, R. Rosales, R. D. Jamora, P. M. Pasco, E. M. Cutiongco-dela Paz, K. Freimann, T. G. Schmidt, D. Dressler, F. J. Kaiser, L. Bertram, J. Erdmann, K. Lohmann, and C. Klein, “New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3),” *Eur. J. Hum. Genet.*, vol. 23, no. 10, pp. 1334–1340, 2015.
[37] S. Ruppert, E. H. Wang, and R. Tjian, “Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation.,” *Nature*, vol. 362, no. 6416, pp. 175–179, Mar. 1993.
[38] M. C. Thomas and C.-M. Chiang, “The general transcription machinery and general cofactors.,” *Crit. Rev. Biochem. Mol. Biol.*, vol. 41, no. 3, pp. 105–178, 2006.
[39] T. H. Kim, L. O. Barrera, M. Zheng, C. Qu, M. a Singer, T. a Richmond, Y. Wu, R. D. Green, and B. Ren, “A high-resolution map of active promoters in the human genome.,” *Nature*, vol. 436, no. 7052, pp. 876–880, 2005.
[40] T. Sekiguchi, Y. Nohiro, Y. Nakamura, N. Hisamoto, and T. Nishimoto, “The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein.,” *Mol. Cell. Biol.*, vol. 11, no. 6, pp. 3317–3325, Jun. 1991.
[41] E. H. Wang and R. Tjian, “Promoter-selective transcriptional defect in cell cycle mutant ts13 rescued by hTAFII250.,” *Science*, vol. 263, no. 5148, pp. 811–814, Feb. 1994.
[42] T. Hayashida, T. Sekiguchi, E. Noguchi, H. Sunamoto, T. Ohba, and T. Nishimoto, “The CCG1/TAFII250 gene is mutated in thermosensitive G1 mutants of the BHK21 cell line derived from golden hamster.,” *Gene*, vol. 141, no. 2, pp. 267–270, Apr. 1994.
[43] J. A. O’Rawe, Y. Wu, M. J. Dörfel, A. F. Rope, P. Y. B. Au, J. S. Parboosingh, S. Moon, M. Kousi, K. Kosma, C. S. Smith, M. Tzetis, J. L. Schuette, R. B. Hufnagel, C. E. Prada, F. Martinez, C. Orellana, J. Crain, A. Caro-Llopis, S. Oltra, S. Monfort, L. T. Jiménez-Barrón, J. Swensen, S. Ellingwood, R. Smith, H. Fang, S. Ospina, S. Stegmann, N. Den Hollander, D. Mittelman, G. Highnam, R. Robison, E. Yang, L. Faivre, A. Roubertie, J.-B. Rivière, K. G. Monaghan, K. Wang, E. E. Davis, N. Katsanis, V. M. Kalscheuer, E. H. Wang, K. Metcalfe, T. Kleefstra, A. M. Innes, S. Kitsiou-Tzeli, M. Rosello, C. E. Keegan, and G. J. Lyon, “TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability,

and Neurological Manifestations," *Am. J. Hum. Genet.*, vol. 97, no. 6, pp. 922–932, 2015.

[44] J. Jambaldorj, S. Makino, B. Munkhbat, and G. Tamiya, "Sustained expression of a neuron-specific isoform of the Taf1 gene in development stages and aging in mice," *Biochem. Biophys. Res. Commun.*, vol. 425, no. 2, pp. 273–277, 2012.

[45] N. Ito, W. T. Hendriks, J. Dhakal, C. A. Vaine, C. Liu, D. Shin, K. Shin, N. Wakabayashi-Ito, M. Dy, T. Multhaupt-Buell, N. Sharma, X. O. Breakefield, and D. C. Bragg, "Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells.," *Dis. Model. Mech.*, vol. 9, no. 4, pp. 451–62, 2016.

[46] W. Sako, R. Morigaki, R. Kaji, I. Tooyama, S. Okita, K. Kitazato, S. Nagahiro, A. M. Graybiel, and S. Goto, "Identification and localization of a neuron-specific isoform of TAF1 in rat brain: implications for neuropathology of DYT3 dystonia.," *Neuroscience*, vol. 189, pp. 100–107, Aug. 2011.

[47] A. W. Dunah, H. Jeong, A. Griffin, Y.-M. Kim, D. G. Standaert, S. M. Hersch, M. M. Mouradian, A. B. Young, N. Tanese, and D. Krainc, "Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease.," *Science*, vol. 296, no. 5576, pp. 2238–2243, Jun. 2002.

[48] M. Watanabe, K. Kobayashi, F. Jin, K. S. Park, T. Yamada, K. Tokunaga, and T. Toda, "Founder SVA retrotransposal insertion in Fukuyama-type congenital muscular dystrophy and its origin in Japanese and Northeast Asian populations.," *Am. J. Med. Genet. A*, vol. 138, no. 4, pp. 344–348, Nov. 2005.

[49] M. Taniguchi-Ikeda, K. Kobayashi, M. Kanagawa, C. Yu, K. Mori, T. Oda, A. Kuga, H. Kurahashi, H. O. Akman, S. DiMauro, R. Kaji, T. Yokota, S. Takeda, and T. Toda, "Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy," *Nature*, vol. 478, no. 7367, pp. 127–131, Oct. 2011.

[50] A. Rakovic, A. Grunewald, J. Kottwitz, N. Bruggemann, P. P. Pramstaller, K. Lohmann, and C. Klein, "Mutations in PINK1 and Parkin impair ubiquitination of Mitofusins in human fibroblasts.," *PLoS One*, vol. 6, no. 3, p. e16746, 2011.

[51] F. J. Kaiser, A. Osmanoric, A. Rakovic, A. Erogullari, N. Uflacker, D. Braunholz, T. Lohnau, S. Orolicki, M. Albrecht, G. Gillessen-Kaesbach, C. Klein, and K. Lohmann, "The dystonia gene DYT1 is repressed by the transcription factor THAP1 (DYT6).," *Ann. Neurol.*, vol. 68, no. 4, pp. 554–559, Oct. 2010.

[52] F. Borovecki, L. Lovrecic, J. Zhou, H. Jeong, F. Then, H. D. Rosas, S. M. Hersch, P. Hogarth, B. Bouzou, R. V Jensen, and D. Krainc, "Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 31, pp. 11023–11028, 2005.

[53] N. Zeltner and L. Studer, "Pluripotent stem cell-based disease modeling: current hurdles and future promise.," *Curr. Opin. Cell Biol.*, vol. 37, pp. 102–110, Dec. 2015.

[54] S. Camnasio, A. Delli Carri, A. Lombardo, I. Grad, C. Mariotti, A. Castucci, B. Rozell, P. Lo Riso, V. Castiglioni, C. Zuccato, C. Rochon, Y. Takashima, G. Diaferia, I. Biunno, C. Gellera, M. Jaconi, A. Smith, O. Hovatta, L. Naldini, S. Di Donato, A. Feki, and E. Cattaneo, "The first reported generation of several induced pluripotent stem cell lines from homozygous and heterozygous Huntington's disease patients demonstrates mutation related enhanced lysosomal activity.," *Neurobiol. Dis.*, vol. 46, no. 1, pp. 41–51, Apr. 2012.

[55] L. N. Munsie, A. J. Milnerwood, P. Seibler, D. A. Beccano-Kelly, I. Tatarnikov, J. Khinda, M. Volta, C. Kadgien, L. P. Cao, L. Tapia, C. Klein, and M. J. Farrer, "Retromer-dependent neurotransmitter receptor trafficking to synapses is altered by the Parkinson's disease VPS35 mutation p.D620N.," *Hum. Mol. Genet.*, vol. 24, no. 6, pp. 1691–1703, Mar. 2015.

[56] S. Kriks, J.-W. Shim, J. Piao, Y. M. Ganat, D. R. Wakeman, Z. Xie, L. Carrillo-

Reid, G. Auyeung, C. Antonacci, A. Buch, L. Yang, M. F. Beal, D. J. Surmeier, J. H. Kordower, V. Tabar, and L. Studer, “Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson’s disease,” *Nature*, vol. 480, no. 7378, pp. 547–551, 2011.

[57] P. Seibler, J. Graziotto, H. Jeong, F. Simunovic, C. Klein, and D. Krainc, “Mitochondrial Parkin recruitment is impaired in neurons derived from mutant PINK1 induced pluripotent stem cells.,” *J. Neurosci.*, vol. 31, no. 16, pp. 5970–5976, 2011.

[58] A. Rakovic, K. Shurkewitsch, P. Seibler, A. Grunewald, A. Zanon, J. Hagenah, D. Krainc, and C. Klein, “Phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (PINK1)-dependent ubiquitination of endogenous Parkin attenuates mitophagy: study in human primary fibroblasts and induced pluripotent stem cell-derived neurons.,” *J. Biol. Chem.*, vol. 288, no. 4, pp. 2223–2237, Jan. 2013.

[59] Y. Yao, X. Cui, I. Al-Ramahi, X. Sun, B. Li, J. Hou, M. Difiglia, J. Palacino, Z.-Y. Wu, L. Ma, J. Botas, and B. Lu, “A striatal-enriched intronic GPCR modulates huntingtin levels and toxicity.,” *Elife*, vol. 4, 2015.

[60] J. A. Doudna and E. Charpentier, “The new frontier of genome engineering with CRISPR-Cas9,” *Science (80-.).*, vol. 346, no. 6213, pp. 1258096–1258096, 2014.

[61] P. D. Hsu, E. S. Lander, and F. Zhang, “Development and applications of CRISPR-Cas9 for genome engineering,” *Cell*, vol. 157, no. 6, pp. 1262–1278, 2014.

[62] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang, “DNA targeting specificity of RNA-guided Cas9 nucleases.,” *Nat. Biotechnol.*, vol. 31, no. 9, pp. 827–32, 2013.

[63] L. Ye, J. Wang, A. I. Beyer, F. Teque, T. J. Cradick, Z. Qi, J. C. Chang, G. Bao, M. O. Muench, J. Yu, J. A. Levy, and Y. W. Kan, “Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 26, pp. 9591–9596, 2014.

[64] X. Huang, Y. Wang, W. Yan, C. Smith, Z. Ye, J. Wang, Y. Gao, L. Mendelsohn, and L. Cheng, “Production of gene-corrected adult beta globin protein in human erythrocytes differentiated from patient iPSCs after genome editing of the sickle point mutation,” *Stem Cells*, p. n/a-n/a, 2015.

[65] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.,” *Science*, vol. 337, no. 6096, pp. 816–821, Aug. 2012.

[66] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, “Multiplex Genome Engineering Using CRISPR/Cas System,” *Science (80-.).*, vol. 339, no. February, pp. 819–824, 2013.

[67] Y. Wu, D. Liang, Y. Wang, M. Bai, W. Tang, S. Bao, Z. Yan, D. Li, and J. Li, “Correction of a genetic disease in mouse via use of CRISPR-Cas9,” *Cell Stem Cell*, vol. 13, no. 6, pp. 659–662, 2013.

[68] J. Wang and S. R. Quake, “RNA-guided endonuclease provides a therapeutic strategy to cure latent herpesviridae infection.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 36, pp. 13157–62, 2014.

[69] G. Schwank, B. K. Koo, V. Sasselli, J. F. Dekkers, I. Heo, T. Demircan, N. Sasaki, S. Boymans, E. Cuppen, C. K. Van Der Ent, E. E. S. Nieuwenhuis, J. M. Beekman, and H. Clevers, “Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients,” *Cell Stem Cell*, vol. 13, no. 6, pp. 653–658, 2013.

[70] M. Schuelke, “An economic method for the fluorescent labeling of PCR fragments.,” *Nat. Biotechnol.*, vol. 18, no. February, pp. 233–234, 2000.
[71] W. S. Oetting, H. K. Lee, D. J. Flanders, G. L. Wiesner, T. A. Sellers, and R. A. King, “Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers.,” *Genomics*, vol. 30, no. 3, pp. 450–458, Dec. 1995.
[72] K. Wang, M. Li, and H. Hakonarson, “ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.,” *Nucleic Acids Res.*, vol. 38, no. 16, p. e164, Sep. 2010.
[73] K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S. F. A. Grant, H. Hakonarson, and M. Bucan, “PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data.,” *Genome Res.*, vol. 17, no. 11, pp. 1665–1674, Nov. 2007.
[74] S. a Bustin, V. Benes, J. a Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele, and C. T. Wittwer, “The MIQE guidelines:Minimum Information for publication of quantitative real-time PCR experiments,” *Clin. Chem.*, vol. 55, no. 4, pp. 611–622, 2009.
[75] K. J. Livak and T. D. Schmittgen, “Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.,” *Methods*, vol. 25, no. 4, pp. 402–408, 2001.
[76] M. Marullo, C. Zuccato, C. Mariotti, N. Lahiri, S. J. Tabrizi, S. Di Donato, and E. Cattaneo, “Expressed Alu repeats as a novel, reliable tool for normalization of real-time quantitative RT-PCR data.,” *Genome Biol.*, vol. 11, no. 1, p. R9, 2010.
[77] A. Rihani, T. Van Maerken, F. Pattyn, G. Van Peer, A. Beckers, S. De Brouwer, C. Kumps, E. Mets, J. Van der Meulen, P. Rondou, C. Leonelli, P. Mestdagh, F. Speleman, and J. Vandesompele, “Effective Alu Repeat Based RT-Qpcr Normalization in Cancer Cell Perturbation Experiments,” *PLoS One*, vol. 8, no. 8, 2013.
[78] T. Hruz, M. Wyss, M. Docquier, M. W. Pfaffl, S. Masanetz, L. Borghi, P. Verbrugghe, L. Kalaydjieva, S. Bleuler, O. Laule, P. Descombes, W. Gruissem, and P. Zimmermann, “RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization.,” *BMC Genomics*, vol. 12, no. 1, p. 156, 2011.
[79] B. S. Stamova, M. Apperson, W. L. Walker, Y. Tian, H. Xu, P. Adamczy, X. Zhan, D. Liu, B. P. Ander, I. H. Liao, J. P. Gregg, R. J. Turner, G. Jickling, L. Lit, and F. R. Sharp, “Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood.,” *BMC Med. Genomics*, vol. 2, p. 49, 2009.
[80] J. T. Leek, R. B. Scharpf, H. C. Bravo, D. Simcha, B. Langmead, W. E. Johnson, D. Geman, K. Baggerly, and R. A. Irizarry, “Tackling the widespread and critical impact of batch effects in high-throughput data,” *Nat Rev Genet*, vol. 11, no. 10, pp. 733–739, Oct. 2010.
[81] V. G. Tusher, R. Tibshirani, and G. Chu, “Significance analysis of microarrays applied to the ionizing radiation response.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 9, pp. 5116–5121, 2001.
[82] M. Jeanmougin, A. de Reynies, L. Marisa, C. Paccard, G. Nuel, and M. Guedj, “Should we abandon the t-Test in the analysis of gene expression microarray data: A comparison of variance modeling strategies,” *PLoS One*, vol. 5, no. 9, pp. 1–9, 2010.
[83] R. Shamir, A. Maron-Katz, A. Tanay, C. Linhart, I. Steinfeld, R. Sharan, Y. Shiloh, and R. Elkon, “EXPANDER--an integrative program suite for microarray data analysis.,” *BMC Bioinformatics*, vol. 6, p. 232, 2005.
[84] I. Ulitsky, A. Maron-Katz, S. Shavit, D. Sagir, C. Linhart, R. Elkon, A. Tanay, R.

Sharan, Y. Shiloh, and R. Shamir, “Expander: from expression microarrays to networks and functions,” *Nat. Protoc.*, vol. 5, no. 2, pp. 303–322, Feb. 2010.

[85] I. Ulitsky and R. Shamir, “Identification of functional modules using network topology and high-throughput data.,” *BMC Syst. Biol.*, vol. 1, p. 8, 2007.

[86] I. Ulitsky, L. C. Laurent, and R. Shamir, “Towards computational prediction of microRNA function and activity,” *Nucleic Acids Res.*, vol. 38, no. 15, pp. e160–e160, 2010.

[87] R. Elkon, C. Linhart, R. Sharan, R. Shamir, and Y. Shiloh, “Genome-Wide In Silico Identification of Transcriptional Regulators Controlling the Cell Cycle in Human Cells,” pp. 773–780, 2003.

[88] A. Subramanian, A. Subramanian, P. Tamayo, P. Tamayo, V. K. Mootha, V. K. Mootha, S. Mukherjee, S. Mukherjee, B. L. Ebert, B. L. Ebert, M. a Gillette, M. a Gillette, A. Paulovich, A. Paulovich, S. L. Pomeroy, S. L. Pomeroy, T. R. Golub, T. R. Golub, E. S. Lander, E. S. Lander, J. P. Mesirov, and J. P. Mesirov, “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 43, pp. 15545–50, 2005.

[89] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, and Q. Morris, “The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function,” *Nucleic Acids Res.*, vol. 38, no. SUPPL. 2, pp. 214–220, 2010.

[90] A. D. Carri, M. Onorati, M. J. Lelos, V. Castiglioni, A. Faedo, R. Menon, S. Camnasio, R. Vuono, P. Spaiardi, F. Talpo, M. Toselli, G. Martino, R. a Barker, S. B. Dunnett, G. Biella, and E. Cattaneo, “Developmentally coordinated extrinsic signals drive human pluripotent stem cell differentiation toward authentic DARPP-32+ medium-sized spiny neurons.,” *Development*, vol. 140, no. 2, pp. 301–12, 2013.

[91] S. Bae, J. Park, and J. S. Kim, “Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases,” *Bioinformatics*, vol. 30, no. 10, pp. 1473–1475, 2014.

[92] L. P. Wong, R. T. H. Ong, W. T. Poh, X. Liu, P. Chen, R. Li, K. K. Y. Lam, N. E. Pillai, K. S. Sim, H. Xu, N. L. Sim, S. M. Teo, J. N. Foo, L. W. L. Tan, Y. Lim, S. H. Koo, L. S. H. Gan, C. Y. Cheng, S. Wee, E. P. H. Yap, P. C. Ng, W. Y. Lim, R. Soong, M. R. Wenk, T. Aung, T. Y. Wong, C. C. Khor, P. Little, K. S. Chia, and Y. Y. Teo, “Deep whole-genome sequencing of 100 southeast Asian malays,” *Am. J. Hum. Genet.*, vol. 92, no. 1, pp. 52–66, 2013.

[93] R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. P. Pant, J. Baccash, A. P. Borcherding, A. Brownley, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. C. Ebert, C. R. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, C. E. McBride, M. Morenzoni, R. E. Morey, K. Mutch, H. Perazich, K. Perry, B. a Peters, J. Peterson, C. L. Pethiyagoda, K. Pothuraju, C. Richter, A. M. Rosenbaum, S. Roy, J. Shafto, U. Sharanhovich, K. W. Shannon, C. G. Sheppy, M. Sun, J. V Thakuria, A. Tran, D. Vu, A. W. Zaranek, X. Wu, S. Drmanac, A. R. Oliphant, W. C. Banyai, B. Martin, D. G. Ballinger, G. M. Church, and C. a Reid, “Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays.,” *Science*, vol. 78, no. November, pp. 1–7, 2009.

[94] R. Daber, S. Sukhadia, and J. J. D. Morrissette, “Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets.,” *Cancer Genet.*, vol. 206, no. 12, pp. 441–448, Dec. 2013.

[95] S. Levy, G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern, B. P. Walenz, N. Axelrod, J. Huang, E. F. Kirkness, G. Denisov, Y. Lin, J. R. MacDonald, A. W. C. Pang, M. Shago, T. B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S. A. Kravitz, D. A. Busam, K. Y. Beeson, T. C. McIntosh, K. A. Remington, J. F. Abril, J. Gill, J. Borman, Y.-H. Rogers, M. E. Frazier, S. W. Scherer, R. L. Strausberg, and J. C. Venter, "The diploid genome sequence of an individual human.," *PLoS Biol.*, vol. 5, no. 10, p. e254, Sep. 2007.

[96] L. Villard, S. Briault, A. M. Lossi, C. Paringaux, J. Belougne, L. Colleaux, D. R. Pincus, E. Woollatt, J. Lespinasse, A. Munnich, C. Moraine, M. Fontes, and J. Gecz, "Two unrelated patients with inversions of the X chromosome and non-specific mental retardation: physical and transcriptional mapping of their common breakpoint region in Xq13.1.," *J. Med. Genet.*, vol. 36, no. 10, pp. 754–758, Oct. 1999.

[97] S. H. Subramony, J. Advincula, S. Perlman, R. L. Rosales, L. V. Lee, T. Ashizawa, and M. F. Waters, "Comprehensive phenotype of the p.arg420his allelic form of spinocerebellar ataxia type 13," *Cerebellum*, vol. 12, no. 6, pp. 932–936, 2013.

[98] E. Rogaeva, J. Johnson, A. E. Lang, C. Gulick, K. Gwinn-Hardy, T. Kawarai, C. Sato, A. Morgan, J. Werner, R. Nussbaum, A. Petit, M. S. Okun, A. McInerney, R. Mandel, J. L. Groen, H. H. Fernandez, R. Postuma, K. D. Foote, S. Salehi-Rad, Y. Liang, S. Reimsnider, A. Tandon, J. Hardy, P. St George-Hyslop, and A. B. Singleton, "Analysis of the PINK1 gene in a large cohort of cases with Parkinson disease.," *Arch. Neurol.*, vol. 61, no. 12, pp. 1898–1904, Dec. 2004.

[99] J. Doostzadeh, J. W. Tetrud, M. Allen-Auerbach, J. W. Langston, and B. Schule, "Novel features in a patient homozygous for the L347P mutation in the PINK1 gene.," *Parkinsonism Relat. Disord.*, vol. 13, no. 6, pp. 359–361, Aug. 2007.

[100] T. Fuchs, S. Gavarini, R. Saunders-Pullman, D. Raymond, M. E. Ehrlich, S. B. Bressman, and L. J. Ozelius, "Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia.," *Nat. Genet.*, vol. 41, no. 3, pp. 286–288, 2009.

[101] R. Saunders-Pullman, T. Fuchs, M. San Luciano, D. Raymond, A. Brashear, R. Ortega, A. Deik, L. J. Ozelius, and S. B. Bressman, "Heterogeneity in primary dystonia: Lessons from THAP1, GNAL, and TOR1A in Amish-Mennonites," *Mov. Disord.*, vol. 29, no. 6, pp. 812–818, 2014.

[102] Y. Tang, M. B. Schapiro, D. N. Franz, B. J. Patterson, F. J. Hickey, E. K. Schorry, R. J. Hopkin, M. Wylie, T. Narayan, T. A. Glauser, D. L. Gilbert, A. D. Hershey, and F. R. Sharp, "Blood expression profiles for tuberous sclerosis complex 2, neurofibromatosis type 1, and Down's syndrome.," *Ann. Neurol.*, vol. 56, no. 6, pp. 808–814, Dec. 2004.

[103] A. D. Strand, A. K. Aragaki, D. Shaw, T. Bird, J. Holton, C. Turner, S. J. Tapscott, S. J. Tabrizi, A. H. Schapira, C. Kooperberg, and J. M. Olson, "Gene expression in Huntington's disease skeletal muscle: A potential biomarker," *Hum. Mol. Genet.*, vol. 14, no. 13, pp. 1863–1876, 2005.

[104] M. R. Dalman, A. Deeter, G. Nimishakavi, and Z.-H. Duan, "Fold change and p-value cutoffs significantly alter microarray interpretations," *BMC Bioinformatics*, vol. 13, no. Suppl 2, p. S11, 2012.

[105] D. B. Allison, X. Cui, G. P. Page, and M. Sabripour, "Microarray data analysis: from disarray to consolidation and consensus.," *Nat. Rev. Genet.*, vol. 7, no. 1, pp. 55–65, 2006.

[106] M. P. White, A. J. Rufaihah, L. Liu, Y. T. Ghebremariam, K. N. Ivey, J. P. Cooke, and D. Srivastava, "Limited gene expression variation in human embryonic stem cell and induced pluripotent stem cell-derived endothelial cells.," *Stem Cells*, vol. 31, no. 1, pp. 92–103, Jan. 2013.

[107] A. Bithell, R. Johnson, and N. J. Buckley, "Transcriptional dysregulation of coding and non-coding genes in cellular models of Huntington's disease.,"

Biochem. Soc. Trans., vol. 37, no. Pt 6, pp. 1270–1275, Dec. 2009.

[108] S. Bhattacharya, X. Lou, P. Hwang, K. R. Rajashankar, X. Wang, J.-A. Gustafsson, R. J. Fletterick, R. H. Jacobson, and P. Webb, "Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 25, pp. 9103–9108, Jun. 2014.

[109] M. J. Friedman, A. G. Shah, Z.-H. Fang, E. G. Ward, S. T. Warren, S. Li, and X.-J. Li, "Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration.," *Nat. Neurosci.*, vol. 10, no. 12, pp. 1519–1528, Dec. 2007.

[110] A. G. Shah, M. J. Friedman, S. Huang, M. Roberts, X.-J. Li, and S. Li, "Transcriptional dysregulation of TrkA associates with neurodegeneration in spinocerebellar ataxia type 17.," *Hum. Mol. Genet.*, vol. 18, no. 21, pp. 4141–4152, Nov. 2009.

[111] T. S. Kuroda, M. Fukuda, H. Ariga, and K. Mikoshiba, "The Slp homology domain of synaptotagmin-like proteins 1-4 and Slac2 functions as a novel Rab27A binding domain.," *J. Biol. Chem.*, vol. 277, no. 11, pp. 9212–9218, Mar. 2002.

[112] M. Fukuda, "Versatile role of Rab27 in membrane trafficking: focus on the Rab27 effector families.," *J. Biochem.*, vol. 137, no. 1, pp. 9–16, Jan. 2005.

[113] D. Fenili, Y.-Q. Weng, I. Aubert, M. Nitz, and J. McLaurin, "Sodium/myo-Inositol transporters: substrate transport requirements and regional brain expression in the TgCRND8 mouse model of amyloid pathology.," *PLoS One*, vol. 6, no. 8, p. e24032, 2011.

[114] I. Ahmed, J. I. Sbodio, M. M. Harraz, R. Tyagi, J. C. Grima, L. K. Albacarys, M. E. Hubbi, R. Xu, S. Kim, B. D. Paul, and S. H. Snyder, "Huntington's disease: Neural dysfunction linked to inositol polyphosphate multikinase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 31, pp. 9751–9756, Aug. 2015.

[115] I. Wieland, S. Jakubiczka, P. Muschke, M. Cohen, H. Thiele, K. L. Gerlach, R. H. Adams, and P. Wieacker, "Mutations of the ephrin-B1 gene cause craniofrontonasal syndrome.," *Am. J. Hum. Genet.*, vol. 74, no. 6, pp. 1209–1215, Jun. 2004.

[116] S. Liu, Y. Zhang, H. Bian, and X. Li, "Gene expression profiling predicts pathways and genes associated with Parkinson's disease.," *Neurol. Sci. Off. J. Ital. Neurol. Soc. Ital. Soc. Clin. Neurophysiol.*, vol. 37, no. 1, pp. 73–79, Jan. 2016.

[117] M. Shen, "Basic Helix-loop-helix Protein DEC1 Promotes Chondrocyte Differentiation at the Early and Terminal Stages," *J. Biol. Chem.*, vol. 277, no. 51, pp. 50112–50120, 2002.

[118] Y. Qian, J. Zhang, B. Yan, and X. Chen, "DEC1, a basic helix-loop-helix transcription factor and a novel target gene of the p53 family, mediates p53-dependent premature senescence.," *J. Biol. Chem.*, vol. 283, no. 5, pp. 2896–2905, Feb. 2008.

[119] F. Coppede and L. Migliore, "DNA damage in neurodegenerative diseases.," *Mutat. Res.*, vol. 776, pp. 84–97, Jun. 2015.

[120] A. H. V Schapira, C. W. Olanow, J. T. Greenamyre, and E. Bezard, "Slowing of neurodegeneration in Parkinson's disease and Huntington's disease: future therapeutic perspectives.," *Lancet (London, England)*, vol. 384, no. 9942, pp. 545–555, Aug. 2014.

[121] P. De Rossi, V. Buggia-prévot, B. L. L. Clayton, J. B. Vasquez, C. Van Sanford, R. J. Andrew, R. Lesnick, A. Botté, C. Deyts, S. Salem, E. Rao, R. C. Rice, A. Parent, S. Kar, B. Popko, P. Pytel, S. Estus, and G. Thinakaran, "Predominant expression of Alzheimer ' s disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts," *Mol. Neurodegener.*, vol. 11, no. 59, pp. 1–21, 2016.

[122] X.-M. Wang, W. Y. Yik, P. Zhang, W. Lu, P. K. Dranchak, D. Shibata, S. J. Steinberg, and J. G. Hacia, "The gene expression profiles of induced pluripotent stem cells from individuals with childhood cerebral adrenoleukodystrophy are consistent with proposed mechanisms of pathogenesis.," *Stem Cell Res. Ther.*, vol. 3, no. 5, p. 39, 2012.
[123] J. P. Quinn and V. J. Bubb, "SVA retrotransposons as modulators of gene expression.," *Mob. Genet. Elements*, vol. 4, no. December, p. e32102, 2014.
[124] D. Spieler, M. Kaffe, F. Knauf, J. Bessa, J. J. Tena, F. Giesert, B. Schormair, E. Tilch, H. Lee, M. Horsch, D. Czamara, N. Karbalai, C. Von Toerne, M. Waldenberger, C. Gieger, P. Lichtner, M. Claussnitzer, R. Naumann, B. Müller-Myhsok, M. Torres, L. Garrett, J. Rozman, M. Klingenspor, V. Gailus-Durner, H. Fuchs, M. H. De Angelis, J. Beckers, S. M. Hölter, T. Meitinger, S. M. Hauck, H. Laumen, W. Wurst, F. Casares, J. L. Gómez-Skarmeta, and J. Winkelmann, "Restless Legs Syndrome-Associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon," *Genome Res.*, vol. 24, no. 4, pp. 592–603, 2014.
[125] D. A. Kleinjan and V. Van Heyningen, "Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease," *Am. J. Hum. Genet*, vol. 76, pp. 8–32, 2005.
[126] R. Yang, J. L. Kerschner, N. Gosalia, D. Neems, L. K. Gorsic, A. Safi, G. E. Crawford, S. T. Kosak, S. H. Leir, and A. Harris, "Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus," *Nucleic Acids Res.*, vol. 44, no. 7, pp. 3082–3094, 2015.
[127] F. Soldner, Y. Stelzer, C. S. Shivalila, B. J. Abraham, J. C. Latourelle, M. I. Barrasa, J. Goldmann, R. H. Myers, R. A. Young, and R. Jaenisch, "Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression.," *Nature*, vol. 533, no. 7601, pp. 1–20, 2016.
[128] K. Williams, J. Christensen, and K. Helin, "DNA methylation: TET proteins-guardians of CpG islands?," *EMBO Rep.*, vol. 13, no. 1, pp. 28–35, Jan. 2012.
[129] K. Li, G. Wang, T. Andersen, P. Zhou, and W. T. Pu, "Optimization of genome engineering approaches with the CRISPR/Cas9 system," *PLoS One*, vol. 9, no. 8, p. e105779, 2014.
[130] J. Schuster, J. Halvardson, L. Pilar Lorenzo, A. Ameur, M. Sobol, D. Raykova, G. Anneren, L. Feuk, and N. Dahl, "Transcriptome Profiling Reveals Degree of Variability in Induced Pluripotent Stem Cell Lines: Impact for Human Disease Modeling.," *Cell. Reprogram.*, vol. 17, no. 5, pp. 327–337, Oct. 2015.

Appendices

Primer sequences and conditions

Table 5 Primers used for haplotype and STRP analyses.

STRP	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temperature (°C)
DXS10015	CAGGCAAAGCACTAACCATTTC	AGAGCGAGACTCCATCTCAG	60
DXS10016	CAGACAAGGACAGAAAGGGGG	GGAGTTTTCCTCCCTCACCA	60
DXS10017	CGTCCTTCCATGAGGCACTTG	GACAGACTTGTTTGTACATGG	55
DXS10018	CGTGCACCACTGCGAGATTTG	CATCCATTAAGCATGAGTTAC	55
DXS559	CATCCTTCACCACTGCCTCCA	CTCCCTGCTCCCATCGCCAA	55

These products were amplified using 35 cycles of denaturation (95°C for 30s), the given annealing temperature (for 30s), and extension (72°C for 45s). An initial 5m of denaturation at 95°C and a final extension (10m at 72°C) were also incorporated before/after cycling.

Table 6 Primers used for Sanger sequencing.

Genetic alteration	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temperature (°C)
DSC1	GTTCAGAGAGTCAAGCATGG	TCTCCAAGTCCCCACTCAAC	60
DSC2	CCAACCTTACCTGTTTCCTC	GCAGGTCATGTTCCTAATTC	58
DSC3	GAAGTGAAGATGCAAGTCTC	GGAGTTCTGCCAGATGAG	56
DSC10	AGTGTCTTAAGGATTCAATG	GGATGTTGCACATTCTCAAC	56
DSC12	TTCTTCCCGTGGCACTGATC	TACTCACTCTGTGTGGCCTC	55
48-bp del*	AGCCAAGATCCCGACACTAC	TCCTCATGGCCAATCAAAAGTG	59

* 10% betaine added

Products were PCR-amplified using 35 cycles of denaturation (95°C for 30s), the given annealing temperature (for 30s), and extension (72°C for 45s). An initial 5m of denaturation (at 95°C) and a final extension (10m at 72°C) were also incorporated before/after cycling.

Table 7 Primers and cycling conditions used for detecting the SVA insertion.

	Forward primer (5'-3')	Reverse primer (5'-3')	Cycling
SVA ins	GTTCCATTGTGTGGTT GTACCAGCGTTTGTTC	CACATGAAAAGATGCCC AACATCATTAGCCATTAG	(5 cycles below) Denaturation 98°C for 10s Extension 74°C for 3m30s (5 cycles below) Denaturation 98°C for 10s Extension 72°C for 3m30s (5 cycles below) Denaturation 98°C for 10s Extension 70°C for 3m30s

			(20 cycles below) Denaturation 98°C for 10s Extension 68°C for 3m30s

Step-down cycling was performed according to the protocol of Kawarai et al. (Ref. [34]). An initial 2m of denaturation (at 94°C) and final extension step (68°C at 7m) were also incorporated before/after cycling.

Table 8 Primers used for Sanger sequencing of novel SNVs.

SNV	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temperature (°C)
ChrX:71262608	CCGCACCTGGCCCTGTTTGTT	ACAGATGGGCACGGTGGCTC	66
ChrX:71301439	AGCAGCCGCACTCCATCCAG	TGAGCTGTGTTTGGGGAACGG	56
ChrX:71339837	TGGCCAACATAGTAAGGCCCATCC	CGTGGCTCAAAATGATCCTCCTGC	65
Chrx:71633571	CGTAGAGGACTAGCAAGAG	CTTTATGGAAGTGCCAACC	58
ChrX:71653235	GAAAACATCACAATCTATGTGC	CCCAAGAGGAAAAAGTGG	58

Products were amplified using 35 cycles of denaturation (95°C for 30s), the given annealing temperature (for 30s), and extension (72°C for 45s). An initial 5m of denaturation (at 95°C) and a final extension (10m at 72°C) were also incorporated before/after cycling.

Alternatively, for primers with high annealing temperatures (>65°C) due to high GC content, GC-rich solution (Roche) and cycling was used. GC-rich mix: 50 ng DNA, 2 ul 5x buffer, 0.2 ul 10 mM dNTPs, 4 pmol each of forward and reverse primers, 1 uL 10 mM GC-rich dNTPs, 0.1 uL enzyme; water to final reaction volume of 10 uL. Cycling: 3m of initial denaturation (95°C), followed by 10 cycles of denaturation (95°C for 30s), annealing (30s), and elongation (72°C for 30s), then 25 cycles of denaturation (95°C for 30s), annealing (30s), and ramp-up elongation, i.e. 72°C for 45s initially then increasing by 5s per cycle. Final elongation is 72°C for 7m.

Coordinates are based on GRCh38.

Table 9 Primers used for qPCR experiments.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temperature (°C)
TAF1	CACAATCTCCTGGGCAGTCT	AGAGTCGGGAGAGCTTTCTG	60
OGT	CATGGAAGGGGCAGTACAAG	CTTTGAGCAGGTTCCCCAG	60
ACRC	ACATGGATGGGTGCAAAAA	CCACTGGTGTCATCACTGCT	60
CXCR3	AACCACAAGCACCAAAGCAG	GGCGTCATTTAGCACTTGGT	60
SYTL2	GAGGCCATCATGAAGGTTTTG	CTTCAGCTGCTGGTCATCCT	60
SLC5A3	CCACAGGATTGTTTTGGGTC	AAGGTGGTGGTTCGAATCTG	60
EFNB1	CAAGAACCTGGAGCCCGTAT	AGATGATGTCCAGCTTGTCTCC	60
MRPS6	CGCTACGAGCTGGCTTTAAT	TGCTCCTCTGTCCATCAGG	60

Gene	Forward primer (5'-3')	Reverse primer (5'-3')	Annealing temperature (°C)
ZADH2	ACGGAGACCTCCTCGTCC	GAGGCTTAACTGAGGGGTCA	60
ATP6V0E2	GCCGTCTGCTGTTACCTCTT	GCACGTACCAGATGGTCTCAT	62
BHLHE40	AGCAGAAAATCATTGCCCTG	CATGTCTGGAAACCTGAGCA	60
ZC3H12A	TCCACTCCCAGAAGAGGAAA	AGAAGACCTCCTTGTTCCCA	57
GZF1	TGTGGTGCAAGATTCACTCAG	AGAAGCAAAACTCTTGCCACA	60
GRIN2D	TATCCCGACATGCACAGCTA	AGCATCGTAGATGAAGGCGT	58
KCND2	CTTCGGCTAGCAAGTTCACC	GCTATGGTTTTTGGCACCAT	60
DBN1	AGCACAGGAGGAAACAGCAG	TGTGGGTCTCTTCCTCCTCA	60
ETV3	TGTAGGCCCAGTCAGGAAAC	GGAGACCGGACCGAAGAC	60
DUSP1	CAGTGGACAAACACCCTTCC	AGGACAACCACAAGGCAGAC	60
FLAD1	AAGGTGTTGGTGTCCTGAGTG	AGGGCCTCTGAACTTTCTCC	60
ZNF672	TTCTGTACCTCACCCCATGC	GTGTCTGAGGAGGCAGTCG	60
FPGS	CGAAGCTATGGCCTGAAGAC	AAGAGCTCAGGACTGATGGG	62
UBC	ATCGCTGTGATCGTCACTTG	TTGCCTTGACATTCTCGATG	60
EAR	GAGGCTGAGGCAGGAGAATCG	GTCGCCCAGGCTGGAGTG	60

Products were amplified using 40 cycles of denaturation (95°C for 10s), annealing at the given temperature (10s), and extension (72°C for 10s). An initial 5m of denaturation at 95°C was included prior to cycling.

Table 10 Primers and protocols used for *TAF1* isoform analyses.

Analysis	Forward primer	Reverse primer	Protocol
34' detection via cloning	AGTATACTAAGACTGCCCAG	CTGTGTTACCTGCTTTTCTGG	PCR using Taq polymerase
nTAF1-specific qPCR	GCCCTACACGCCTCTGG	GGAGCACTGGGAATATC	qPCR using SYBR Green I 60°C annealing temp
cTAF1-specific qPCR	GCCCTACACGCCTCTGC	GGAGCACTGGGAATATC	qPCR using SYBR Green I 60°C annealing temp
Full-length *TAF1* cloning	GCTAGCATGGGACCCGGTGCGATTTG	GCTAGCTCATTCATCAGAGTCCAAGTCA	PCR using Q5® high fidelity polymerase

For amplification using Taq polymerase, products were amplified using 35 cycles of denaturation (95°C for 30s), annealing (60°C for 30s), and extension (72°C for 45s). An initial 5m of denaturation at 95°C and a final extension (10m at 72°C) were also incorporated before/after cycling.

For amplification using SYBR Green I chemistry, products were amplified using 40 cycles of denaturation (95°C for 10s), annealing (60°C for 10s), and extension (72°C for 10s). An initial 5m of denaturation at 95°C was included prior to cycling.

For amplification using Q5® high-fidelity polymerase (NEB), products were amplified using 35 cycles of denaturation (98°C for 10s), annealing (60°C for 10s), and extension (72°C for 5m). An initial 30s of denaturation at 98°C and a final extension (5m at 72°C) were also incorporated before/after cycling. To set up the PCR reaction, 50 ng of DNA was mixed with 12.5 ul of Q5® Mix (which contains buffer, Q5® polymerase, and dNTPs), 1.25 ul each of 10 pmol/ul forward and reverse primers; water was added to a final reaction volume of 25 ul.

Table 11 Sequencing primers used for analysis of full-length *TAF1* cDNA.

Exon/Direction	Primer (5'-3')
Exon 1 (forward)	ATGGGACCCGGCTGCGATTTG
Exon 3 (forward)	GTTAGGAGTACAGAAGATGCTGTG
Exon 4 (reverse)	CAATCAATGTCTTCACAATCAGCATC
Exon 10 (forward)	CAAGGAGAGTAAGAAGGAATCA
Exon 11 (forward)	CCATGGAATCTCTCCAATGATGAG
Exon 11 (reverse)	CCATGGATCTTTCACTTCTGGCTGAGAC
Exon 14 (reverse)	CAATTGGCCAGGATGGAGAG
Exon 14-15 (forward)	CAATTGCTGCAAGCATTTGAG
Exon 20 (forward)	CGTAAATTTGGTGTGCCTGAGG
Exon 20 (reverse)	CCTCAGGCACACCAAATTTAC
Exon 25 (forward)	CATGTGGTGCCATTGGACAC
Exon 32 (forward)	AGAGTCGGGAGAGCTTTCTG
Exon 33 (reverse)	CACAATCTCCTGGGCAGTCT
Exon 36 (reverse)	GGAGTGGCACTGGGAATATC
Exon 39 (reverse)	CTCATTCATCAGAGTCCAAGTCA

Overlapping fragments were sequenced using the primers above to verify the full sequence of *TAF1*.

Reagents and solutions

Table 12 Other reagents and solutions used in the experiments.

Reagent	Manufacturer	Purpose
Ampicillin	Sigma-Aldrich	Antibiotic selection of plasmids
Bromphenol blue	Riedel-de-Haën	Loading buffer for gel electrophoresis
Deoxyribonucleotide (dNTP) mix	Amersham Biosciences	Incorporated into elongating PCR fragments
Dimethylsulfoxide (DMSO)	Sigma-Aldrich	Preservation agent for cell pellets
DNA ladder 100 bp and 1 kb	Invitrogen	Comparator of fragment size after gel electrophoresis
Ethanol (C_2H_5OH)	JT Baker	Precipitation of DNA/RNA prior to sequencing
Ethidium bromide ($C_{21}H_{20}BrN_3$)	ICN Biomedicals	Intercalates into DNA, allowing visualization in a gel
Exonuclease I (ExoI)	Fermentas	Degrades excess ssDNA (i.e. primers) in a dsDNA mixture
Alkaline phosphatase (FastAP)	Fermentas	Removes phosphate groups from DNA
Glycerol ($C_3H_5(OH)_3$)	Sigma-Aldrich	Preservation agent used for freezing bacteria
Highly deionized (HiDi) formamide	Applied Biosystems	Resuspends DNA in solution after precipitation
LE Agarose	Biozym	Preparation of gels for electrophoretic separation
$MgCl_2$	Sigma-Aldrich	Co-factor for Taq polymerase in PCR reaction
Midori Green	Nippon Genetics	Intercalates into DNA, allowing visualization in a gel
POP-7™ polymer	Applied Biosystems	Coats capillary of sequencer, controlling osmotic flow
Penicillin/Streptomycin (P/S)	Invitrogen	Inhibits bacterial growth in cell culture medium
Phosphate buffered saline (PBS)	Invitrogen	Washing cellular debris from culture plates
Sodium acetate (NaAc)	Sigma-Aldrich	Precipitation of DNA prior to sequencing
Taq DNA polymerase	Qbiogene	DNA polymerase for PCR amplification of fragments
TE buffer 10x (to make 1L) Tris 108 g Boric acid 55 g EDTA 7.4 g Deionized water to make 1L	Sigma-Aldrich	Solubilizes nucleic acids, also protects against degradation by binding cations required by DNAse
LB broth/agar (to make 1L) NaCl 10 g Tryptone 10 g Yeast extract 10 g (Agar 20 g) Dionized water to make 1L	Bacto™ (BD)	Nutrient-rich media used to culture bacteria during cloning experiments

Gene names

ACRC – Acid repeat-containing gene

ATP6V0E2 – ATPase, H+ transporting, lysosomal, 9-kD, V0 subunit E2

BHLHE40 – Basic helix-loop-helix family, member E40

CXCR3 – Chemokine, CXC motif, receptor 3

DBN1 – Drebrin E

DUSP1 – Dual-specificity phosphatase 1

EFNB1 – Ephrin B1

ETV3 – ETS variant gene 3

FLAD1 – Flavin adenine dinucleotide synthetase

FPGS – Folylpolyglutamate synthetase

GNAL – Guanine nucleotide-binding protein, alpha-activating activity polypeptide, olfactory

GRIN2D – Glutamate receptor, ionotropic, N-methyl-D-aspartate, subunit 2D

GZF1 – GDNF-inducible zinc finger protein 1

KCNC3 – Potassium channel, voltage-gated, shaw-related subfamily, member 3

KCND2 – Potassium voltage-gated channel, Shal-related subfamily, member 2

MRPS6 – Mitochondrial ribosomal protein S6

NONO – Non-POU domain containing, octamer-binding protein

OGT – O-linked N-acetylglucosamine transferase

PINK1 – PTEN-induced putative kinase 1

SLC5A3 – Solute carrier family 5 (inositol transporter), member 3

SYTL2 – Synaptotagmin-like 2

TAF1 – TATA-box binding protein-associated factor 1

THAP1 – THAP domain-containing protein 1

TOR1A – Torsin 1A

UBC – Ubiquitin C

ZADH2 – Zinc alcohol dehydrogenase

ZC3H12A – Zinc finger CCCH domain-containing protein 12A

ZNF672 – Zinc finger protein 672

Abbreviations

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

CNV – Copy Number Variant

DEG – Differentially Expressed Gene

DSB – Double-Stranded Break

DSC – Disease-specific Single-nucleotide Change

DYT3 – Dystonia 3 (old name for XDP based on dystonia locus symbols)

EAR – Expressed Alu Repeats

FDR – False Discovery Rate

GO – Gene Ontology

GSEA – Gene Set Enrichment Analysis

iPS – Induced Pluripotent Stem (cells)

KEGG – Kyoto Encyclopedia of Genes and Genomes

MIQE – Minimum Information for publication of Quantitative real-time PCR Experiments

MTS – Multiple Transcript System

PAM – Protospacer Associated Motif

PCR – Polymerase Chain Reaction

qPCR – Quantitative Polymerase Chain Reaction

RIN – RNA Integrity Number

SAM – Significance Analysis of Microarrays

SNP – Single-Nucleotide Polymorphism

SNV – Single-Nucleotide Variant

SVA – Short interspersed nuclear element, Variable number of tandem repeats, and Alu composite

STRP – Short Tandem Repeat Polymorphic markers

XDP – X-linked Dystonia-Parkinsonism

Index of tables

Index of figures

Acknowledgments

My endless gratitude goes to Prof. Christine Klein, for inviting me to come to the Institute, and for the support and trust that you gave to me and to my projects. The advancement of my career was always on your mind throughout these 4.5 years (even though I myself was not thinking of it), and because of that, it was a very fruitful stay in the lab.

I am also forever grateful to Ana Westenberger, for your support, trust, and motherly love. From you, I learned a special kind of work ethic. If one day I can be called a good scientist, it will be because of you. Completing this thesis would not have been possible without you.

I would also like to thank the Deutscher Akademischer Austauschdienst (DAAD) for the scholarship which allowed me to pursue PhD studies in Germany, as well as the Deutsche Dystonie-Gesselschaft for recognizing my work.

I am grateful to all patients with XDP who participated in our studies. You are the inspiration why we keep working hard to solve this mystery. This thesis is dedicated to you and to all colleagues in the Philippines who work tirelessly (and sometimes without compensation) for XDP.

To all the people in the lab/Institute who made lab work fun: all the postdocs, all the students, and all the technicians. In particular those who I've become good friends with. Ana, Karen, Eva, Victor, Thomas, Christoph Max, Marija, Melissa, Felix, Felix, Martje, Johanne, Katja, Aleks R, Alex M, Norbert, Karin, Britta, Hauke, Heike, Frauke: I am very thankful for the *Willkommenskultur* that you extended. I learned a lot about your amazing country and culture in the past 4.5 years; I consider Germany a second home because of all of you.

To my family and to Ian, thank you for your support from thousands of kilometers away, and for your long-distance love. It's comforting to know that whatever happens in Germany, I would be welcomed home with loving arms – with or without thesis. I cannot wait to come back.

Declaration

Hereby I declare that I have written this Dissertation completely on my own. Furthermore, I confirm that no other sources and resources have been used than those cited. This Dissertation, in the same or in a similar form, has not been submitted to any other committee for the purpose of obtaining a doctoral degree.

Aloysius Domingo
Lübeck, 25 October 2016