

## ARTICLE

# New insights into the genetics of X-linked dystonia-parkinsonism (XDP, DYT3)

Aloysius Domingo<sup>1,2,3,14</sup>, Ana Westenberger<sup>1,14</sup>, Lillian V Lee<sup>2</sup>, Ingrid Brønne<sup>4</sup>, Tian Liu<sup>5,6</sup>, Inga Vater<sup>7</sup>, Raymond Rosales<sup>2,8</sup>, Roland Dominic Jamora<sup>2,9</sup>, Paul Matthew Pasco<sup>2,9</sup>, Eva Maria Cutiongco-dela Paz<sup>2,10</sup>, Karen Freimann<sup>1</sup>, Thomas GPM Schmidt<sup>1</sup>, Dirk Dressler<sup>11</sup>, Frank J Kaiser<sup>12</sup>, Lars Bertram<sup>5,13</sup>, Jeanette Erdmann<sup>7</sup>, Katja Lohmann<sup>1</sup> and Christine Klein<sup>\*,1</sup>

X-linked recessive dystonia-parkinsonism is a rare movement disorder that is highly prevalent in Panay Island in the Philippines. Earlier studies identified seven different genetic alterations within a 427-kb disease locus on the X chromosome; however, the exact disease-causing variant among these is still not unequivocally determined. To further investigate the genetic cause of this disease, we sequenced all previously reported genetic alterations in 166 patients and 473 Filipino controls. Singly occurring variants in our ethnically matched controls would have allowed us to define these as polymorphisms, but none were found. Instead, we identified five patients carrying none of the disease-associated variants, and one male control carrying all of them. In parallel, we searched for novel single-nucleotide variants using next-generation sequencing. We did not identify any shared variants in coding regions of the X chromosome. However, by validating intergenic variants discovered via genome sequencing, we were able to define the boundaries of the disease-specific haplotype and narrow the disease locus to a 294-kb region that includes four known genes. Using microarray-based analyses, we ruled out the presence of disease-linked copy number variants within the implicated region. Finally, we utilized *in silico* analysis and detected no strong evidence of regulatory regions surrounding the disease-associated variants. In conclusion, our finding of disease-specific variants occurring in complete linkage disequilibrium raises new insights and intriguing questions about the origin of the disease haplotype, the existence of phenocopies and of reduced penetrance, and the causative genetic alteration in XDP.

*European Journal of Human Genetics* (2015) 23, 1334–1340; doi:10.1038/ejhg.2014.292; published online 21 January 2015

## INTRODUCTION

Due to advances in sequencing technology, genetic variations causative for dystonia are being discovered more rapidly in recent years. Although a number have already been validated in independent studies, whether some putative disease-causing variants are truly specific has yet to be confirmed, and the challenge of proving causation remains for some diseases. X-linked dystonia-parkinsonism (XDP, DYT3, OMIM #314250) is a well-described X-linked recessive syndrome of combined dystonia-parkinsonism, with an underlying genetic cause that has not yet been unequivocally determined.<sup>1</sup> The condition is indigenous to Panay Island in the Philippines, and all cases described so far have been linked to Filipino ancestry, suggesting a single genetic founder and genetic homogeneity. Although extremely rare globally, the prevalence of XDP in the Philippines is 0.31 per 100 000; and in Panay Island, 5.74 per 100 000. Ninety-five percent of affected individuals are males; the average age is 44 years (20–70 years); and the average age at onset is 39 years (12–64 years).<sup>2</sup>

The DYT3 disease locus has been mapped to an ~427-kb region on Xq13.1 flanked by DXS10015 proximally and DXS559 distally,<sup>3</sup> however, previous studies did not reveal mutations in the coding regions of known genes inside this interval.<sup>3–6</sup> Instead, five disease-specific single-nucleotide changes (DSC1, DSC2, DSC3, DSC10, DSC12), one 48-bp deletion, and one 2627-bp SVA (SINE-VNTR-Alu element) retrotransposon insertion, all found either within ‘deep intronic’ or intergenic DNA segments, or in a nonconventional exon of the nearby *TAFI* gene, have been identified only in affected individuals.<sup>7,8</sup> None of these disease-associated genetic variants are located in protein-coding segments based on current annotations of the human genome, although the regulation of neuron-specific *TAFI* isoforms or of other genes in the region has been postulated as disease mechanisms.<sup>7–9</sup> Nevertheless, the exact disease-causing variant has been a matter of debate,<sup>9–12</sup> and theoretically, only one—if any—is functionally related to disease causation, whereas the remaining are merely benign variants.

<sup>1</sup>Institute of Neurogenetics, University of Lübeck, Lübeck, Germany; <sup>2</sup>XDP Study Group, Philippine Children’s Medical Center, Quezon City, Philippines; <sup>3</sup>Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Lübeck, Germany; <sup>4</sup>Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany; <sup>5</sup>Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany; <sup>6</sup>Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany; <sup>7</sup>Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany; <sup>8</sup>Faculty of Neurology and Psychiatry, University of Santo Tomas, Manila, Philippines; <sup>9</sup>Department of Neurosciences, University of the Philippines Manila—Philippine General Hospital, Manila, Philippines; <sup>10</sup>Institute of Human Genetics, National Institutes of Health—University of the Philippines Manila, Manila, Philippines; <sup>11</sup>Movement Disorders Section, Department of Neurology, Hannover Medical School, Hannover, Germany; <sup>12</sup>Section of Functional Genetics, Institute of Human Genetics, University of Lübeck, Lübeck, Germany; <sup>13</sup>School of Public Health, Faculty of Medicine, Imperial College, London, UK

<sup>14</sup>These authors contributed equally to this work.

\*Correspondence: Dr C Klein, Institute of Neurogenetics, University of Lübeck, Maria Goeppert Str. 1, 23562 Lübeck, Germany. Tel: +49 451 2903354; Fax: +49 451 2903355; E-mail: christine.klein@neuro.uni-luebeck.de

Received 3 July 2014; revised 17 November 2014; accepted 9 December 2014; published online 21 January 2015

As it is difficult to conceive that seven independent genetic variants arose *de novo* simultaneously in a single individual to produce the first XDP case (ie, the genetic founder), we pursued genotyping of all currently known genetic variants associated with XDP in a large group of patients and in a carefully chosen cohort of geographically and ethnically matched controls. Our aim was to detect recombination events in patients or single occurrences of any of the genetic variants in controls, which would inform us through segregation analysis which variant/s is/are truly specific to the phenotype. We took great effort to include informative controls fulfilling two criteria: (1) absence of family history of dystonia or parkinsonism; and (2) ancestry from Panay Island, as it was possible that in previous sequencing studies, mixed-race controls or Filipinos not from Panay Island were investigated. Thus, the great majority of our controls were offspring of the population in which the XDP founder also belonged, as these individuals were to our mind the most likely to harbor benign variants found also in the genomes of XDP patients.

Parallel to genotyping the known genetic variants, we used next-generation sequencing (NGS) to discover new potentially disease-causing single-nucleotide variants (SNV) within and outside the linked region on the X chromosome. We used non-segregating variants that we identified via genome sequencing to define the boundaries of the XDP haplotype and narrow the disease locus. Furthermore, we screened for copy number variants (CNV) within the implicated region and genome-wide using complementary microarray-based analyses. Finally, as none of the genetic variants are translated into protein, we used available online databases to investigate the possibility that the variants alter regulatory regions within the XDP locus.

## MATERIALS AND METHODS

### Samples

The patient group included 163 adult males (average age  $\pm$  SD:  $49 \pm 9.6$  years, range: 31–73 years) from the Philippines who were clinically diagnosed with XDP upon assessment by movement disorders specialists. In addition, three affected females were analyzed (average age: 62 years). We previously reported on the genetic underpinnings of the manifestation of XDP, an X-linked recessive syndrome, in two of these females, that is, extremely skewed X chromosome inactivation,<sup>13</sup> and Turner mosaicism.<sup>14</sup> From these patients, five males and one (homozygous) female were selected for genome sequencing. Apart from two samples being a mother and son pair, no other individuals selected for NGS were obviously related. Of note, the patients included in this study represent already a fourth of all cases ever documented in the Philippine XDP Study Group's registry.

The control cohort consisted of 452 neurologically normal individuals from Panay Island. An additional 21 controls composed of six sons of XDP males and 15 healthy Filipinos who do not come from Panay Island were included. Thus, in total, we investigated 473 controls representing 695 healthy X-chromosomes (251 male, 222 female, average age  $\pm$  SD:  $37 \pm 12.5$  years in males and  $38 \pm 13.5$  years in females). All subjects gave written informed consent prior to genetic analysis and institutional review boards from Germany and from the Philippines approved the study.

### Sanger sequencing and short tandem repeat (STR) polymorphism analysis

We genotyped samples via Sanger sequencing using an ABI3500XL Genetic Analyzer (Applied Biosystems, Darmstadt, Germany); the 48-bp deletion and the SVA insertion were detected via PCR amplification and analysis on agarose gels as previously described.<sup>15</sup> Haplotype analysis using STR markers within or surrounding the disease locus (DXS10015, DXS10016, DXS10017, DXS10018, and DXS559)<sup>7</sup> were analyzed on polyacrylamide gels or on an ABI3130XL Genetic Analyzer running GeneMapper 4.0 software (Applied Biosystems). Primer sequences and conditions are available upon request.

### Genome sequencing

One genome (G01) was sequenced by Complete Genomics (Mountain View, CA, USA). The five other genomes (G02–G06) were sequenced by Knome Inc. (Boston, MA, USA) on an Illumina platform. Annotation and filtering of variant calls from both service providers were performed using Annovar,<sup>16</sup> and using the following criteria to determine variants of interest: (1) rare variants (minor allele frequency  $< 0.01$  or undetermined) based on the 1000 Genomes Project (Annovar version 1000g2012apr), the Exome Variant Server (Annovar version 6500), or the dbSNP database (Build 137); (2) not within known regions of segmental duplication; and (3) not found in three in-house genomes from unrelated individuals without XDP. Variants of interest were submitted to the Leiden Open Variant Database for *TAF1* at <http://databases.lovd.nl/shared/genes/TAF1>, with screening number 0000019895.

### Microarray-based copy number analysis

To search for the presence of novel or disease-specific CNVs near the implicated XDP region and elsewhere, we subjected 121 individuals to microarray-based SNP genotyping using two different technologies: the Genome-wide Human SNP Array 6.0 (Affymetrix, Inc., Santa Clara, CA, USA), and the HumanOmniExpress-24 Beadchip Array (Illumina, Inc., San Diego, CA, USA). Detailed methods for both analyses can be found in the accompanying Supplementary Methods.

### Database search for regulatory regions

RegulomeDB (<http://regulome.stanford.edu/>)<sup>17</sup> was used to search for predicted regulatory elements within the XDP locus. We further annotated each DSC ( $\pm 100$  bases) using the UCSC Browser (<http://genome.ucsc.edu>). To predict regulatory regions within the SVA insertion, we first obtained the sequence from GenBank (AB191243.1) then used the BLAT function of the UCSC Browser to search for similar regions in the genome. Chromosomal regions with  $> 99\%$  sequence identity to the SVA insertion were then annotated using the UCSC Browser.

## RESULTS

### Sequencing of previously described genetic variants

We found that all disease-associated genetic variants, that is, the five DSCs, the 48-bp deletion, and the SVA insertion, completely cosegregate in a single haploblock (an 'XDP haplotype') in all but five XDP patients tested ('XDP haplotype-carriers': XDP001-XDP158, Table 1); in these samples, there was no disease-associated genetic variant that was observed without the co-occurrence of the others. However, five patients did not carry any of these genetic variants ('non-carriers': XDP159-163, clinical features described in Supplementary Table 1). The three affected females that we included also carried all the variants in homozygous (XDP164) or heterozygous (XDP165-166) states.

Analysis of STR polymorphisms revealed that the 158 XDP haplotype-carriers shared alleles at DXS10016, DXS10017, and DXS10018, except for differences accountable by 'slippage mutations'. DXS10015 is excluded proximally, and DXS559 distally, from this haploblock, subtending a 427.4-kb region. Notably, these markers were reported by Németh *et al.*<sup>3</sup> previously to include a 350-kb region, however, coordinates have since changed and using the current human genome assembly (GRCh38) calculates the included region to 427 kb.

Among controls, 2 of 473—one 35-year-old male and one female—carried the entire haploblock of XDP-associated genetic variants (C472–C473 in Table 1). At the time the male control was examined, he was 4 years below the average disease onset; it is unknown whether he has exhibited symptoms since then. We did not find any control carrying isolated or singly occurring disease-associated variants.

**Table 1 Genotypes at STR markers, previously described, and novel variants in the XDP-linked region in affected and control individuals**

Samples	rs41484056		rs41377154		rs41532445		DSC12		DSC10		DSC1		DSC3		DSC2		rs41438158	
	A>G	DXS10015	DXS10016	G>A	(novel)	G>A	T>G	SVA insertion	C>T	del	T>A	C>T	G>A	G>A	C>T	C>T	C>T	ChrX:71653235
XDP001 = G01	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP002 = G02	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP003 = G03	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP004-144 (140)	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
			129															242
XDP145	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	C	C	C
XDP146	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	C	C	C
XDP147	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	C	C	C
XDP148	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	C	C	C
XDP149	G	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	C	C	C
XDP150	A	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP151	A	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP152	A	137	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP153	G	141	127	A	Del	A	G	Ins	T	Del	A	T	A	A	T	T	T	T
XDP154	A	141	127	A	G	A	A	Ins	T	Del	A	T	A	A	T	T	T	T
XDP155	A	141	127	A	G	A	A	Ins	T	Del	A	T	A	A	T	T	T	T
XDP156	A	141	127	A	G	A	A	Ins	T	Del	A	T	A	A	T	T	T	T
XDP157	G	137	127	G	G	G	G	Het	C/T	Del	A	T	A	A	T	T	T	T
XDP158 (XXY)	A/G	137/149	127	G/G	G/G	G/G	G/G	Het	C/T	Het	T/A	C/T	G/A	G/A	C/T	C/T	C/T	238/230
XDP159 = G05	A	133	135	G	G	G	G	wt	C	wt	T	C	G	G	C	C	C	C
XDP160 = G06	A	144	127	G	G	G	G	wt	C	wt	T	C	G	G	C	C	C	C
XDP161	A	133	135	G	G	G	G	wt	C	wt	T	C	G	G	C	C	C	C
XDP162	A	144	127	G	G	G	G	wt	C	wt	T	C	G	G	C	C	C	C
XDP163	A	133	135	G	G	G	G	wt	C	wt	T	C	G	G	C	C	C	C
XDP164 = G04 (XX)	G/G	Hom	127	A/A	Del	A/A	A/A	Ins	T/T	Del	A/A	T/T	A/A	Hom	Hom	T/T	T/T	Hom
XDP165 (XX)	A/G	Het	Het	G/A	G/Del	G/A	G/A	Het	C/T	Het	T/A	C/T	G/A	Het	Het	C/T	C/T	Het
XDP166 (XX)	A/G	Het	Het	G/A	G/Del	G/A	G/A	Het	C/T	Het	T/A	C/T	G/A	Het	Het	C/T	C/T	Het
C001-C470 (470)	A	n.d.	n.d.	G	G	G	G	wt	C	wt	T	C	G	n.d.	n.d.	C	C	n.d.
C471*	A	n.d.	n.d.	G	Del	G	G	wt	C	wt	T	C	G	n.d.	n.d.	C	C	n.d.
C472 (XY)	G	n.d.	n.d.	A	Del	A	A	Ins	T	Del	A	T	A	n.d.	n.d.	T	T	n.d.
C473 (XX)	A/G	n.d.	n.d.	G/A	G/Del	G/A	G/A	Het	C/T	Het	T/A	C/T	G/A	n.d.	n.d.	C/T	C/T	n.d.

Abbreviation: Del, deletion; Het, heterozygous; Hom, homozygous; Ins, insertion; n.d., not done; wt, wild-type. Shaded cells represent linked loci based on segregation analysis in patients and genotyping of variants discovered via genome sequencing. Darker area represents the narrowed XDP haplotype.  
\*Son of an XDP patient. He is unaffected and inherited the deletion from his mother who was heterozygous in chrX:71301439delG. This in combination with five patients who are wt for the variant (XDP154-158) excludes this variant from being disease-causing.

### Genome sequencing and segregation analysis

We then selected three XDP haplotype-carriers (XDP001–003 = G01–G03) and two of the four non-carriers (XDP159–160 = G05–G06) for genome sequencing; we also included the homozygous female patient (XDP164 = G04). Genome sequences were obtained with mean coverages of 51x, 36x, 47x, 35x, 31x, and 39x (G01–G06) and with an average of 97% of known sites mapped. Average read depth for the X chromosome was 30x, and was 23x for the region between DXS10015 and DXS559. Variant search within the linked region in the four genomes carrying the XDP haplotype (G01–G04) revealed 15 variants of interest shared by at least three genomes (Table 2). Comparing G01–G04 with the two sequences not carrying the XDP haplotype (G05–G06), only rs4844149 (a repeat polymorphism) was found shared. Later validation by Sanger sequencing confirmed that this allele was also present in controls (5/43 X-chromosomes). Notably, the paucity of variants found within the linked region is similar to previous sequencing studies<sup>7,8</sup> which used traditional methods (ie, not NGS) to sequence this locus. Outside of the linked region, we found no exonic SNV located on the X chromosome that is shared between G01–G04 and either G05 or G06.

Out of the 15 variants shared by G01–G04 within the linked region, only two were completely novel (chrX.GRCh38:g.71301439delG, chrX.GRCh38:g.71653235C>T). The remaining variants included the five DSCs, rs4844149, and seven SNVs previously detected in other studies but have not yet been validated in controls. As we had already determined that the DSCs cosegregate in patients, we then identified six SNVs that flank the DSCs and genotyped these in all of our patient and control samples. The aim was to determine the boundaries of the XDP haplotype through recombination analysis. Non-segregating alleles in five patients establish the distal recombination event at rs41438158 (chrX.GRCh38:g.71633571C>T). In the proximal part of the linked region, alternative alleles at four variants place the boundary of the linked region downstream to rs41532445 (chrX.GRCh38:g.71339837G>A)(Table 1). Thus, we successfully narrowed the XDP disease locus to a 294.7-kb region (chrX.GRCh38:g.71339837\_71633571) by validating polymorphisms

discovered via genome sequencing in our large patient and control cohorts. This narrowed XDP haplotype subtends a region that includes the genes *TAF1*, *OGT*, *ACRC*, and *CXCR3* (Figure 1).

### Analysis of copy number variations

To screen for potentially disease-associated CNVs, we first subjected one XDP haplotype-carrier (G02), and one non-carrier (G05) to microarray-based genotyping using the Affymetrix 6.0 array. Although known gains around Xq21.31–21.32 were detected in both individuals, we detected no disease-specific CNV in either sample on chromosome X (Supplementary Figure 1) and genome-wide (data not shown). We then extended the search by genotyping 120 XDP haplotype-carriers using the Illumina OmniExpress array. This extended analysis revealed CNV events in 14 of 105 samples (merged into eight distinct CNV regions of which seven were deletions) (Supplementary Figure 2). All but one loci (ie, Xq21.1) detected in one sample had been reported previously, and the loci detected were only present in a small subset of the patients. Furthermore, no CNV loci were observed within the region of the XDP haplotype on the X chromosome (71.21–71.61Mb, GRCh38). Collectively, the results rule-out CNVs as the molecular cause underlying XDP in these individuals.

### Search for regulatory regions

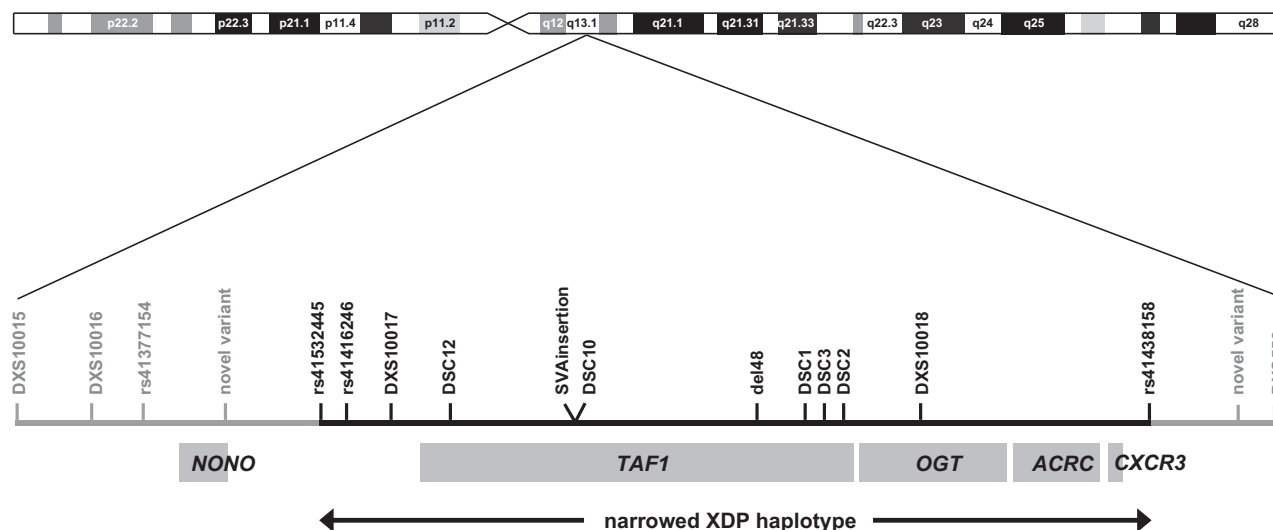
Finally, we used RegulomeDB to search for predicted regulatory elements within the XDP haplotype we have narrowed. This online tool computes a score predicting DNA-acting elements based on multiple public data sets and from the Encyclopedia of DNA Elements (ENCODE) Project. Regions surrounding the DSCs showed either no hit or a score of 6, equivalent to the lowest level of evidence/weak or no data supporting regulatory functions. The only SNV within the linked region with a reasonably high score (2b, evidence of transcription factor binding and a DNaseI footprint) was rs5981113, which is not among the DSCs, and is 10 000-bp away from the nearest XDP-specific genetic variant. We further annotated by hand each DSC using the UCSC Browser, which also collates predicted regulatory functions based on data from ENCODE. Among the DSCs, only the region

**Table 2** Variants in the linked region common among four genomes (G01–G04) sequenced

Position on ChrX	dbSNP ID	Reference	Variant	Alleles in genomes						Comments
				G01	G02	G03	G04	G05	G06	
71212570	rs41484056	A	G	G	G	G	G/G	A	A	
71219031	rs41483845	A	C	C	C	C	C/C	A	A	
71262608	rs41377154	G	A	A	A	A	A/A	G	G	
71268784	rs41339045	A	T	T	T	T	T/T	A	A	
71301439		G	Del	Del	G	G	G/G	G	G	Novel; Sanger sequencing later showed the deletion in G02, G03, and G04.
71339837	rs41532445	G	A	A	A	A	A/A	G	G	
71344555	rs41416246	C	T	T	T	T	T/T	C	C	
71391195		T	G	G	G	G	G/G	T	T	DSC12
71441489		C	T	C	T	T	T/T	C	C	DSC10; Sanger sequencing later showed T for G01.
71483780	rs4844149	C	A	C <sup>a</sup>	A	A	A/A	A	A	
71513660		T	A	A	A	A	A/A	T	T	DSC1
71529785		C	T	T	T	T	T/T	C	C	DSC3
71531433		G	A	A	A	A	A/A	G	G	DSC2
71633571	rs41438158	C	T	T	T	T	T/T	C	C	
71653235		C	T	T	T	T	T/T	C	C	Novel

G05 and G06 are genomes from patients who do not carry the XDP haplotype. We validated six variants using Sanger sequencing (shaded rows) in all other male patients and in all control samples. Positions are based on GRCh38.

<sup>a</sup>Possible genotyping error due to variant being in a repeat polymorphism; allele also found in controls.



**Figure 1** Narrowed XDP haplotype on chromosome X. Validation of variants discovered through genome sequencing enabled us to narrow the disease locus and define the exact boundaries of the XDP haplotype (from 427 kb between DXS10015 and DXS559 to 294 kb between rs41416246 and rs41438158, respectively), subtending a region that includes four known genes: *TAF1*, *OGT*, *ACRC*, and *CXCR3*.

surrounding *DSC1* harbors a DNaseI hypersensitivity site based on experiments done in a lymphoblastoid cell line. Otherwise, the loci of the different *DSCs* contain no chromatin immunoprecipitation and sequencing (ChIP-Seq) peaks, nor is there evidence of brain histone modification in these regions. There seem to be areas of CpG island methylation based on weak methylated DNA immunoprecipitation and sequencing (MeDIP-Seq) signals but in cell lines other than neuronal.

To predict regulatory regions within the SVA insertion, three chromosomal regions with >99% sequence similarity to the SVA insertion sequence were annotated (chr19.GRCh38:g.40107194\_40109853, chr11.GRCh38:g.64170742\_64173483, chr4.GRCh38:g.150658643\_150661524). Raw neuronal MeDIP-Seq signals were seen in all three regions, indicating sites of CpG island methylation.

## DISCUSSION

The combination of conventional sequencing, NGS, CNV analysis, and database search for regulatory functions that we performed in this study did not identify a novel, unique, and unambiguous genetic alteration that can be said to be the pathogenic mutation in XDP. Instead, what we observed is a disease-specific haplotype in almost all XDP patients, consisting of genetic variants occurring together in complete linkage disequilibrium. Thus, the first question raised by our findings is: how did these variants come about in the genetic founder?

It is difficult to formulate a scenario wherein seven distinct genetic variants arose simultaneously in the first XDP case. More plausibly, the original proband was a member of a healthy population within whose genome six of seven XDP-associated variants were already present, and that a subsequent, seventh variant was introduced into the genome *de novo*. However, among the large number of ethnically and geographically matched controls that we included in this study, we were not able to find any member or offspring of this pre-XDP population carrying only the original six genetic variants. As such, an intriguing hypothesis is that a partial haplotype was brought into Panay Island from another geographical region, and this, in combination with a *de novo* variant occurring locally later, gave rise to the

disease haplotype in the XDP founder. Interestingly, there is one well-known Filipino legend that tells the story of 10 foreigners from the island of Borneo (part of present-day Malaysia) who traveled to and eventually settled in Panay Island. Irregularities, however, have contested the accuracy of these oral histories, and they are currently not considered reliable from an anthropological point of view.<sup>18</sup> Nevertheless, we looked at publicly available genome sequences,<sup>19</sup> including those from a cohort of 100 Malaysians,<sup>20</sup> but found none of the XDP-associated genetic variants in these data sets, although one has to keep in mind the limited number of individuals investigated in these studies against the rarity of XDP-associated alleles.

Based on current knowledge, a likely explanation for the non-carriers of XDP haplotype is that they represent phenocopies. In two, we were able to search genome sequences for possible disease-causing variants in known genes for isolated dystonia (*THAP1*, *GNAL*, *TH/GCHI/SPR*), dystonia-parkinsonism syndromes (*ATP13A2*, *FBXO7*), and an adult-onset ataxia (*KCNC3*) that was previously described in a Filipino pedigree from Panay Island.<sup>21</sup> We found no changes that point to another genetic cause in either. Although XDP is at the moment considered to be due to a founder mutation, and all patients are expected to share the same genetic alteration, genetic heterogeneity has also been described in other forms of dystonia occurring in other homogeneous communities/genetic isolates. In Amish-Mennonites with primary dystonia in whom the founder mutation in *THAP1* has not been found, other variants in the same gene (but not the original indel), and in other genes for primary dystonia (*GNAL* and *TOR1A*) were later discovered to be causative for the phenotype.<sup>22</sup> Of note, in the original linkage study<sup>3</sup> and in the article that first described XDP-specific genetic variants,<sup>7</sup> there was also one patient who did not carry any of the genetic variants reported. This was, however, dismissed as a misdiagnosed case. An explanation for the unaffected male carrying the haplotype could be late age at onset, or reduced penetrance, although this is not supported by the pedigrees we have analyzed to date.

Alternatively, if one considers that the genetic variants we and others have reported are present in the great majority of patients but not in all (ie, non-carriers), while also considering that we have

detected the variants in controls, one arrives at the possibility of a yet undiscovered genetic variant that is truly specific for XDP. It is furthermore conceivable that the XDP haplotype and the variants within merely denote the location of this currently unknown genetic alteration. In this study, we used the complementary methods of NGS and microarray-based analyses to detect both SNVs and CNVs; however, other genetic alterations that escape our methods have not been completely ruled out, for example, translocations, chromosomal inversions, and mosaicism. Of note, one study has detected an 8.5-Mb inversion in Xq13.1,<sup>23</sup> that spans and includes the entirety of the XDP haplotype, including 57 genes (esv1584074, chrX.GRCh38:g.63249949\_71722146). Furthermore, a region in Xq13.1 centromeric to the XDP haplotype has been found to harbor breakpoints for chromosomal inversions associated with X-linked mental retardation.<sup>24</sup> Although no formal experiments have been done to detect inversions of and within the haplotype, Makino *et al.* previously used overlapping contigs and was able to obtain an entire sequence spanning 463 kb between *GJB1* (centromeric to DXS10015) and *CXCR3*,<sup>7</sup> likely indicating that no small inversions exist within. Southern blotting of *ACRC* also has not revealed rearrangements within this gene.<sup>5</sup> Interestingly, all five patients who are non-carriers of the disease-associated variants have the same alleles at the DXS10017 marker. Also, two non-carriers of the XDP haplotype share the distal STR marker (DXS10018) with haplotype-carriers.

It must be noted that all disease-associated genetic variants that others and we have identified so far lie in noncoding, deep intronic, or intergenic regions. In particular, the previously described variants are situated either within introns of *TAF1* (eg, the SVA retrotransposon insertion in intron 32), or in an untranslated region termed the 'multiple transcript system'.<sup>7</sup> In theory, though not protein-coding themselves, any of the genetic variants within the XDP haplotype could be altering the function of silencers or enhancers of protein-coding genes within or interacting with the region. Although sequence analyses of *OGT* have revealed no XDP-specific genetic variants,<sup>3</sup> this gene has been linked by functional analysis to DYT6, by means of a common THAP motif that associates with it.<sup>25</sup> Our own database search, however did not reveal strong evidence of regulatory elements among the DSCs, although this conclusion is highly limited by the paucity of studies available on neuronal tissue, and specifically targeting this region of the X chromosome. We also predicted regulatory regions within the SVA insertion and found possible sites of CpG island methylation in chromosomal regions with similarity to the SVA insertion sequence. One previous study likewise reports that the SVA insertion itself is hypermethylated in XDP patients and further hypothesizes that this status affects the function of a *cis*-regulatory element within the region.<sup>8</sup> However, the purported regulatory element has not been found as of yet, and its function unverified. Thus, studies characterizing the regulatory and functional consequences of all disease-specific variants within the XDP haplotype are warranted, and are part of our future investigations.

In summary, we describe here a narrowed disease-specific haplotype for XDP, inclusive of untranslated variants within introns or non-conventional exons of *TAF1*. These genetic variants occur in linkage disequilibrium and are found in almost all XDP patients in this study. Any alteration within the haplotype may be the disease-causing variant, but at the moment, the genetic cause of XDP is far from completely understood.

## CONFLICT OF INTEREST

CK is a medical counselor to Centogene, a commercial genetic testing center based in Rostock, Germany. The remaining authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank the Bachmann-Strauss Dystonia & Parkinson Foundation through a Jake's Ride for Dystonia Research Grant (to AW) and the Fritz Thyssen Foundation (to AW) for funding our experiments. AD is the recipient of a scholarship from the German Academic Exchange Program (DAAD). CK is a Schilling Professor of Neurogenetics (Herman and Lilly Schilling Foundation). We would like to thank the following members of the XDP Study Group of the Philippines who assisted in collecting XDP patients and assembling control cohorts: Dr Abundio Balgos, Dr Cid Czarina Diesta, Dr Criscely Go, Dr Arlene Ng, Dr Anthony Piano, Dr Marita VT Reyes, Dr Sonia Sarcia, and Mercy Joyce Monding-Yabres.

## ABBREVIATIONS

XDP, X-linked dystonia-parkinsonism; DSC, disease-specific single-nucleotide change; SNV, single-nucleotide variation; CNV, copy number variant; STR, Short-tandem repeat; NGS, next-generation sequencing, in this text refers to genome sequencing; *TAF1*, TATA box-binding protein-associated factor 1 gene; *OGT*, O-linked  $\beta$ -N-acetylglucosamine transferase gene; *ACRC*, acidic repeat-containing protein gene; *CXCR3*, chemokine receptor 3 gene; SVA, SINE-VNTR-Alu element.

- 1 Klein C: Genetics in dystonia. *Parkinsonism Relat Disord* 2014; **20**(Suppl 1): S137–S142.
- 2 Lee LV, Rivera C, Teleg RA *et al*: The unique phenomenology of sex-linked dystonia-parkinsonism (XDP, DYT3, "Lubag"). *Int J Neurosci* 2011; **121**(Suppl): 3–11.
- 3 Németh AH, Nolte D, Dunne E *et al*: Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). *Genomics* 1999; **60**: 320–329.
- 4 Peters U, Haberhausen G, Kostrzewa M: PFX1 and p54nrb: fine mapping, genomic structure, and exclusion as candidate genes of X-linked dystonia parkinsonism. *Hum Genet* 1997; **100**: 569–572.
- 5 Nolte D, Ramser J, Niemann S, Lehrach H, Sudbrak R, Müller U: ACRC codes for a novel nuclear protein with unusual acidic repeat tract and maps to DYT3 (dystonia parkinsonism) critical interval in Xq13.1. *Neurogenetics* 2001; **3**: 207–213.
- 6 Deng H, Le WD: Joseph Jankovic: Genetic study of an American family with DYT3 dystonia (lubag). *Neurosci Lett* 2008; **448**: 180–183.
- 7 Nolte D, Niemann S, Müller U: Specific sequence changes in multiple transcript system DYT3 are associated with X-linked dystonia parkinsonism. *Proc Natl Acad Sci USA* 2003; **100**: 10347–10352.
- 8 Makino S, Kaji R, Ando S *et al*: Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* 2007; **80**: 393–406.
- 9 Herzfeld T, Nolte D, Grznarova M, Hofmann A, Schultze JL, Müller U: 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* 2013; **22**: 941–951.
- 10 Müller U, Herzfeld T, Nolte D: The TAF1/DYT3 multiple transcript system in X-linked dystonia parkinsonism (letters to the editor). *Am J Hum Genet* 2007; **81**: 415–417.
- 11 Tamiya G, Makino S, Kaji R: TAF1 as the most plausible disease gene for XDP/DYT3 (letters to the editor). *Am J Hum Genet* 2007; **81**: 417–418.
- 12 Sako W, Morigaki R, Kaji R *et al*: Identification and localization of a neuron-specific isoform of TAF1 in rat brain: implications for neuropathology of DYT3 dystonia. *Neuroscience* 2011; **189**: 100–107.
- 13 Domingo A, Lee LV, Brüggemann N *et al*: Case report of a female with X-linked recessive dystonia-parkinsonism: a clue to the epidemiology of parkinsonism in Filipino women? *JAMA Neurol* 2014; **71**: 1177–1180.
- 14 Westenberger A, Rosales R, Heintz S *et al*: X-linked Dystonia-Parkinsonism manifesting in a female patient due to atypical turner syndrome. *Mov Disord* 2013; **28**: 675–678.
- 15 Kawarai T, Pasco PMD, Teleg RA *et al*: Application of long-range polymerase chain reaction in the diagnosis of X-linked dystonia-parkinsonism. *Neurogenetics* 2013; **14**: 167–169.
- 16 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nuc Acids Res* 2010; **38**: e164.

- 17 Boyle AP, Hong EL, Hariharan M *et al*: Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–1797.
- 18 Scott WH: Cracks in the parchment curtain and other essays in Philippine history. Manila, Quezon City: New Day Publishers, 1982. 300 pp.
- 19 Drmanac R, Sparks AB, Callow MJ *et al*: Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 2010; **327**: 78–81.
- 20 Wong LP, Ong RTH, Poh WT *et al*: Deep whole-genome sequencing of 100 southeast Asian Malays. *Am J Hum Genet* 2013; **92**: 52–66.
- 21 Subramony SH, Advincula J, Perlman S *et al*: Comprehensive phenotype of the p.Arg420his allelic form of spinocerebellar ataxia type 13. *Cerebellum* 2013; **12**: 932–936.
- 22 Saunders-Pullman R, Fuchs T *et al*: Heterogeneity in primary dystonia: lessons from THAP1, GNAL, and TOR1A in Amish-Mennonites. *Mov Disord* 2014; **29**: 812–818.
- 23 Mazars R, Gonzalez-de-Peredo A, Cayrol C *et al*: The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. *J Biol Chem* 2010; **285**: 13364–13371.
- 24 Levy S, Sutton G, Ng P *et al*: The diploid genome sequence of an individual human. *PLoS Biol* 2007; **5**: e254.
- 25 Villard L, Briault S, Am Lossi *et al*: 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* 1999; **36**: 754–758.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)