



Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in *TAF1*

D. Christopher Bragg^{a,b,1}, Kotchaphorn Mangkalaphiban^{a,b}, Christine A. Vaine^{a,b}, Nichita J. Kulkarni^{a,b}, David Shin^{a,b}, Rachita Yadav^{a,c}, Jyotsna Dhakal^{a,b}, Mai-Linh Ton^{a,b}, Anne Cheng^{a,b}, Christopher T. Russo^{a,b}, Mark Ang^d, Patrick Acuña^{a,b}, Criscely Go^e, Taylor N. Franceour^{a,b}, Trisha Multhaupt-Buell^{a,b}, Naoto Ito^{a,b}, Ulrich Müller^f, William T. Hendriks^{a,b}, Xandra O. Breakefield^{a,b,g}, Nutan Sharma^{a,b}, and Laurie J. Ozelius^{a,b,1}

^aThe Collaborative Center for X-linked Dystonia-Parkinsonism, Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129; ^bHarvard Brain Science Initiative, Harvard Medical School, Boston, MA 02114; ^cCenter for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114; ^dDepartment of Pathology, College of Medicine, University of the Philippines, Manila, Manila, Philippines; ^eDepartment of Neurology, Jose Reyes Memorial Medical Center, Quezon City, Philippines; ^fInstitut für Humangenetik, Justus Liebig University Giessen, D-35392 Giessen, Germany; and ^gCenter for Molecular Imaging Research, Department of Radiology, Massachusetts General Hospital, Charlestown, MA 02129

Edited by Solomon H. Snyder, Johns Hopkins University School of Medicine, Baltimore, MD, and approved October 24, 2017 (received for review July 13, 2017)

X-linked dystonia-parkinsonism (XDP) is a neurodegenerative disease associated with an antisense insertion of a SINE-VNTR-Alu (SVA)-type retrotransposon within an intron of *TAF1*. This unique insertion coincides with six additional noncoding sequence changes in *TAF1*, the gene that encodes TATA-binding protein-associated factor-1, which appear to be inherited together as an identical haplotype in all reported cases. Here we examined the sequence of this SVA in XDP patients ($n = 140$) and detected polymorphic variation in the length of a hexanucleotide repeat domain, (CCCTCT)_n. The number of repeats in these cases ranged from 35 to 52 and showed a highly significant inverse correlation with age at disease onset. Because other SVAs exhibit intrinsic promoter activity that depends in part on the hexameric domain, we assayed the transcriptional regulatory effects of varying hexameric lengths found in the unique XDP SVA retrotransposon using luciferase reporter constructs. When inserted sense or antisense to the luciferase reading frame, the XDP variants repressed or enhanced transcription, respectively, to an extent that appeared to vary with length of the hexamer. Further in silico analysis of this SVA sequence revealed multiple motifs predicted to form G-quadruplexes, with the greatest potential detected for the hexameric repeat domain. These data directly link sequence variation within the XDP-specific SVA sequence to phenotypic variability in clinical disease manifestation and provide insight into potential mechanisms by which this intronic retroelement may induce transcriptional interference in *TAF1* expression.

effort to identify the pathogenic gene variant (7, 12–22). The puzzling feature of XDP is that all reported patients inherit an apparently identical founder haplotype within a 294-kb genomic segment that spans seven sequence variants: five single-nucleotide substitutions, designated disease-specific sequence change (DSC) 1, 2, 3, 10, and 12; a 48-bp deletion; and a SINE-VNTR-Alu (SVA)-type retrotransposon insertion (7, 20, 22). These variants fall within noncoding regions in and around the *TAF1* gene, which encodes TATA-binding protein-associated factor-1 (TAF1), a component of the transcription factor II D (TFIID) complex involved in transcriptional regulation (23–27). To date there have been no reports of recombination events creating partial haplotypes in affected individuals, nor have any of these variants ever been found in ethnically matched or other control subjects (22).

Even though no study has yet demonstrated a functional role for any of these variants in XDP pathogenesis, the SVA is a compelling candidate given the growing list of human diseases associated with mobile DNA elements (28, 29). Nearly half of the human genome is made up of sequences derived from different classes of retroelements, three of which remain active and capable of retrotransposition: Alu, long interspersed nuclear element-1 (L1), and SVA (28–31). SVAs are specific to hominids and are the youngest family of retrotransposons, with ~3,600 annotated

XDP | DYT3 | dystonia | TAF1 | Parkinson's disease

X-linked dystonia-parkinsonism (XDP) is a progressive neurodegenerative disorder that affects individuals with maternal ancestry originating from the island of Panay, Philippines (1–3). The most frequently documented clinical phenotype consists of an initial focal dystonia that spreads to multiple body regions over time and combines with, or is replaced by, parkinsonism (2–6). Symptoms typically emerge in adulthood at an average age of 39.7 y, although a relatively wide age range for disease onset (12–79 y) has been reported in various patient cohorts (2, 4, 5, 7). Neuropathological analyses of a limited number of XDP brains at autopsy have demonstrated a selective loss of medium spiny neurons (MSNs) within the striatum (8, 9), resembling the lesion observed in Huntington's disease (HD) even though chorea, the characteristic feature of HD, is not typically observed in XDP. The involvement of other brain regions cannot yet be ruled out, however, as recent neuroimaging studies suggest that functional alterations in XDP are not limited to the striatum but also involve the thalamus and corticospinal tract (10, 11).

Although the disease was first described clinically more than 40 y ago, its genetic basis has remained elusive despite considerable

Significance

The genetic basis of X-linked dystonia-parkinsonism (XDP) has been difficult to unravel, in part because all patients inherit the same haplotype of seven sequence variants, none of which has ever been identified in control individuals. This study revealed that one of the haplotype markers, a retrotransposon insertion within an intron of *TAF1*, has a variable number of hexameric repeats among affected individuals with an increase in repeat number strongly correlated with earlier age at disease onset. These data support a contributing role for this sequence in disease pathogenesis while further suggesting that XDP may be part of a growing list of neurodegenerative disorders associated with unstable repeat expansions.

Author contributions: D.C.B., C.A.V., R.Y., N.J., X.O.B., and L.J.O. designed research; K.M., C.A.V., N.J.K., D.S., R.Y., J.D., M.-L.T., A.C., C.T.R., and W.T.H. performed research; P.A., C.G., T.N.F., T.M.-B., U.M., and N.S. contributed new reagents/analytic tools; D.C.B., C.A.V., R.Y., M.A., and L.J.O. analyzed data; and D.C.B., X.O.B., and L.J.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: bragg@helix.mgh.harvard.edu or lozelius@partners.org.

elements in the human reference sequence (31) which can be subdivided into seven subtypes, A–F and F1 (32, 33). Their composite structure consists of (5'–3') hexameric CCCTCT repeats, two antisense Alu-like fragments, a variable number of GC-rich tandem repeats (VNTR), sequences from the *env* gene of a human endogenous retrovirus (HERV), and a poly(A) tail (30). At least 85 human diseases are currently linked to active retroelements, with at least 12 associated specifically with SVAs (28, 34–45). These SVA insertions have been associated with various types of transcriptional interference, which may be due to their capacities to activate cryptic splice sites, generate transcripts via intrinsic promoter activity, and form G-quadruplex (G4) structures that inhibit the progression of RNA polymerase II (RNAP II) (32, 46–54). The functional consequences of the XDP-specific SVA are not yet known. Previous studies have reported aberrant transcription of *TAF1* exons in XDP caudate (7) as well as in cultured XDP fibroblasts (55, 56) and induced pluripotent stem cell (iPSC)-derived neural stem cells (NSCs) (55). Whether these transcriptional defects may be directly linked to the disease-specific SVA insertion has not yet been determined.

Because SVAs contain known polymorphic sequences, we examined the XDP-specific insertion by sequencing a BAC clone bearing the XDP haplotype region from a single proband. Comparison of the disease-specific SVA sequence in this individual with one previously reported (7) revealed a striking difference in the number of hexanucleotide (CCCTCT) repeats at the 5' end of the SVA. Based on that observation, we characterized variation in hexameric repeat length within the disease-specific SVA in XDP patients seen in clinics in North America and in the Philippines and in archival DNA specimens that were previously sequenced (20). To determine the potential functional significance of this

variation, we further examined the correlation between repeat length and age at disease onset (AO), as well as the effects of repeat length on transcriptional activity of the SVA in a reporter gene assay, given that the hexameric repeat domain in other SVAs is a major determinant of intrinsic promoter activity (52). The results from these analyses reveal that the disease haplotype is not identical in all affected XDP individuals. Rather, the disease-specific SVA insertion is polymorphic with respect to the number of hexameric repeats, and the length of this domain shows a highly significant inverse correlation with AO. In reporter assays in cultured cells, this variation modulates at least one functional property of the SVA, which in other retroelements has been associated with transcriptional interference of the surrounding host gene. Collectively these results further support the hypothesis that the disease-specific SVA insertion in intron 32 of *TAF1* is a major determinant of pathogenesis in individuals with XDP.

Results

Sequence Variation Within the XDP SVA and Correlation with AO. Fig. 1 depicts the genomic segment on chromosome Xq13.1 that was first associated with XDP through linkage analyses (12, 13) and the positions of the seven known disease-specific haplotype markers, including the unique SVA insertion (7, 20). Given the length and highly repetitive nature of SVAs, it can be difficult to determine their complete sequence using conventional short-read technologies. To overcome that limitation, we generated a BAC clone bearing an ~200-kb genomic segment spanning the XDP-unique SVA insertion site from a single proband and used Pacific Biosciences long-read single-molecule real-time sequencing (PacBio SMRT) to determine its sequence. That analysis successfully assembled the complete SVA sequence from this individual, which

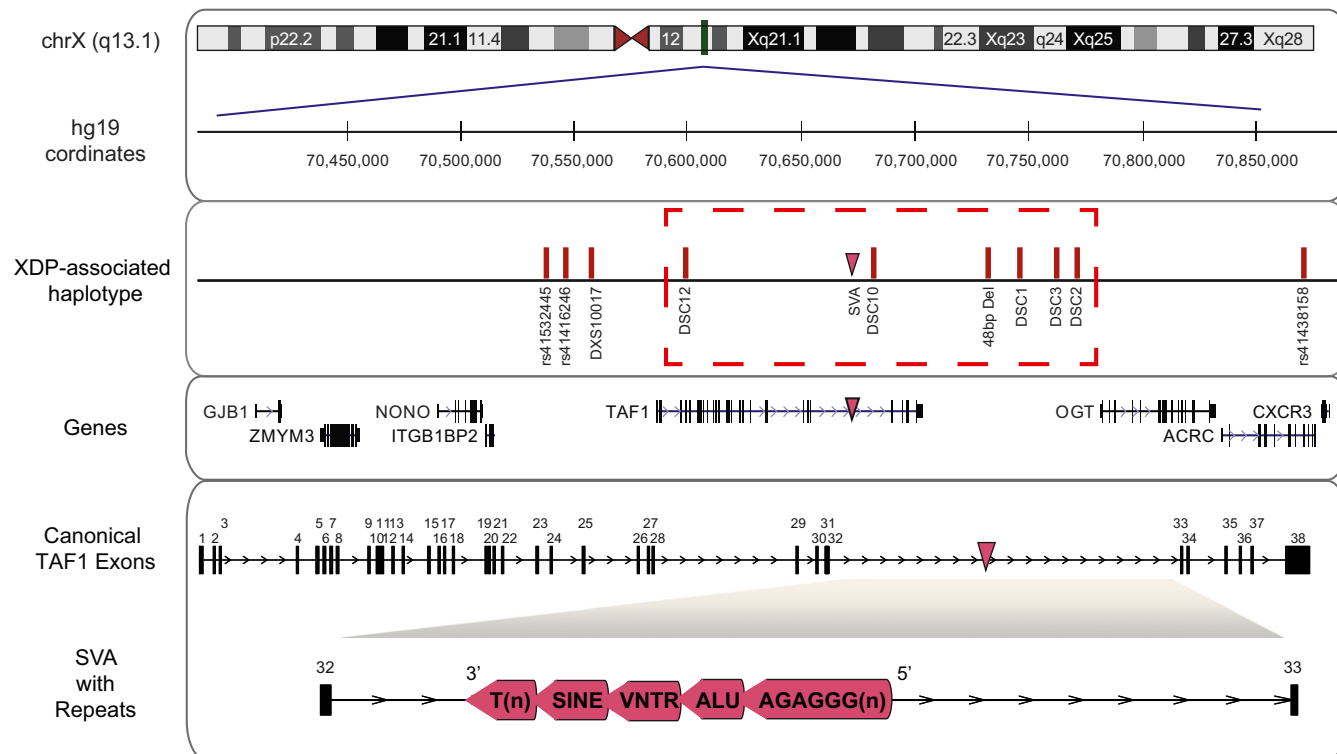


Fig. 1. (Upper) The genomic segment previously associated with XDP on chromosome Xq13.1 with hg19 coordinates, seven known XDP-specific variants that comprise the disease haplotype (boxed region), and flanking markers used to narrow the region. Haplotype variants consist of five single-nucleotide substitutions annotated as DSC-1, 2, 3, 10, and 12; a 48-bp deletion (48 bp Del), and a SVA-type retrotransposon insertion. Eight genes are shown within the broader linkage region, including *TAF1*. (Lower) Canonical exons of *TAF1*, the relative position of the SVA inserted antisense to *TAF1*, and the domain structure of the SVA consisting of (5'–3') a hexameric repeat (CCCTCT) of variable length, an Alu-like domain, a VNTR, a SINE domain, and a poly(A) tail.

Table 1. Clinical characteristics of XDP subjects

Cohort	<i>N</i>	Mean AO, y	Range, y	% DYT*	% PD*	% both*	% unknown*
United States	14	39.43	32–58	64	29	0	7
Philippines	67	41.79	25–61	64	15	20	1
Archival	59	39.98	23–60	61	2	0	37
Total	140	40.79	23–61	63	11	9	17

*Percent of cases with dystonia (DYT), parkinsonism (PD), or both as initial presentation and percent of cases for which initial symptoms are unknown.

we compared with one previously reported (7). Alignment of the two XDP SVAs revealed nearly complete homology except in the number of CCCTCT hexameric repeats: 35 in this study vs. 18 in the SVA identified by Makino et al. (7).

Based on that observation, we hypothesized that the SVA hexameric domain may be capable of expansion in XDP individuals and that its length may correlate with clinical disease manifestation, as reported for other disorders associated with repeat sequence expansions (57, 58). To test that possibility, we quantified hexameric repeat length in genomic DNA (gDNA) samples obtained from 140 individuals, including patients seen in both North America and in the Philippines, as well as from archival DNA specimens previously analyzed (20). Within this cohort, the length of the SVA hexamer ranged from 35 to 52 repeats. We then asked if this variation correlated with clinical features in these individuals (summarized in Table 1). The metric consistently reported for all affected subjects, as well as the archival specimens, was AO. We therefore performed linear regression analysis to determine the relationship between AO and repeat length, which revealed a highly significant inverse correlation ($R^2 = 0.507$; $P = 3.54 \times 10^{-23}$) (Fig. 2A). This significant correlation was also apparent when examining each subgroup (North America, Philippines, and archival DNA) individually (Fig. 2B–D). As an index of disease duration, we also calculated the time between the reported AO and age at sample collection for each individual, but this measure did not correlate with repeat length ($R^2 = 0.000422$; $P = 0.8351$).

SVA Sequence Variation Within Pedigrees and Differential Transmission.

Some of the subjects in this cohort were members of pedigrees with multiple haplotype-positive individuals. In these cases we noted that the SVA hexameric repeat length varied even within families (Fig. 3A). Furthermore, in the limited number of cases for which we could compare intergenerational transmission of the SVA from male vs. female haplotype carriers, the change in repeat length appeared to differ depending on the sex of the parent. Fig. 3B shows the percentage of cases in which SVA repeat length in a child increased, remained stable, or decreased when inherited from a mother vs. a father. When inherited from the mother, the SVA repeat primarily increased in length or remained stable, whereas with inheritance from fathers there was a more frequent tendency for the repeat to decrease when transmitted to daughters. χ^2 analysis revealed a significant difference in these distributions ($\chi^2 = 23.35$, $df = 2$; $P < 0.0001$), suggesting that the sex of the parent may be a critical determinant of how the hexameric repeat changes during inheritance.

Given these observations, we tested whether gDNA isolated from buccal cells of XDP patients ($n = 9$) showed repeat sizes similar to those in gDNA isolated directly from blood and saw no differences. We also assessed whether changes in SVA hexameric repeat length were found in XDP fibroblasts, iPSCs, and iPSC-differentiated NSCs during continuous propagation in culture. In most cases, the hexameric repeat length for each individual was equivalent in gDNA isolated from blood, saliva, and cell lines, with only minor variations in length occasionally detected in a subset of cell lines during propagation. Thus, significant slippage of the

SVA hexamer did not typically occur in these cell lines under the growth conditions tested, nor was there evidence of significant mosaicism among the cell types profiled.

Effects of SVA Hexamer Length on Intrinsic Promoter Activity. To determine if variation in hexameric repeat length alters a functional property of the XDP-specific SVA, we performed a reporter gene assay in which the firefly luciferase (FLuc) coding sequence was placed under the transcriptional control of the disease-specific SVA bearing 35, 41, or 52 hexameric repeats in either orientation (Hex35, Hex41, or Hex52) (Fig. 4A). As an additional comparison, we also generated an SVA construct lacking a hexameric repeat domain (Δ Hex). Although we did not detect significant slippage of the endogenous hexameric repeat sequence during continuous culture of XDP cell lines, plasmid constructs bearing these SVAs did at times exhibit substantial changes in the number of hexameric repeats following growth in bacteria. That effect was mitigated by propagating plasmid vectors in bacterial strains that exhibit reduced frequencies of homologous recombination. Before functional testing, the hexameric repeat length in each SVA–luciferase reporter construct was confirmed using the PCR sizing assay.

The intrinsic promoter activity of each SVA variant was assayed in human SH-SY5Y neuroblastoma cells and U2OS osteosarcoma cells, which have been shown to support particularly high levels of retrotransposition by SVAs (59). Although the

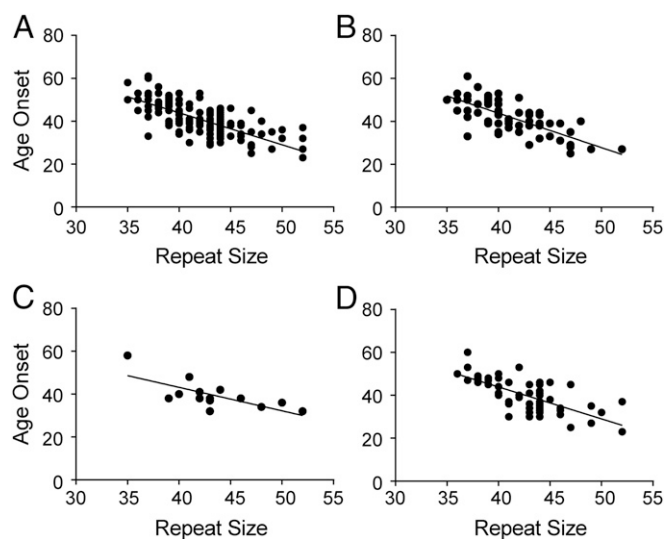


Fig. 2. Length of the hexameric repeat is polymorphic in affected XDP individuals and is inversely correlated with AO based on linear regression analysis. (A) Correlation between repeat length and AO in the entire cohort; $n = 140$, $R^2 = 0.507$, $P = 3.54 \times 10^{-23}$. (B–D) Analysis of individual subgroups revealed similar correlations in probands seen in Philippines clinics ($n = 67$, $R^2 = 0.5073$, $P = 1.39 \times 10^{-11}$) (B), probands seen in a US clinic ($n = 14$, $R^2 = 0.519$, $P = 0.003658$) (C), and archival DNA samples (20) ($n = 59$, $R^2 = 0.5055$, $P = 1.79 \times 10^{-10}$) (D).

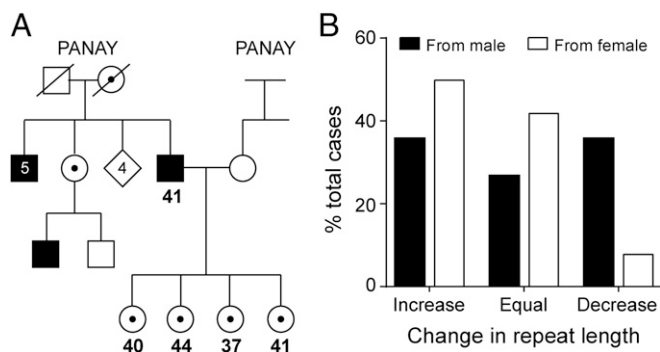


Fig. 3. (A) A representative XDP pedigree with multiple haplotype-positive individuals, including an affected proband (black box) with 41 repeats of the SVA hexamer and four daughters (circles) with different tract lengths ranging from 37 to 44 hexameric repeats. (B) Contingency table of intergenerational pairs depicting the change in hexameric repeat tract length during transmission through the male (solid bars) vs. female (open bars) germ lines. Because the repeat is present on the X chromosome, mothers transmitted the allele to both sons and daughters, whereas fathers transmitted only to daughters. When inherited from females ($n = 42$), the hexameric repeat length primarily remained the same or increased, but when inherited from males ($n = 11$) there was an increased frequency of contractions. χ^2 analysis revealed a significant difference in these distributions ($\chi^2 = 23.35$, $df = 2$; $P < 0.0001$), suggesting that parental sex may influence expansion vs. contraction of the hexamer.

pGL3b vector itself lacks a defined promoter element, it produced a moderate level of basal FLuc activity in both cell types. In SH-SY5Y cells, the SVAs with different hexameric repeat lengths all significantly repressed this basal activity when inserted in the forward direction, i.e., sense to the FLuc reading frame (Fig. 4B). Conversely, when inserted in the reverse orientation (i.e., antisense to FLuc), the longer Hex41 and Hex52 variants significantly increased FLuc activity above basal levels, but the shorter Hex35 variant produced only minimal effect. The truncated Δ Hex SVA in both orientations was less effective at modulating basal FLuc activity produced by pGL3b alone, suggesting that most of the transcriptional activity of the SVA in SH-SY5Y cells derived from the hexameric repeat domain. In both orientations, the long Hex52 SVA produced a greater effect on average than the shorter variants, but these differences did not achieve statistical significance with correction for multiple hypothesis testing.

In U2OS cells, the Hex52 and Hex35 SVAs significantly repressed luciferase activity when inserted in the forward direction (Fig. 4C), although the Hex41 variant was less effective. In contrast, the truncated Δ Hex SVA in the same orientation enhanced FLuc activity, suggesting that in U2OS cells other domains within the SVA may also contribute to intrinsic promoter activity in different ways. When inserted in the reverse direction, all SVA variants increased FLuc activity to levels ~ 5 – 10 -fold higher than that detected in pGL3b-transfected control cells (Fig. 4C). Unlike the pattern detected in SH-SY5Y cells, the short Hex35 variant produced the greatest increase in luciferase activity. Although there was an apparent inverse relationship between repeat length and the effect on luciferase activity, the differences between SVA constructs did not achieve significance with correction for multiple hypothesis testing. Thus, in this assay, the SVA in either orientation (sense or antisense) modulated expression of the FLuc reporter in a manner that varied with cell type. Consistent with previous reports, we detected a significant effect on reporter activity induced by the presence of the hexameric repeat domain, based on comparisons with the truncated variant lacking this motif, as well as trends suggesting

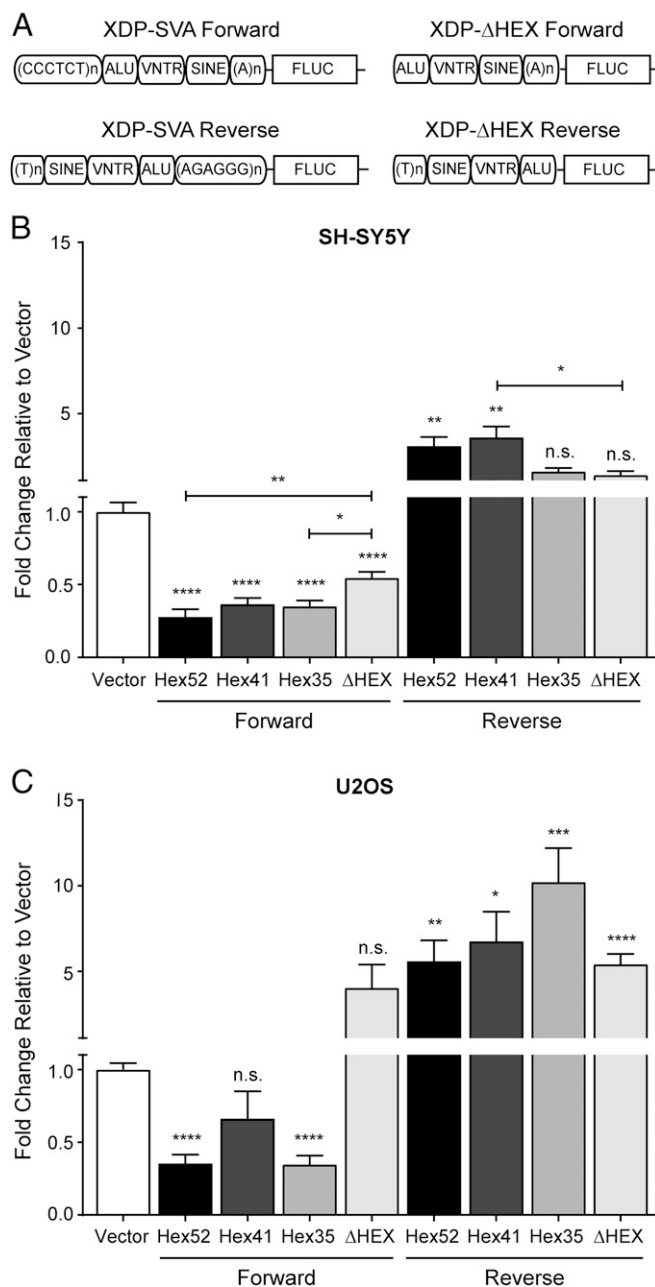


Fig. 4. Luciferase reporter assay to quantify intrinsic promoter activity of the XDP-specific SVA. (A) Schematic depicting the reporter constructs. Three versions of SVA forward and reverse orientations were generated, representing hexamers with 52, 41, or 35 repeats (Hex52, Hex41, and Hex35, respectively). A truncated SVA lacking the hexameric repeat domain (Δ Hex) was also generated in both forward and reverse orientations. (B) Fold changes in luciferase activity in SH-SY5Y cells produced by SVA constructs relative to basal level induced by the pGL3b vector alone. Data represent fold-change values averaged across four replicate experiments, shown as SEM. Significance was assessed by one-way ANOVA followed by post hoc Student's t tests with Bonferroni correction. Asterisks above bars denote the significance for the indicated SVA construct compared with vector alone. Asterisks above lines indicate the significance for additional comparisons. n.s., not significant. All four SVA forward constructs significantly repressed luciferase activity, although the Δ Hex variant was less effective than the constructs bearing hexameric repeats. In the reverse orientation, the longer SVAs (Hex52 and Hex41) significantly increased luciferase activity, whereas the shorter variants (Hex35 and Δ Hex) exhibited minimal effect. (C) In U2OS cells, the Hex52 and Hex35 SVAs in the forward orientation significantly repressed activity, whereas the truncated Δ Hex SVA significantly increased it. Increased activity was also produced by all SVA variants in the reverse orientation, with the greatest increase produced by the Hex35 variant. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

that the length of the repeat tract may influence its transcriptional behavior in some cellular contexts.

Predicted G4 Formation. G4 motifs are structures formed by stacked guanine tetrads that can significantly influence transcriptional dynamics and are often found within promoter elements and 5' UTRs of human genes (60, 61). Because previous studies have shown that SVAs also exhibit a particularly strong tendency to form G4 structures (51), we aligned the sequence obtained from the XDP BAC clone with consensus Alu and SINE sequences to define the boundaries for these domains and then analyzed the G4 potential of each position using the *in silico* prediction tool QGRS Mapper (60). The program computes a G-score (0–105) based on the number of possible guanine tetrads and the length of the gaps between them within a given nucleotide sequence. Fig. 5 depicts G-scores for each nucleotide position within the different domains of the SVA, indicating the likelihood that each position may form part of a stable quadruplex. Scores are plotted for both the sense and antisense strands of the SVA bearing 35 repeats of the hexamer CCCTCT or AGAGGG, respectively (Fig. 5 *A* and *B*). For both strands, the central VNTR domain contained multiple guanine-rich regions predicted to form G4 structures. On the antisense strand, the hexameric AGAGGG repeat scored more highly than any other position within the SVA. Increasing the number of AGAGGG repeats from 35 to 41 or 52 maintained this G-score across a quadruplex extending the full length of the repeat (Fig. 5*C*).

Discussion

Over the past four decades, clinical studies of XDP cohorts have observed various degrees of heterogeneity within different measures of clinical disease, including AO, duration of illness, and the pattern and temporal evolution of dystonic vs. parkinsonian symptoms (2, 4–7, 18). Of these measures, AO in particular has been shown to vary widely, although a basis for this heterogeneity has not been determined. In this study we demonstrate that the disease-specific SVA in patients is polymorphic with respect to the length of its hexameric repeat domain and that this variation in repeat length accounts for ~50% of the variance in AO within our cohort. Similar DNA repeat expansions have been implicated in more than 40 human diseases, many of which affect the nervous system, such as Huntington's disease, frontotemporal dementia, amyotrophic lateral sclerosis, fragile X syndrome, myotonic dystrophy, spinal and bulbar muscular atrophy, and the spinocerebellar ataxias (58, 62–65). A common finding in these disorders is an inverse correlation between repeat length and AO (57, 58, 66–70). In some cases, repeat length has also been correlated with additional measures, including penetrance (71), age at death (72), neuropathology (73), and severity of particular symptoms (70, 74). In this study we did not have sufficient phenotypic data on all subjects to probe for correlations between repeat length and other clinical parameters aside from AO, although future prospective studies of larger XDP cohorts can now address these questions.

The XDP-specific SVA repeat also exhibited intergenerational instability, which is another common feature of expansion disorders (63). Pathogenic repeats which consistently expand during inheritance can result in decreased AO and/or increased disease severity in subsequent generations, a phenomenon known as “genetic anticipation” (63). However, previous clinical studies have not reported obvious signs of possible anticipation in XDP, which may be consistent with our observation that, in our cohort, the SVA repeat underwent both expansion and contraction during inheritance. This instability appeared to vary during transmission through the male vs. female germ lines, as has also been documented for other repeat expansion disorders (75–78). However, given the relatively small number of intergenerational pairs within our cohort, this finding should be interpreted with caution.

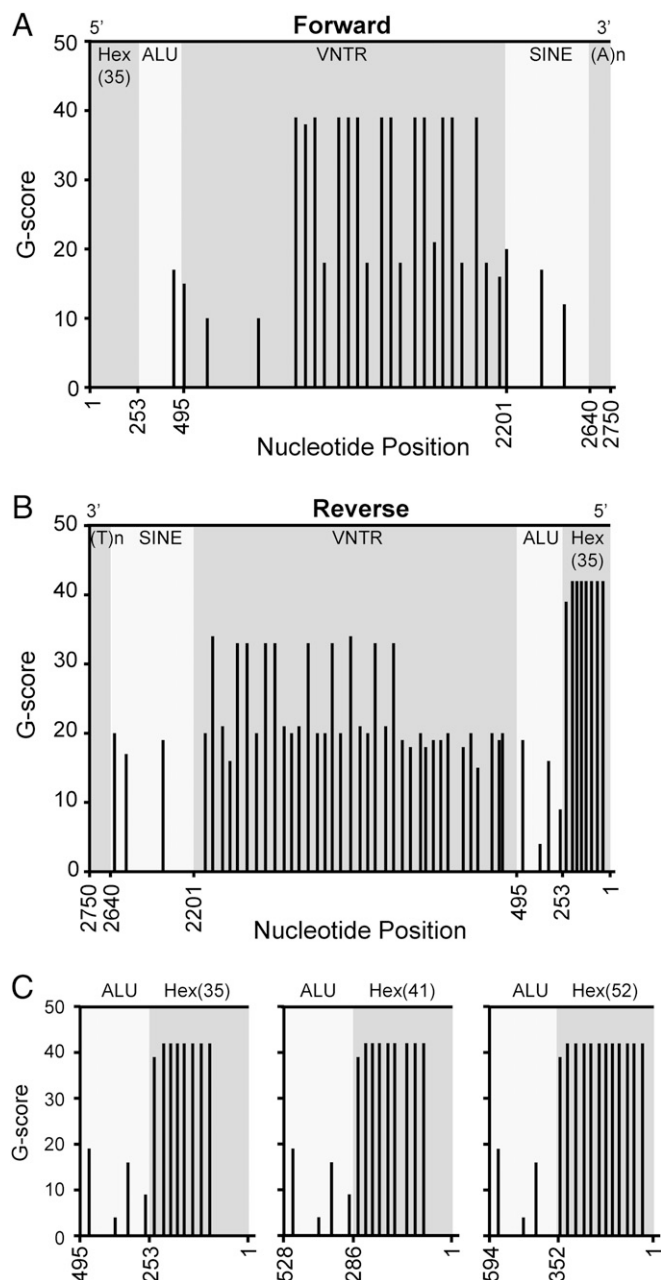


Fig. 5. (*A* and *B*) Predicted G4 formation by each position in the XDP-specific SVA on both forward (sense; 5'–3') (*A*) and reverse (antisense; 3'–5') (*B*) strands. Predicted boundaries are designated for the SVA functional domains: the hexameric repeat (Hex), Alu-like domain (Alu), VNTR, SINE-R domain, and a poly(A) or poly(T) sequence. The x axis depicts the nucleotide position within the SVA sequence of the individual bearing 35 hexameric repeats, as derived from the BAC clone. The y axis represents the G-score computed by QGRS Mapper software (60). The central VNTR region in both orientations includes positions predicted to form G4 structures, whereas the AGAGGG, hexamer in the reverse orientation exhibits greater G4 potential than any other position in the SVA. (*C*) Increasing the number of hexameric repeats to 41 or 52 extends the length of the predicted quadruplex as shown.

While XDP may share features with other neurodegenerative diseases linked to DNA repeats, it is distinctive in that the sequence expansion occurs within a retrotransposon. SVA-type retrotransposons occupy ~0.2% of the human genome, with approximately one third of these insertions residing in genic regions and new insertions occurring at an estimated rate of 1 in

916 births (30, 79, 80). Most of these insertion events are apparently benign, but at least a dozen have been associated with various hereditary syndromes and cancers (28, 29, 36, 39, 43–45). These insertions have been shown to disrupt gene expression in multiple ways, including (i) deletion of host gene sequences, as in neurofibromatosis type 1 (40) and leukemia (39); (ii) altered splicing events, including exon skipping as occurs in X-linked agammaglobulinemia (38), hereditary elliptocytosis, and hereditary pyropoikilocytosis (34); and (iii) exonization of SVA sequences, which has been observed in Fukuyama-type muscular dystrophy (35), autosomal recessive hypercholesterolemia (37), neutral lipid storage disease with myopathy (41), and Lynch syndrome (42). Other nonpathogenic SVA insertions may instead regulate transcription by modulating existing promoter elements and/or by creating new promoters via their intrinsic activities (32, 48, 49, 51–54, 81). The latter may be due in part to the high GC content of SVAs, which may establish CpG island-like domains capable of recruiting transcription factors and altering local chromatin structure (32, 54, 82).

We assessed the intrinsic promoter activity of the XDP-specific SVA in a conventional luciferase reporter assay, which revealed that it may potentially act as a bidirectional transcriptional regulatory unit. The extent and nature of this regulation (e.g., repression or enhancement) varied with cell type, consistent with similar analyses of SVAs inserted within the *PARK7* (51) and *FUS* (52) genes. In many cases, the introduction of a functional promoter element within an intron, as may occur with a retroelement insertion, has been shown to interfere with transcription of surrounding genomic segments (50). This transcriptional interference may develop due to collisions between a primary RNAP II complex transcribing the host gene and a competing RNAP II and/or transcription factors bound to the intronic element (50, 83, 84). Such collisions may stall or dislodge the primary RNAP II complex, thereby disrupting the proper splicing and/or elongation of the host gene transcript (50, 83, 84). In addition to obstacles created by DNA-bound factors, the DNA sequence itself may impede transcription by forming secondary structures such as G4 quadruplexes which can inhibit progression of RNAP II (51, 60, 61). Retroelements are a major source of G4 structures within the human genome, and SVAs reportedly have the highest proportion of G4-forming sequences of all retroelement families (51, 85). Consistent with that observation, in silico mapping of the XDP-specific SVA predicted G4 formation by both the hexameric repeat and multiple sites within the central VNTR domain. These potential barriers to transcription are collectively summarized in Fig. 6, illustrating how an RNAP II complex transcribing *TAF1* exons could be hindered by factors and sequences associated with the intronic, XDP-specific SVA.

In addition to the potential effects on transcription, the capacity of the XDP hexamer to form an uninterrupted, contiguous quadruplex may also underlie its genomic instability. Changes in repeat tract length have been linked mechanistically to their tendency to form secondary structures, including stacked G4 tetrads, which promote errors by DNA polymerase (62, 63, 86). Moreover, this instability is not limited to germ-line transmission, as it can be triggered potentially during any event that unwinds DNA, including DNA repair in nondividing cells such as neurons (62, 63, 86). In this study we detected only minor variation in repeat length among the cell types profiled, but for many neurodegenerative disorders, somatic mosaicism in the nervous system may be a critical factor in pathogenesis (62, 65, 87–89). These somatic repeat expansions may be modulated by multiple cellular factors, including components of the DNA mismatch repair, base excision repair, and oxidative stress pathways (62, 87, 90). Most intriguingly, recent work in model systems suggests that some of these factors might be targeted therapeutically by agents that pharmacologically restrict somatic repeat expansion (65, 87). Whether such agents might prove beneficial in XDP remains to be determined, as further

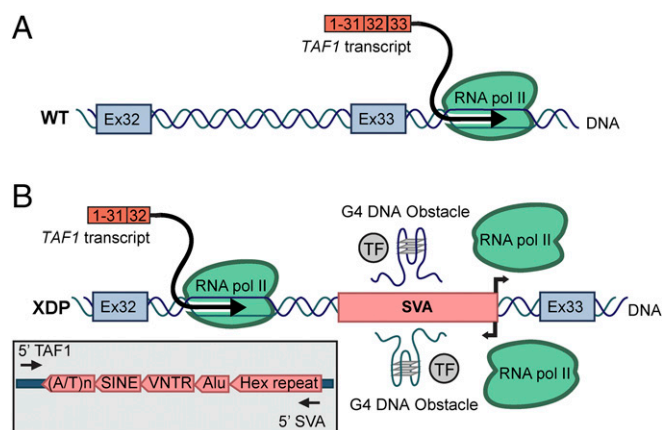


Fig. 6. Hypothesized sources of transcriptional interference associated with the XDP-specific SVA insertion in an intron of *TAF1* between flanking exons 32 (Ex32) and 33 (Ex33). (A) In wild-type cells, RNAP II successfully traverses the intron, generating a transcript that splices sequences derived from exons 32 and 33. (B) In XDP cells, the intronic SVA insertion may create multiple barriers to RNAP II transcribing *TAF1*, including transcription factors and/or competing RNAP II complexes associated with the SVA on both strands as well as G4 formation by the VNTR and hexameric domains. Obstacles that slow or prevent progression of RNAP II transcribing *TAF1* may potentially decrease transcription of downstream exons and/or alter splicing at these loci. (Inset) Organization of SVA inserted antisense to *TAF1*.

study is required to assess the extent to which somatic expansion of the SVA hexamer occurs within the brain.

Taken together the results from this study contrast with the consensus that has evolved from the genomic analyses of XDP reported to date. Those studies suggested that the five disease-specific single-nucleotide substitutions, the 48-bp deletion, and the unique SVA insertion are inherited together as an identical haplotype in all XDP probands (7, 20, 22). We find instead that the disease-specific SVA in probands includes a polymorphic, unstable repeat expansion, the length of which correlates with a measure of clinical disease in fashion similar to that documented in other repeat-expansion disorders. These data support the hypothesis that the SVA plays a causal role in XDP pathogenesis and further suggest that sequence variation within the SVA may be a critical disease modifier. Future studies to discern the mechanism (s) by which the expanded SVA hexamer affects transcription in patient cells are now a particular priority for understanding the physiological basis of XDP.

Materials and Methods

XDP Subjects and Sample Collection. Subjects recruited for this study included individuals with XDP evaluated at Massachusetts General Hospital (Boston), Jose R. Reyes Memorial Medical Center (Manila, Philippines), and regional clinics on the island of Panay (Philippines). All participants provided written informed consent, and the study was approved by all institutional review boards. Clinical evaluation included comprehensive neurological examinations with recorded scores for standard scales: Burke–Fahn–Marsden, Tsui–Torticollis, Toronto Western Spasmodic Torticollis Rating, and Voice Disability Index (91). Blood was collected from all patients, and a subset also provided saliva specimens. gDNA was extracted from blood using the Gentra Puregene kit (Qiagen) and from saliva using the Oragene Discover kit (DNA Genotek). Enrolled subjects were confirmed to be positive for six of the seven known haplotype markers (five DSCs and the SVA) by PCR amplification of blood gDNA followed by Sanger sequencing of amplicons as previously described (55). The collection methods and the clinical characterization of donor subjects who provided archival DNA specimens included in this study have been previously described (20).

BAC Sequencing Analysis. The generation, screening, sequencing, and assembly of XDP BAC clones was performed by Amplicon Express. Briefly, gDNA from a confirmed XDP-affected male was digested with BamHI and used to generate a BAC library in vector pBACe3.6 (92). Candidate positive clones

Table 2. Primers and PCR conditions

Target/application	Primer	Reagent	PCR conditions
XDP SVA Hexamer sizing	5'-[FAM]AGCAGTACAGTCCAGCTTTGGC-3' 5'-CTCAAGCCTTATTACAATGCCAGT-3'	PrimeSTAR GXL	94 °C 2 min; 30× (98 °C 10 s, 64 °C 35 s)
SVA-FLuc forward	5'-TAATACAAGCTTACCACCTGTTCCCTGTT- CCAC-3' 5'-GTATTAGCTAGCCTCAAGCCTTATTACA- AT GCCAGT-3'	PrimeSTAR GXL	98 °C 3 min; 30× (98 °C 10 s, 60 °C 15 s, 68 °C 3.5 min); 68 °C 5 min
SVA-FLuc reverse	5'-TAATACGCTAGCCACCACCTGTTCCCTGTT- CCACT-3' 5'-TGCAGAGAACTGGATCA-3'	PrimeSTAR GXL	98 °C 3 min; 30× (98 °C 10 s, 60 °C 15 s, 68 °C 3.5 min); 68 °C 5 min
SVAΔHEX-FLuc forward	5'-TAATACAAGCTTGTCCATTGTGTGGTT- GTA CCAGCGTTTGTTC-3' 5'-GTATTAGCTAGCGCCAAAGCTGGACTGT- AC TGCT-3'	PrimeSTAR GXL	98 °C 3 min; 30× (98 °C 10 s, 68 °C 3.5 min)
SVAΔHEX-FLuc reverse	5'-TAATACGCTAGCGTTCATTGTGTGGTT- GTA CCAGCGTTTGTTC-3' 5'-GTATTAAAGCTTGCCAAAGCTGGACTGT- AC TGCT-3'	PrimeSTAR GXL	98 °C 3 min; 30× (98 °C 10 s, 68 °C 3.5 min)
SVA-FLuc sequencing primers	5'-CTAGCAAAATAGGCTGTCCC-3' 5'-TTGTTAAACAGATGCTTGAAGGCAG-3' 5'-TGTCTCCACAAAACCAGTCAG-3' 5'-CCTTATGCAGTTGCTCTCC-3'	PrimeSTAR GXL	98 °C 3 min; 30× (98 °C 10 s, 68 °C 3.5 min)

were screened using either a probe downstream of the SVA (chrX: 70,674,877–70,675,733; between exons 34 and 35 of *TAF1*) or a probe upstream of the SVA (chrX: 70,613,608–70,613,809; between exons 21 and 22 of *TAF1*). Two clones were identified that were confirmed to bear the other XDP haplotype markers by PCR and then were subjected to long-read sequencing (93) for further verification and contig assembly. A library was generated using the PacBio 20-kb library preparation (Pacific Biosciences) and was sequenced on the PacBio RS II instrument (DNA Link). Sequencing data were then used to assemble a single 201,921-bp contig corresponding to hg19 positions 70,546,230 bp to 70,747,084 bp on the X chromosome.

Quantification of SVA Hexameric Repeat Length. To interrogate the number of hexameric repeats in the SVA of different individuals, we developed a fluorescence PCR-based assay. Blood and salivary gDNA samples from affected hemizygous males and heterozygous carrier females were PCR amplified with the primers and conditions listed in Table 2. One primer was labeled with a FAM tag to allow sizing of the repeats. Following PCR, aliquots of each product were resolved via electrophoresis to confirm amplification of the SVA sequence. To size the repeats, 0.8 μL of FAM-tagged product was mixed with 9.5 μL of loading buffer consisting of 9.4 μL of Hi-Di formamide and 0.1 μL of GeneScan500 LIZ (Thermo Fisher Scientific) as an internal size standard. The products were denatured for 5 min at 95 °C, and capillary electrophoresis was performed on the ABI 3500xL Genetic Analyzer (Applied Biosystems) under the fragment-analysis protocol. Raw data were processed using GeneMapper v5 software (Applied Biosystems).

Cell Culture. We evaluated potential slippage of the XDP hexameric repeat during continuous propagation in culture of XDP primary fibroblasts, iPSCs, and iPSC-differentiated NSCs. Fibroblast derivation, iPSC reprogramming, and NSC differentiation of iPSCs have all been previously reported (55). Additional experiments were performed using the human neuroblastoma (SH-SY5Y) and osteosarcoma (U2OS) cell lines (American Type Culture Collection). Fibroblasts and SH-SY5Y and U2OS cells were cultured in DMEM supplemented with either 10% (SH-SY5Y and U2OS cells) or 20% (fibroblasts) FBS and 1× penicillin/streptomycin L-glutamine. iPSCs were propagated on Geltrex-coated tissue-culture plates in mTESR-1 medium (STEMCELL Technologies). NSC lines were cultured on Geltrex-coated plates in Neurobasal medium [advanced DMEM/F12 (1:1) with 2% PSC neural induction supplement]. Fibroblasts and SH-SY5Y and U2OS cells were passaged using trypsin, whereas iPSCs and NSCs were collected via Accutase (Sigma). All cell lines were maintained in a humidified incubator at 37 °C with 5% CO₂. All media, supplements, and ancillary reagents were obtained from Thermo Fisher Scientific, except where noted.

Luciferase Reporter Assays. XDP-specific SVA variants bearing 35, 41, or 52 repeats of the hexamer were amplified by PCR from blood gDNA

obtained from probands. Amplification was performed as previously described (55) using primers flanking the SVA (Table 2) and PrimeSTAR GXL reagents (Takara Bio). Amplicons were cloned into the NheI and HindIII sites of the pGL3-basic (pGL3b) vector encoding FLuc (Promega) in either the sense or antisense orientation relative to the FLuc ORF. Additional constructs were generated bearing truncated SVAs lacking the hexameric repeat domain (ΔHex).

To evaluate transcriptional activity of these SVA constructs, cultured SH-SY5Y and U2OS cells were cotransfected with a pGL3b-FLuc reporter plasmid and a separate plasmid encoding *Renilla* luciferase (RLuc) under the control of the constitutive thymidine kinase promoter (TK-RLuc) (Promega). Transfection was performed using Lipofectamine 3000 (Thermo Fisher Scientific) as recommended, and mock-transfected cells were included as a negative control. Cells received a complete medium exchange the next day and were maintained in culture for an additional 24 h. FLuc and RLuc activities were quantified in cell lysates using the Dual-Glo Luciferase Assay System (Promega) as recommended on a Synergy HTX luminometer (BioTek). For each sample, relative light unit (RLU) values for FLuc and RLuc were adjusted for background activity based on counts measured in mock-transfected samples. The adjusted FLuc/RLuc ratio was calculated for each SVA variant, expressed as a fold-change relative to the ratio obtained for the pGL3b vector alone within each experiment, and averaged across at least four independent experiments.

SVA Functional Domains and G4 Prediction. The nucleotide sequence obtained from the XDP BAC clone was aligned to consensus sequences for human SINE-R (94) and Alu (95) elements to predict the boundaries for these domains relative to the VNTR and the hexameric repeat domain. The sequence was further analyzed using Quadruplex-forming G-Rich Sequence (QGRS) Mapper Software (60). Computed G-scores, indicating the potential for a given sequence motif to form a quadruplex structure, were compared for each position within the XDP disease-specific SVA on both forward (sense) and reverse (antisense) strands.

Statistical Analyses. Linear regression analysis was performed using R (<https://www.R-project.org>). One-way ANOVA, Student's *t* tests with Bonferroni correction, and χ^2 analyses were all performed using GraphPad Prism v7 software (GraphPad Software).

ACKNOWLEDGMENTS. We thank Dr. Winnie Xin (Massachusetts General Hospital, MGH) and Ms. Rosemary Barone (MGH) for assistance with quantification of hexameric repeat tract length in DNA samples. Funding for this study was provided by the MGH Collaborative Center for X-Linked Dystonia-Parkinsonism (D.C.B., N.S., and X.O.B.) and by NIH Grants 5P01NS087997 (to L.J.O., D.C.B., N.S., and X.O.B.) and R01NS102423 (to D.C.B. and L.J.O.).

1. Lee LV, Pascasio FM, Fuentes FD, Viterbo GH (1976) Torsion dystonia in Panay, Philippines. *Adv Neurol* 14:137–151.
2. Lee LV, Kupke KG, Caballar-Gonzaga F, Hebron-Ortiz M, Müller U (1991) The phenotype of the X-linked dystonia-parkinsonism syndrome. An assessment of 42 cases in the Philippines. *Medicine (Baltimore)* 70:179–187.
3. Lee LV, et al. (2011) The unique phenomenology of sex-linked dystonia parkinsonism (XDP, DYT3, "Lubag"). *Int J Neurosci* 121:3–11.
4. Lee LV, et al. (2002) The natural history of sex-linked recessive dystonia parkinsonism of Panay, Philippines (XDP). *Parkinsonism Relat Disord* 9:29–38.
5. Evidente VG, et al. (2002) Phenomenology of "Lubag" or X-linked dystonia-parkinsonism. *Mov Disord* 17:1271–1277.
6. Evidente VG, Gwinn-Hardy K, Hardy J, Hernandez D, Singleton A (2002) X-linked dystonia ("Lubag") presenting predominantly with parkinsonism: A more benign phenotype? *Mov Disord* 17:200–202.
7. Makino S, et al. (2007) Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* 80:393–406.
8. Goto S, et al. (2005) Functional anatomy of the basal ganglia in X-linked recessive dystonia-parkinsonism. *Ann Neurol* 58:7–17.
9. Goto S, et al. (2013) Defects in the striatal neuropeptide Y system in X-linked dystonia-parkinsonism. *Brain* 136:1555–1567.
10. Brüggemann N, et al. (2016) Neuroanatomical changes extend beyond striatal atrophy in X-linked dystonia parkinsonism. *Parkinsonism Relat Disord* 31:91–97.
11. Walter U, et al. (2017) Sonographic alteration of substantia nigra is related to parkinsonism-predominant course of X-linked dystonia-parkinsonism. *Parkinsonism Relat Disord* 37:43–49.
12. Kupke KG, Lee LV, Müller U (1990) Assignment of the X-linked torsion dystonia gene to Xq21 by linkage analysis. *Neurology* 40:1438–1442.
13. Wilhelmssen KC, et al. (1991) Genetic mapping of "Lubag" (X-linked dystonia-parkinsonism) in a Filipino kindred to the pericentromeric region of the X chromosome. *Ann Neurol* 29:124–131.
14. Kupke KG, Graeber MB, Müller U (1992) Dystonia-parkinsonism syndrome (XDP) locus: Flanking markers in Xq12-q21.1. *Am J Hum Genet* 50:808–815.
15. Graeber MB, Kupke KG, Müller U (1992) Delineation of the dystonia-parkinsonism syndrome locus in Xq13. *Proc Natl Acad Sci USA* 89:8245–8248.
16. Müller U, et al. (1994) DXS106 and DXS559 flank the X-linked dystonia-parkinsonism syndrome locus (DYT3). *Genomics* 23:114–117.
17. Haberhausen G, et al. (1995) 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* 57:644–650.
18. Wilhelmssen KC, et al. (1998) Molecular genetic analysis of Lubag. *Adv Neurol* 78:341–348.
19. Németh AH, et al. (1999) Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). *Genomics* 60:320–329.
20. Nolte D, Niemann S, Müller U (2003) Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism. *Proc Natl Acad Sci USA* 100:10347–10352.
21. Herzfeld T, Nolte D, Müller U (2007) Structural and functional analysis of the human TAF1/DYT3 multiple transcript system. *Mamm Genome* 18:787–795.
22. Domingo A, et al. (2015) New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3). *Eur J Hum Genet* 23:1334–1340.
23. Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 41:105–178.
24. Papai G, Weil PA, Schultz P (2011) New insights into the function of transcription factor TFIID from recent structural studies. *Curr Opin Genet Dev* 21:219–224.
25. Anandapadamanaban M, et al. (2013) High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. *Nat Struct Mol Biol* 20:1008–1014.
26. Grünberg S, Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* 38:603–611.
27. Malkowska M, Kokoszynska K, Rychlewski L, Wyrwicz L (2013) Structural bioinformatics of the general transcription factor TFIID. *Biochimie* 95:680–691.
28. Kaer K, Speck M (2013) Retroelements in human disease. *Gene* 518:231–241.
29. Hancks DC, Kazazian HH, Jr (2016) Roles for retrotransposon insertions in human disease. *Mob DNA* 7:9.
30. Hancks DC, Kazazian HH, Jr (2010) SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol* 20:234–245.
31. Stewart C, et al.; 1000 Genomes Project (2011) A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet* 7:e1002236.
32. Wang H, et al. (2005) SVA elements: A hominid-specific retroposon family. *J Mol Biol* 354:994–1007.
33. Bantysh OB, Buzdin AA (2009) Novel family of human transposable elements formed due to fusion of the first exon of gene MAST2 with retrotransposon SVA. *Biochemistry (Mosc)* 74:1393–1399.
34. Hassoun H, et al. (1994) A novel mobile element inserted in the alpha spectrin gene: Spectrin dayton. A truncated alpha spectrin associated with hereditary elliptocytosis. *J Clin Invest* 94:643–648.
35. Kobayashi K, et al. (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* 394:388–392.
36. Legoux P, et al. (2000) Molecular characterization of germline NF2 gene rearrangements. *Genomics* 65:62–66.
37. Wilund KR, et al. (2002) Molecular mechanisms of autosomal recessive hypercholesterolemia. *Hum Mol Genet* 11:3019–3030.
38. Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH, Jr (2005) Two independent retrotransposon insertions at the same site within the coding region of BTK. *Hum Mutat* 25:324–325.
39. Takasu M, et al. (2007) Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. *Tissue Antigens* 70:144–150.
40. Szpakowski S, et al. (2009) Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. *Gene* 448:151–167.
41. Akman HO, et al. (2010) Neutral lipid storage disease with subclinical myopathy due to a retrotransposal insertion in the PNPLA2 gene. *Neuromuscul Disord* 20:397–402.
42. van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT (2012) Insertion of an SVA element, a nonautonomous retrotransposon, in PMS2 intron 7 as a novel cause of Lynch syndrome. *Hum Mutat* 33:1051–1055.
43. Vogt J, et al. (2014) SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. *Genome Biol* 15:R80.
44. Nakamura Y, et al. (2015) SVA retrotransposon in exon 6 of the coagulation factor IX gene causing severe hemophilia B. *Int J Hematol* 102:134–139.
45. Stacey SN, et al. (2016) Insertion of an SVA-E retrotransposon into the CASP8 gene is associated with protection against prostate cancer. *Hum Mol Genet* 25:1008–1018.
46. Vorechovsky I (2010) Transposable elements in disease-associated cryptic exons. *Hum Genet* 127:135–154.
47. Hancks DC, Ewing AD, Chen JE, Tokunaga K, Kazazian HH, Jr (2009) Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* 19:1983–1991.
48. Kim DS, Hahn Y (2010) Human-specific antisense transcripts induced by the insertion of transposable element. *Int J Mol Med* 26:151–157.
49. Kim DS, Hahn Y (2011) Identification of human-specific transcript variants induced by DNA insertions in the human genome. *Bioinformatics* 27:14–21.
50. Kaer K, Speck M (2012) Intronic retroelements: Not just "speed bumps" for RNA polymerase II. *Mob Genet Elements* 2:154–157.
51. Savage AL, Bubb VJ, Breen G, Quinn JP (2013) Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* 13:101.
52. Savage AL, et al. (2014) An evaluation of a SVA retrotransposon in the FUS promoter as a transcriptional regulator and its association to ALS. *PLoS One* 9:e90833.
53. Quinn JP, Bubb VJ (2014) SVA retrotransposons as modulators of gene expression. *Mob Genet Elements* 4:e32102.
54. Gianfrancesco O, Bubb VJ, Quinn JP (2017) SVA retrotransposons as potential modulators of neuropeptide gene expression. *Neuropeptides* 64:3–7.
55. Ito N, et al. (2016) Decreased N-TAF1 expression in X-linked dystonia-parkinsonism patient-specific neural stem cells. *Dis Model Mech* 9:451–462.
56. Domingo A, et al. (2016) Evidence of TAF1 dysfunction in peripheral models of X-linked dystonia-parkinsonism. *Cell Mol Life Sci* 73:3205–3215.
57. Koshy BT, Zoghbi HY (1997) The CAG/polyglutamine tract diseases: Gene products and molecular pathogenesis. *Brain Pathol* 7:927–942.
58. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci* 30:575–621.
59. Hancks DC, Mandal PK, Cheung LE, Kazazian HH, Jr (2012) The minimal active human SVA retrotransposon requires only the 5'-hexamer and Alu-like domains. *Mol Cell Biol* 32:4718–4726.
60. Kikin O, D'Antonio L, Bagga PS (2006) QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* 34:W676–W682.
61. Hänzel-Hertsch R, Di Antonio M, Balasubramanian S (2017) DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 18:279–284.
62. Schmidt MH, Pearson CE (2016) Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst)* 38:117–126.
63. Mirkin SM (2007) Expandable DNA repeats and human disease. *Nature* 447:932–940.
64. Jones L, Houlden H, Tabrizi SJ (2017) DNA repair in the trinucleotide repeat disorders. *Lancet Neurol* 16:88–96.
65. Gomes-Pereira M, Monckton DG (2006) Chemical modifiers of unstable expanded simple sequence repeats: What goes up, could come down. *Mutat Res* 598:15–34.
66. Wexler NS, et al.; U.S.-Venezuela Collaborative Research Project (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci USA* 101:3498–3503.
67. van de Warrenburg BP, et al. (2002) Spinocerebellar ataxias in the Netherlands: Prevalence and age at onset variance analysis. *Neurology* 58:702–708.
68. Gijssels I, et al. (2016) The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol Psychiatry* 21:1112–1124.
69. Igarashi S, et al. (1992) Strong correlation between the number of CAG repeats in androgen receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. *Neurology* 42:2300–2302.
70. Yum K, Wang ET, Kalsotra A (2017) Myotonic dystrophy: Disease repeat range, penetrance, age of onset, and relationship between repeat size and phenotypes. *Curr Opin Genet Dev* 44:30–37.
71. McNeil SM, et al. (1997) Reduced penetrance of the Huntington's disease mutation. *Hum Mol Genet* 6:775–779.
72. Keum JW, et al. (2016) The HTT CAG-expansion mutation determines age at death but not disease duration in Huntington disease. *Am J Hum Genet* 98:287–298.
73. Hadzi TC, et al. (2012) Assessment of cortical and striatal involvement in 523 Huntington disease brains. *Neurology* 79:1708–1715.
74. Costa MdoC, Paulson HL (2012) Toward understanding Machado-Joseph disease. *Prog Neurobiol* 97:239–257.
75. Aziz NA, van Belzen MJ, Coops ID, Belfroid RD, Roos RA (2011) Parent-of-origin differences of mutant HTT CAG repeat instability in Huntington's disease. *Eur J Med Genet* 54:e413–e418.

