Published in final edited form as:

Hum Mutat. 2020 January ; 41(1): 150–168. doi:10.1002/humu.23902.

Xq22 deletions and correlation with distinct neurological disease traits in females: further evidence for a contiguous gene syndrome

Hadia Hijazi¹, Fernanda S. Coelho^{2,3}, Claudia Gonzaga-Jauregui⁴, Laura Bernardini⁵, Soe S. Mar⁶, Melanie A. Manning^{7,8}, Andrea Hanson-Kahn^{9,10}, Sakku Bai Naidu^{11,12}, Siddharth Srivastava¹³, Jennifer A. Lee¹⁴, Julie R. Jones¹⁴, Michael J. Friez¹⁴, Thomas Alberico¹⁵, Barbara Torres⁵, Ping Fang¹⁶, Sau Wai Cheung^{1,17}, Xiaofei Song¹, Angelique Davis-Williams¹⁵, Carly Jornlin¹⁵, Patricia A. Wight¹⁸, Pankaj Patyal¹⁸, Jennifer Taube¹⁵, Andrea Poretti¹¹, Ken Inoue¹⁹, Feng Zhang²⁰, Davut Pehlivan^{1,21}, Claudia M.B. Carvalho¹, Grace M. Hobson^{15,*}, James R. Lupski^{1,22,23,24,*}

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

Corresponding authors: James R. Lupski, M.D., Ph.D., D.Sc. (hon), FAAP, FACMG, FANA, FAAAS, FAAS, Department of Molecular and Human Genetics, Baylor College of Medicine, 1 Baylor Plaza, Room 604B, Houston, TX 77030, jlupski@bcm.edu, Tel: (713) 798-6530, Fax: (713) 798-5073; Grace M. Hobson, PhD, Nemours Biomedical Research, Nemours/Alfred I. duPont Hospital for Children, 1600 Rockland Road, RC1, Wilmington, DE 19803, USA, Tel: +1-302-838-1771, ghobson@nemours.org. Authors' Contributions

HH, CMBC, GMH, and JRL participated in the design of the study. HH performed the majority of data analyses and experiments. HH, CMBC, DP, and GH prepared the manuscript with JRL. FSC participated in the design of the study and performed some experiments and analyses. CGJ developed the HSIR computational method while at Baylor College of Medicine. LB referred a patient to this study, provided their clinical information, and performed the XCI studies on the referred case. SSM, MAM and AHK referred patients to this study and provided their clinical information. JL participated in the design of the study, performed some experiments, and together with SWC and PF facilitated anonymized sample studies and performed the XCI studies on three of the female samples. SN and SS referred a patient to this study and provided their clinical information. PP and PW performed some experiments and analyses. TA and KS performed molecular studies on one sample. JRJ and MF performed molecular studies and XCI on one samples and participated in data interpretation. JT performed in silico analyses. AP performed and interpreted MRI on one patient. XS performed the microhomeology and breakpoint-junction sequence-similarity analyses. ADW, and CJ performed molecular experiments on two samples. KI reviewed the clinical information and helped with clinical interpretation. FZ participated in the design of the study and performed some experiments. GMH provided samples and data for this study, participated in data interpretation and critically revised the manuscript. DP reviewed and analyzed the clinical information, prepared the clinical table and critically revised the manuscript. CMBC performed experiments, participated in data interpretation and critically revised the manuscript. JRL participated in data analysis and interpretation, oversaw the manuscript preparation, and critically revised the manuscript. All authors approved the final submitted version of the manuscript, and are accountable for the accuracy and integrity of the work.

*James R. Lupski and Grace M. Hobson should be considered joint senior authors

Declarations

Competing Interests

Baylor College of Medicine (BCM) and Miraca Holdings have formed a joint venture with shared ownership and governance of the Baylor Genetics (BG), which performs clinical microarray analysis and clinical exome sequencing. CG-J is currently a full-time employee of Regeneron Pharmaceuticals Inc. and receives stock options as part of compensation. J.R.L. serves on the Scientific Advisory Board of the BG. J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from molecular genetic testing offered in the Baylor Genetics Laboratories. Nemours Biomedical Research derives revenue from molecular genetic testing offered in the Nemours Molecular Diagnostics Laboratory. The Greenwood Genetic Center receives revenue from diagnostic testing performed in the GGC Molecular Diagnostic Laboratory. The other authors declare no competing financial interests.

Availability of data and material

Data generated during this study and the datasets supporting the conclusions of this article are included within the article and its additional files. The GTEx data used for the analyses described in this manuscript were obtained from: Gene expression, GTEx Batch Gene Query, the [GTEx Portal](https://www.gtexportal.org/) ("GTEx Portal - The Broad Institute of MIT and Harvard," n.d.) on 11/11/18.

- ²Programa de Pós-Graduação em Genética Departamento de Biologia Geral – UFMG, Belo Horizonte, MinasGerais, Brazil
- ³Instituto René Rachou – FIOCRUZ, Belo Horizonte, MG, Brazil
- ⁴Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY
- ⁵Cytogenetics Unit, “Casa Sollievo della Sofferenza” Foundation, San Giovanni Rotondo (FG), Italy
- ⁶Department of Neurology, Washington University School of Medicine, St. Louis, MO
- ⁷Stanford University School of Medicine, Department of Pediatrics, Division of Medical Genetics, Palo Alto, CA
- ⁸Stanford University School of Medicine, Department of Pathology, Palo Alto, CA
- ⁹Stanford University School of Medicine, Department of Genetics, Palo Alto, CA
- ¹⁰Lucile Packard Children’s Hospital, Department of Pediatrics, Division of Medical Genetics, Palo Alto, CA
- ¹¹Departments of Neurology and Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD
- ¹²Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD
- ¹³Department of Neurology, Boston Children’s Hospital, Boston, MA
- ¹⁴Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC
- ¹⁵Nemours Biomedical Research, Nemours/Alfred I. duPont Hospital for Children, Wilmington, DE
- ¹⁶WuXi NextCODE, Cambridge, MA
- ¹⁷Baylor Genetics, Houston, TX
- ¹⁸Department of Physiology and Biophysics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR
- ¹⁹Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan
- ²⁰Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering at School of Life Sciences, Institute of Reproduction and Development, Fudan University, Shanghai 200011, China
- ²¹Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, TX
- ²²Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX
- ²³Department of Pediatrics, Baylor College of Medicine, Houston, TX
- ²⁴Texas Children’s Hospital, Houston, TX

Abstract

Xq22 deletions that encompass *PLP1* (Xq22-*PLP1*-DEL) are notable for variable expressivity of neurological disease traits in females ranging from a mild late-onset form of Spastic Paraplegia type 2 [MIM# 312920], sometimes associated with skewed X-inactivation, to an early-onset neurological disease trait (EONDT) of severe developmental delay, intellectual disability and behavioral abnormalities. Size and gene content of Xq22-*PLP1*-DEL vary and were proposed as potential molecular etiologies underlying variable expressivity in carrier females where two smallest regions of overlap (SROs) were suggested to influence disease. We ascertained a cohort of eight unrelated patients harboring Xq22-*PLP1*-DEL and performed high-density array Comparative Genomic Hybridization (aCGH) and breakpoint-junction sequencing. Molecular characterization of Xq22-*PLP1*-DEL from 17 cases (8 herein and 9 published) revealed an overrepresentation of breakpoints that reside within repeats (11/17, ~65%) and the clustering of ~47% of proximal breakpoints in a genomic instability hotspot with characteristic non-B DNA density. These findings implicate a potential role for genomic architecture in stimulating the formation of Xq22-*PLP1*-DEL. The correlation of Xq22-*PLP1*-DEL gene content with neurological disease trait in female cases enabled refinement of the associated SROs to a single genomic interval containing six genes. Our data support the hypothesis that genes contiguous to *PLP1* contribute to EONDT.

Keywords

BEX3; RAB40A; TCEAL4; TCEAL3; TCEAL1; MORF4L2

Background

Xq22 encompasses an approximately 10 Mb genomic interval on the X chromosome in which ~100 annotated genes map; *PLP1* is one of the 14 disease-associated genes mapping to this interval. It is a dosage-sensitive gene as evidenced by both copy-number gains and copy-number losses resulting in disease traits (Inoue, 2005, 2017; Torisu et al., 2012; Woodward et al., 2005). In addition to copy number variants (CNV), rare point variants in *PLP1*, including nucleotide substitutions, small insertions or deletions (indels) and regulatory region alterations, have also been associated with neurological disease (Inoue, 2005, 2017; Lee, Madrid, et al., 2006; Osorio & Goldman, 2018). Duplication of *PLP1* is the most commonly observed variant allele in clinically ascertained cases and it is usually associated with Pelizaeus-Merzbacher Disease (PMD) [MIM# 312080], a dysmyelinating leukodystrophy affecting the central nervous system with characteristic clinical features including nystagmus, spastic quadriplegia, ataxia, and developmental delay, in males. In contrast to duplications, complete or partial deletion of *PLP1* is the rarest mutational event being observed in less than 5% of clinically ascertained cases with *PLP1* variants.

Hemizygous Xq22 deletions encompassing *PLP1* are typically small in size (<600 kb) relative to other *PLP1* CNVs, such as heterozygous deletions and duplications, and they can be associated with PMD or with Spastic Paraplegia type 2 (SPG2) [MIM# 312920] a progressive lower extremity spasticity disease. Hemizygous deletion, in addition to central nervous system disease, can also involve peripheral neuropathy in some cases (Garbern et al., 1997; Inoue, 2005, 2017). In the heterozygous state, deletions are notable for extensive

variability in their size (range: 33 kb - 5 Mb) and the associated phenotypes are characterized by variable expressivity of neurological disease ranging from a mild age-dependent penetrant form of SPG2 to a severe early-onset neurological disease trait (defined here as EONDT) (Inoue et al., 2002; Matsufuji et al., 2013; Raskind, Williams, Hudson, & Bird, 1991; Torisu et al., 2012; Yamamoto et al., 2014). The associated EONDT is characterized by severe infancy-onset developmental delay/intellectual disability (DD/ID), behavioral abnormalities, hypotonia, strabismus and potentially an emerging recognizable pattern of craniofacial dysmorphology (Yamamoto et al., 2014).

To date, nine non-recurrent Xq22 deletions encompassing *PLP1* (Xq22-*PLP1* deletions) have been reported, of which five were identified in hemizygous male probands and their less severely affected heterozygous female relatives, while three were identified in heterozygous female probands with EONDT and one was identified in a heterozygous female proband with an unclassified severe DD/ID phenotype (Inoue et al., 2002; Matsufuji et al., 2013; Raskind et al., 1991; Torisu et al., 2012; Yamamoto et al., 2014). Notably, none of the published families with EONDT affected females have reported male carriers of the same Xq22-*PLP1* deletion. The increased severity and distinctive disease nature observed in EONDT female cases, as opposed to the mild SPG2 manifestation in affected female relatives of male probands, were attributed to the unique gene content of large Xq22 deletions found exclusively in female cases. Two smallest regions of overlap (SRO) within large deletions were suggested to underlie the disease manifestation, chrX:102,233,526-102,957,289 (~730 kb) and chrX:102,993,719-103,982,269 (~990 kb) (Yamamoto et al., 2014). The study of additional rare Xq22 deletion CNV in unrelated families, and EONDT cases in females, may facilitate further understanding of this disease-gene locus and disease-trait manifestation as well as provide insight into sex influenced neurodevelopmental disease expression.

SV mutagenesis mechanistic studies of Xq22 duplication and triplication CNVs suggest a role for Low-Copy Repeats (LCRs) in stimulating rearrangement formation via replicative repair mechanisms (Beck et al., 2015; Carvalho et al., 2012; Carvalho et al., 2011; Lee, Carvalho, & Lupski, 2007; L. Zhang et al., 2017). Genomic studies of five Xq22-*PLP1* deletions did not reveal evidence for the potential involvement of LCRs or other paralogous genomic structures that might render genomic instability (Inoue et al., 2002; Matsufuji et al., 2013; Torisu et al., 2012). From breakpoint-junction studies, microhomology was the only persistent finding at Xq22-*PLP1* deletions junctions allowing one to surmise potential nonhomologous recombination mechanism as contributing to deletion formation (Inoue et al., 2002). Although previous molecular studies of Xq22-*PLP1* gains, duplications and triplications, have provided data revealing *PLP1* is a dosage-sensitive gene and some insight into genomic instability and rearrangement mechanism, more molecular and clinical data from rare Xq22-*PLP1* deletion cases in both males and females are warranted to infer potential upstream mechanisms for Xq22 deletion CNV formation and downstream genic mechanism(s) mediating trait manifestation.

We investigated eight unrelated patients carrying Xq22 deletions that partially or fully encompass *PLP1* (5 female probands, and 3 male probands). We describe the clinical details of affected females manifesting EONDT and examine potential genotype/phenotype

correlation by mapping the breakpoint-junction and studying the size, extent and gene content of their deletions. At the molecular level, we show the results of an analysis of the genomic features from 17 individuals, including nine published cases, carrying Xq22-*PLP1* deletions. We computationally and experimentally define a novel class of intrachromosomal repeats and provide evidence for intrinsic genome architecture fomenting genomic instability. These data implicate the involvement of large inverted repeat structures and short tandem repeats in the formation of Xq22-*PLP1* deletions. Our genotype/phenotype correlation findings support the contention that deletion gene content may provide prognostic information in female cases.

Materials and Methods

Editorial Policies and Ethical Considerations

The clinical information for families HOU1008, HOU1025, and HOU2971 was ascertained and studied using protocol H-16213 approved by the Institutional Review Board for research involving human subjects at Baylor College of Medicine while the information for families HOU3003, HOU3004 and HOU4615 was ascertained using protocol IRB#2002-142 approved by the Institutional Review Board for research involving human subjects at Nemours/Alfred I. duPont Hospital for Children. Cases BAB2615 and BAB2614 were ascertained through studies of anonymized data from clinical microarray CMA screening at Baylor Genetics (Institutional Review Board approved protocol H-36612). Informed consent for inclusion in the study was obtained from all adult participants and from parents/guardians of minors. This study was conducted in accordance with the Declaration of Helsinki.

Subjects

Eight cases were investigated, five of which were unrelated affected female probands, BAB2595/FamilyHOU1008, BAB2614, BAB2650/FamilyHOU1025, BAB8120/FamilyHOU2971, BAB12522/FamilyHOU4615 while three were unrelated affected male probands, BAB2615, BAB8201/FamilyHOU3003, and BAB8204/FamilyHOU3004 (Figure 1). Clinical information was available for six of the eight cases (4 females and 2 males), BAB2595, BAB2650, BAB8120, BAB8201, BAB8204 and BAB12522. Genomic DNA for all experiments was isolated from blood according to standard procedures.

Array Comparative Genomic Hybridization (aCGH)

Two tiling-path high density custom oligonucleotide microarrays were designed. The first interrogates the copy-number of a 15.5 Mb segment encompassing *PLP1* on chromosome Xq22. A total of 40,208 60-mer oligonucleotide probes spanning chrX: 97,915,511-113,400,000 (NCBI build 36) were selected with an average distribution of 1 probe per 386 bp (format 4 X 44K) using the Agilent eArray website (<https://earray.chem.agilent.com/suredesign/>) (Carvalho et al., 2012). The second interrogates the copy-number of an ~10 Mb segment encompassing *PLP1* on chromosome Xq22 with a total of 24,643 60-mer probes spanning chrX:97,835,000-107,855,000 (NCBI build 35) (format 4X 44K) (Lee et al., 2007). The experimental procedures were performed according to the manufacturer's protocol (Agilent Oligonucleotide Array-Based CGH for Genomic DNA

Analysis, Version 7.2, Agilent Technologies) with some modifications as described (Carvalho et al., 2011; Carvalho et al., 2009). Sex-matched controls (GM10851 as male control and GM15510 as female control, both from Coriell Institute) were used for all samples. Genomic copy number at a given locus, or map position as evidenced by interrogating oligonucleotide probe, was defined by analysis of the normalized $\log_2$ (Cy5/Cy3) ratio average of the CGH signal. Agilent Feature Extraction software (version10) and Agilent Genomic Workbench (version 7.0.4.0) were used to process scanned array images and analyze extracted files, respectively. Average $\log_2$ ratio thresholds of -1 and -4 were considered as heterozygous female and hemizygous male deletions, respectively.

PLP1 Sequencing

Coding exons and intron 3 of *PLP1* (NM_000533.4) were amplified via PCR in all five female cases harboring heterozygous *PLP1* deletion. PCR products were purified and analyzed by Sanger di-deoxy DNA sequencing (primer sequences are provided in Supp. Table S1). Variants with minor allele frequency (MAF) of 0.001 in the 1000 Genomes Project database and the genome aggregation database (gnomAD) were considered rare (Genomes Project et al., 2015; Karczewski et al., 2019).

Mapping Breakpoints and Characterizing Microhomology/Microhomeology

Deletion breakpoint-junctions, or ‘join-points’ where two discontinuous sequences (i.e., rearrangement substrate sequences) in the human reference genome are joined during the rearrangement process, were amplified for all 8 cases, using inward-facing primers for PCR amplification, and Sanger sequenced. Primers were designed at the apparent boundaries of each deleted segment as defined by transitions from copy-number neutral to copy-number loss by aCGH analysis (primer sequences are provided in Supp. Table S1). Junction sequences were screened for single nucleotide variants (SNVs) flanking the junction. The identified Xq22-*PLP1* deletions and SNVs detected in their junction sequences were deposited in the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

The two reference genome loci to which the proximal and distal breakpoints mapped were extended and aligned to examine for microhomology/microhomeology at the join-points and sequence similarity flanking the join-points. Nucleotide identity of 100% between the proximal and distal reference strands leading to the junction was considered microhomology when found in a subtractive state in the patient’s DNA sequence (i.e. identical sequence found once only in the patient’s DNA at the deletion junction) (Carvalho & Lupski, 2016). Microhomeology was defined as imperfect matches leading to the junction (cutoff of 70% identity with a maximum two-nucleotides gap followed by at least two perfectly matched nucleotides) (adapted from Beck et al., 2019; Liu et al., 2017). Microhomeology at the breakpoint-junctions was recently reported as a feature associated with replication-based mechanism of SV formation (Bahrambeigi et al., 2019; Liu et al., 2017). We also analyzed the similarity of DNA sequences surrounding breakpoint-junctions by alignment and similarity computation using the R Biostrings package in samples that had < 220 bp of microhomeology on either side of the junction (adapted from Gu et al., 2015); 150 nucleotides of the reference sequence was obtained on either end of a breakpoint-junction and the two 300-nucleotide reference sequences were then centralized at the junction-

microhomology or blunt end and aligned together using the Needleman-Wunsch algorithm Biostrings package. Sequence similarity was calculated within a 20 bp moving window as the percentage of aligned bases over the total count of non-gap sequences, for which orientation relies on the alignment with sample sequence. A heatmap was plotted for each event.

Identifying and Analyzing Highly-Similar Intrachromosomal Repeats (HSIR)

Computational analyses of the haploid reference human genome were performed to identify highly-similar repeated sequences across the human genome that may contribute to local genomic instability. Our analyses aimed to identify all >700 bp intrachromosomal repeat sequences that are highly similar (95%-100% identical) in the human genome reference assembly (GRCh37/Hg19). We used the LASTZ (Local Alignment Search Tool, blastZ-like) algorithm (release 1.02.00) that performs pairwise alignment of large DNA sequences to identify highly-similar intrachromosomal sequence comparisons (<http://www.bx.psu.edu/~rsharris/lastz/>). This approach is distinct from other methods of repeat structure detection such as utilized for PMD-LCRs (Lee, Inoue, et al., 2006), Segmental Dups (Bailey et al., 2002; Bailey, Yavor, Massa, Trask, & Eichler, 2001), Self Chain (Chiaromonte, Yap, & Miller, 2002; Kent, Baertsch, Hinrichs, Miller, & Haussler, 2003; Schwartz et al., 2003) and Repeat Masker (Jurka, 2000) in that it: 1) detects repeats by intrachromosomal pairwise alignment (i.e. self-comparison of each chromosome in the reference assembly), 2) imposes no size restriction on detected repeat fragments, 3) allows 95%-100% identity levels between pairs, 4) involves no filtering or masking of repetitive elements and common repeats, and 5) imposes no distance restrictions between matching pairs. Using this approach, we identified a total of 58,469 highly-similar intrachromosomal repeat (HSIR) pairs with a median size of 6 kb (minimal=793bp; maximal=406,322bp) and an average identity of 96.8%; of these pairs, 29,503 are in direct orientation and 28,966 are in inverted orientation in the hg19 reference assembly. The HSIR dataset is deposited as Additional file 1 and 2, direct and inverted pairs, respectively.

We then further investigated HSIRs for coinciding deletion breakpoints (base pair-level resolution of breakpoint coordinates were used when available or aCGH-derived approximated coordinates depending on availability of the junction sequence); a total of six HSIRs were identified coincident with deletion breakpoints and these HSIRs were annotated with respect to four repeat datasets: 1) the PMD-LCRs dataset for LCRs at the *PLP1* locus (chrX:98,551,855-106,551,854, coordinates lifted to the GRCh37/hg19 build), 2) the Segmental Dups dataset for genome-wide identified paralogous repeats, 3) the SelfChain dataset for gapped genome self-alignments and 4) the RepeatMasker dataset for SINEs, LINEs, *Alu* elements, and other repetitive elements in the genome.

Breakpoints were also examined for repeat structures that were not part of the HSIR dataset but were found in any one of the other four repeat datasets, i.e. the PMD-LCRs dataset, the Segmental Dups dataset, the SelfChain dataset, and the RepeatMasker dataset. Details of breakpoint-coinciding HSIRs and other breakpoint-coinciding repeat structures are provided in Supp. Table S2.

Non-B DNA Motifs

Three test/control groups of genomic intervals near the cluster of Xq22-*PLP1* deletion breakpoints were examined for the presence of seven non-B DNA motifs via the publicly available non-B database (non-B DB) and its associated tools (Cer et al., 2013) (<https://nonb-abcc.ncifcrf.gov/apps/nBMST/default/>). The three tested groups were: 1) the reference sequence embedded within the proximal and distal halves of the HSIR RepX-i1010 (chrX:102,874,647-102,944,647 and chrX:102,944,648-103,014,647), 2) the reference sequence embedded within 140 kb of RepX-i1010 (chrX:102,874,647-103,014,647) as well as two similarly sized control fragments flanking the start and end of this repeat structure region (chrX:102,734,646-102,874,646 and chrX:103,014,648-103,154,648), and 3) the reference sequence embedded within the identified 90 kb genomic instability hotspot (chrX:102,954,647-103,044,647), as well as two similarly-sized (90 kb) control fragments flanking the start and end of this hotspot region. The distribution of a motif in any one of the investigated regions was defined as the number of motif start sites found in 10 kb moving frames within the region; the 7-frames distribution of the distal half of RepX-i1010 was compared to the 7-frames distribution of the proximal half, the 14-frames distribution of RepX-i1010 was compared to the two sets of 14-frames flanking control distributions, and the 9-frames distribution of the hotspot was compared to the two sets of 9-frames flanking control distributions. Statistical conclusions were based on the results obtained from the Wilcoxon Rank Sum double-sided or one-sided test considering non-normal distributions. A heatmap plot was generated via the heatmap.2 function from the CRAN gplot R package for quantitative visualization of the three distributions.

X-Chromosome Inactivation (XCI)

Genomic DNA isolated from blood was used for implementing XCI studies according to an established protocol with modification (Allen, Zoghbi, Moseley, Rosenblatt, & Belmont, 1992). This protocol is designed to detect methylated *HpaII* and *HhaI* sites near the polymorphic CAG repeat in the human androgen receptor gene, *AR* [MIM# 313700]. Sequences were analyzed using GeneMapper v3.7 (Applied Biosystems) software.

Glossary of abbreviations is provided in Table 1.

Results

Clinical Findings in the Investigated Subjects

Four female (BAB2595, BAB2650 and BAB8120, BAB12522) and 2 male (BAB8201 and BAB8204) probands were examined clinically; pedigrees shown in Figure 1. Clinical features are summarized in Table 2 and further detailed information is provided in Additional file 3. All female patients had clinical features consistent with the early onset neurological disease trait reported by Yamamoto et al. (2014) and defined here as EONDT including hypotonia at birth, severe DD/ID, and neurobehavioral abnormalities; BAB2595, BAB8120 and BAB12522 had autistic features and BAB2650 had anxiety. A facial photograph of one subject (BAB2650) has been published (Brender, Wallerstein, Sum, & Wallerstein, 2015) and shows dysmorphic features concordant with what has been observed in three female patients reported by Yamamoto et al. (2014) including the triangular face,

broad forehead, strabismus and prominent jaw; however, another subject (BAB12522) did not have dysmorphic features. Brain MRIs revealed different degrees of increased T2 signal in the white matter of all four female subjects; BAB2650 and BAB12522 were noted to have a thin corpus callosum.

The two male patients (BAB8201 and BAB8204) presented with DD and spasticity consistent with SPG2. Additionally, BAB8201 had clinical seizures responsive to carbamazepine while BAB8204 had an abnormal EEG. Brain MRI was available on BAB8201 in which diffuse hyper-intensity in the white matter on T2-weighted images suggestive of hypomyelination, thin corpus callosum and reduced brain volume were detected. Clinical information from the family of BAB8201 with positive family history was published separately (Hisama et al., 2001).

All six patients had strabismus while BAB2650 was also noted to have had nystagmus. Five out of the six patients (BAB2650, BAB8120, BAB8201, BAB8204 and BAB12522) had gastrointestinal issues including gastrointestinal reflux disorder (GERD) and poor weight gain. Abnormal brain auditory evoked potential (BAEP) was reported in one female case, BAB8120, and in one male case, BAB8201. Information on BAEP testing in the second male patient, BAB8204, was lacking.

Size, Genomic Extent, and Gene Content of Xq22 Deletions

Deletions were detected by aCGH and the exact coordinates for breakpoints were determined (Figure 2 and Table 3). The size range in males was determined to be from 6.7 kb to 854 kb (3 male probands) and in females was from 693 kb to 5.6 Mb (5 female probands) (Table 3). Deletion alignment to the haploid reference genome and mapping of selected genes within the deletion intervals are shown in Figure 3A. Six of the deletions include the entire *PLP1* gene (NM_000533.4) while two (in unrelated males, BAB8201 and BAB8204) encompass the first exon and part of intron 1 only (Figure 3B). BAB8201 harbors the smallest of all *PLP1* deletions to date (6.7 kb) involving no consensus gene transcripts other than part of *PLP1*, while BAB2615 harbors the largest hemizygous deletion (854 kb) in which 23 genes in addition to *PLP1* are deleted; nine of the deleted protein-coding genes appear to be highly expressed in the brain as found on the GTEx portal (<https://gtexportal.org/home/>) and as shown in Supp. Figure S1. Female subject BAB8120 harbors the largest heterozygous deletion (5.6 Mb) with 63 genes deleted in addition to *PLP1*, one of which is a gene associated with primary ciliary dyskinesia, *PIH1D3* [MIM# 300933] (Figure 3A). Physical examination of this patient at 3.5 years of age revealed no clinical evidence for a ciliopathy phenotype that could potentially result from haploinsufficiency of *PIH1D3*. All five heterozygous deletions (BAB2595, BAB2650, BAB8120, BAB2614 and BAB12522) include *GLRA4* that has been previously associated with intellectual disability (Labonne et al., 2016) and notably they all extend further centromeric and include a gene that is differentially expressed in the brain, *BEX3* [MIM# 300361] (also known as *NGFRAP1*) and that has been suggested to contribute to disease (Yamamoto et al., 2014) (Figure 3A).

Genotype-Phenotype Correlations in Females

We examined for unique genotypic aberrations in EONDT females that could potentially contribute to the distinctive nature and increased severity of their disease as opposed to females deleted for *PLP1* that manifest a neurological phenotype consistent with SPG2. We initially screened for rare (i.e. MAF of 0.001 in the 1000 Genomes Project database and the gnomAD database), likely damaging SNVs in the non-deleted *PLP1* allele in all five females via PCR and Sanger sequencing of coding exons and of intron 3. We aimed to explore the possibility of biallelic *PLP1* variation, however no evidence was found for rare and likely damaging variants in the examined females with heterozygous Xq22-*PLP1*-DEL (data not shown).

Next, we examined gene content among all 14 known heterozygous-*PLP1*-deletion cases with clinical phenotypes available and attempted to elucidate an SRO that is explicitly shared between EONDT cases, three published (Yamamoto et al., 2014) and four herein, while excluded from the deletions of all known female SPG2 cases. Of note, Yamamoto et al. (2014) reported four Xq22-*PLP1* deletion cases, the fourth case, Patient 5, harbored a small (85 kb) deletion that completely overlapped with the deletion of an SPG2 family (family of BAB1684 from Inoue et al. (2002)) whose affected individuals (male and female) did not show features of EONDT. The original authors proposed that the distinct phenotype in Patient 5 cannot be explained by the Xq22 deletion and that it is likely due to a pathogenic variant that cannot be detected by aCGH such as copy-number neutral genomic rearrangements or Indels/SNVs (Yamamoto et al., 2014). Hence the Xq22 deletion of Patient 5 was excluded from our SRO analysis, as it is not expected to encompass genes that could potentially contribute to EONDT features.

We found an SRO upstream of *PLP1* within a larger region that was previously suggested to contribute to disease in EONDT females (Yamamoto et al., 2014) and we referred to this region as the EONDT-SRO, chrX:102,615,641-102,957,288 (Figure 3A, green magnification). This region is ~342 kb in size and contains six genes: *BEX3*, *RAB40A*, *TCEAL4*, *TCEAL3*, *TCEAL1*, and *MORF4L2*, all of which are expressed in the brain albeit at different levels (Figure 3C) (data sourced from the GTEx portal). These results parsimoniously enabled the possible exclusion of the second previously proposed SRO (downstream of *PLP1* and containing *IL1RAPL2*) (Figure 3A, grey box) from contributing to disease as it was not encompassed by two of the EONDT deletions in the current cohort, BAB2650 and BAB2595.

Breakpoint-Junction Sequence Features

Potential mechanisms of SV formation, i.e. upstream rearrangement mechanisms, can be surmised or inferred from the study of breakpoint-junction sequence features and mutational signatures such as insertional complexities, SNVs, microhomology, microhomeology (highly similar but imperfect matches) and flanking sequence similarity (Beck et al., 2019; Conrad et al., 2010; Drier et al., 2013; Liu et al., 2017; Ottaviani, LeCain, & Sheer, 2014; Yang et al., 2013). We aligned the eight deletion junction sequences to the reference haploid genome and examined them for the aforementioned features.

Seven of the eight junctions yielded simple alignments, i.e. were resolved by a single join-point between two non-contiguous sequences, while one, the junction observed in BAB8120, was complex with at least two join-points linking a 22-bp insertion to the deletion breakpoints. The insertion in BAB8120 does not appear to have a perfect match in the reference haploid genome consistent with a non-templated insertion or a bi-product of iterative template switching. Notably, due to the presence of the same 5-bp sequence unit twice in the patient's junction sequence flanking the 22-bp insertion (Figure 2F), two alignments to the reference haploid genome are feasible in this case, each exploiting different features and inferring distinct potential mechanism/s of formation (Supp. Figure S2); thus, the status of microhomology cannot be conclusively determined in BAB8120.

Most deletion junctions (7/8) were apparently simple DNA end-joining with 6 out of those 7 (~85%) revealing microhomology (≥ 1 , range: 1-124 bp) and two (~29%), BAB2614 and BAB12522, revealing microhomology (Table 3). Increased flanking sequence similarity was observed in BAB2614 and in one of two alignments of BAB8120 (Supp. Figure S3).

Junction sequences (range: 112-730 bp) were examined for SNVs. Variants were found in five instances involving 4 cases, BAB2614, BAB8201, BAB8204 and BAB12522 (Supp. Table 3 and Figure 2); three of the identified SNVs (in BAB2614, BAB8201 and BAB12522) were rare (1000Genomes MAF: 0, 0 and 0.001 respectively), while two (in BAB2614 and BAB8201) were polymorphic (1000Genome MAF: 0.083 and 1 respectively). Notably, the variant identified in BAB8201 was present in the junction sequence of his mother and grandmother (BAB8202 and BAB8203 respectively) who also carry the Xq22-*PLP1* deletion, hence indicating *in-cis* inheritance with the deletion (pedigree shown in Figure 1). We were unable to explore for the possible *de novo* nature of the identified rare SNVs due to unavailability of parental/grandparental samples. Examined junction sequences are provided in Supp. Table S3.

Overall, junction sequences studied herein reveal some common features with the five previously published Xq22-*PLP1* deletion junction sequences (Inoue et al., 2002; Lee et al., 2007; Matsufuji et al., 2013; Torisu et al., 2012) including microhomology (≥ 1 , range: 1-124 bp) and insertional complexities; out of all 13 sequenced junctions to date, explicit microhomology was observed in 10/13 (~77%) and insertional complexities were observed in 3/13 (~23%) (Table 3).

Highly-Similar Intrachromosomal Repeats and other Large Repeat Structures at Deletion Breakpoints

Genome architectural features at CNV breakpoints can potentially reflect intrinsic structures rendering the region susceptible to genomic instability, DNA single and/or double stranded breaks, and to being a potential rearrangement substrate for DNA break repair (Carvalho & Lupski, 2016; J. M. Chen, Cooper, Ferec, Kehrer-Sawatzki, & Patrinos, 2010). We explored the genome architecture at deletion breakpoints in the haploid reference computationally via our in-house developed method for detecting highly-similar intrachromosomal repeats (HSIRs) as described in the Methods section. Breakpoints were also investigated for coincidence with repeat structures that were not part of the HSIR dataset, but were found on any one of the PMD-LCRs dataset (Lee, Inoue, et al., 2006), the Segmental Dups dataset

(Bailey et al., 2002; Bailey et al., 2001), or the RepeatMasker dataset (Jurka, 2000). This analysis was applied to a compiled cohort of 17 Xq22-*PLP1* deletions (eight new and nine from the published literature).

Nine deletions (out of 17, ~53%) had breakpoints embedded within HSIRs with the greater part of the nine (5/9, 55%) involving a single HSIR, RepX-i1010. This HSIR is 142 kb in size and pair-aligns in an inverted orientation (> 95% identity) to an equally-sized fragment on the p arm of chromosome X (chrX:52,103,260-52,245,416). In addition to those found within RepX-i1010, six other breakpoints were found nearby this HSIR at telomeric distances of ~ 2 kb, ~13 kb, ~17 kb, 20 kb, ~22 kb and ~28 kb (Figure 4A). Taken together, ~32% (11/34) of all Xq22-*PLP1* deletion breakpoints are clustered within or near the telomeric half of RepX-i1010 in a region that is < 90 kb in size. Such dense clustering of breakpoints is not seen anywhere else in the genomic vicinity implicated in Xq22-*PLP1* deletions. We therefore propose this region to be a genomic instability hotspot (GI-hotspot) for DNA breakage (Figure 4A).

Other HSIRs were also identified at deletion breakpoints however they were not as recurrently implicated as RepX-i1010. The distribution of HSIRs and PMD-LCRs in the interval covering *PLP1* deletions (chrX:100,650,000-106,703,000) and their coincidence with deletion breakpoints are illustrated in Supp. Figure S4. Notably, HSIRs that are part of inverted-repeat-pairs are more enriched in this locus than direct-repeat-pair HSIRs (Supp. Figure S4). Furthermore, all HSIRs that were identified at deletion breakpoints were part of inverted repeat pairs while no evidence was found for the involvement of any of the directly-oriented HSIR pairs in this region.

Three of the 17 Xq22-*PLP1* deletion cases had breakpoints embedded within repeat-element pairs, including two from this cohort, BAB2614 and BAB12522. The deletion breakpoint-junction in BAB2614 merges a pair of directly-oriented *AluY* elements yielding a chimeric *Alu* hybrid (Figure 2G) thus defining an *Alu*-*Alu* recombination (Song et al., 2018); the implicated *AluY* pair share 86.8% identity with 0.6% gaps as detected by the Smith-Waterman pair-alignment algorithm. Whereas, the deletion junction in BAB12522 merges a pair of directly oriented LINEs (Figure 2H).

Overall, out of the 17 cases of Xq22-*PLP1* deletions known to date, the majority (12/17, ~71%) have at least one breakpoint residing in or near (maximum separation = 16.5 kb) an HSIR, four have breakpoints mapping to L1 LINE elements and two have breakpoints mapping to a pair of *Alu* elements (Supp. Table S2).

Non-B DNA Motifs in RepX-i1010 and the GI-Hotspot

Non-B DNA motifs, i.e. sequences that are < 100 nucleotide in size and have the potential to adopt non-B DNA structure, have been associated with genomic instability in several previous studies (Bacolla & Wells, 2009; Boyer, Grigurevic, Cazaux, & Hoffmann, 2013; X. Chen et al., 2013; Madireddy & Gerhardt, 2017; Oren et al., 2016; Wang & Vasquez, 2014; Wells, 2007; Wojciechowska, Napierala, Larson, & Wells, 2006; Zhao, Bacolla, Wang, & Vasquez, 2010). In particular, short tandem repeats, and inverted repeats (including palindromic or potential cruciform forming sequences) have been found to be enriched near

recurrent and non-recurrent deletion CNV breakpoints (Akgun et al., 1997; X. Chen et al., 2013; Oren et al., 2016). We investigated RepX-i1010 and the GI-hotspot for seven non-B DNA motifs including G-Quadruplex (GQ) forming sequences, Short Tandem Repeats (STR), A-Phased Repeats (APR), Mirror Repeats (MR), Inverted Repeats (IR), Direct Repeats (DR), and Z-DNA forming sequences (Cer et al., 2013; Wells, 2007). Non-B DNA motifs, specifically STRs, within large repeat structures (e.g. LCR) have been previously proposed to incite genomic instability and the formation of non-recurrent-but-clustered CNV breakpoints (Li, Yen, & Shapiro, 1992; Liu et al., 2011). To understand the distribution and enrichment of non-B DNA motifs throughout the RepX-i1010 locus and the GI-hotspot region, we compared the density of these motifs in 10-kb intervals among three different reference genome sequence groups: 1) the distal half (70 kb) of RepX-i1010 where breakpoints cluster against the proximal half (70 kb) of it (Figure 4A), 2) RepX-i1010 against two flanking similarly sized (140 kb) control regions (Figure 4B), and 3) the GI-hotspot region (90 kb in total, the distal 60 kb from RepX-i1010 + 30 kb telomeric to that) against two flanking similarly-sized (90 kb) control regions (Figure 4C). Heatmap visualization of the density of motifs in the three tested groups is illustrated in Figure 4.

Analysis of the first group indicated no significant enrichment in the distal half of RepX-i1010 as compared to the proximal half (one-sided Wilcoxon Rank Sum test), however, the distribution of IR motifs was significantly distinct between the two halves (two-sided Wilcoxon Rank Sum test, P-value = 0.017). Analysis of the second group revealed a significant enrichment of STR motifs within RepX-i1010 as compared to both flanking control regions (one-sided Wilcoxon Rank Sum test, P-value = 0.021 with proximal control, and 0.009 with distal control). MR and DR motifs were also enriched within RepX-i1010, however only in comparison to the distal control region (one-sided Wilcoxon Rank Sum test, P-value = 0.042, and 0.021 respectively). The third group analysis yielded enrichment of STR motifs in the GI-hotspot as compared to the downstream control region (one-sided Wilcoxon Rank Sum test, P-value = 0.022) and enrichment of Z-DNA motifs in the GI-hotspot as compared to the upstream control region (one-sided Wilcoxon Rank Sum test, P-value = 0.028). IR motifs were distributed distinctively across the GI-hotspot and the proximal control region (two-sided Wilcoxon Rank Sum test, P-value = 0.014). P-values of other tests are indicated in Table 4.

In summary, we found evidence for the potential role of STR and Z-DNA motifs in inciting instability and leading to Xq22-*PLP1*-DEL formation as they were significantly enriched at the proposed genomic instability hotspot where ~47% of proximal breakpoints appeared to be clustered.

XCI Status in Blood Tissue of Female Probands

XCI testing on peripheral blood-derived DNA using microsatellite trinucleotide polymorphic alleles of the androgen receptor gene, *AR*, revealed heterogeneous results among the five studied female cases. Four of them (BAB2614, BAB2595, BAB2650 and BAB12522) showed random patterns of XCI in peripheral blood, while the fourth (BAB8120) revealed XCI skewing where >90% inactivation of the deletion-bearing X-chromosome (paternal

chromosome as revealed by SNP-phasing, data not shown) was evident. Data summarized in Table 3.

Discussion

We investigated the molecular etiology and genotype-phenotype correlation in unrelated subjects from a cohort of Xq22-*PLP1* deletions consisting of eight cases, five females and three males (Figure 1). Deletions of *PLP1* are rare and this study perhaps represents the largest investigated cohort of Xq22 deletions extending over *PLP1* to date. Female cases presented with clinical findings that were manifest as an entity defined here as early onset neurological disease trait (EONDT) consistent with clinical findings associated with large heterozygous Xq22-*PLP1* deletions in the literature (Yamamoto et al., 2014) (Table 2). We found evidence for the potential contribution to disease severity and EONDT in female cases of a genomic interval upstream of *PLP1* – a region previously implicated (Figure 3A). Our genomic analysis of deletion junction sequences further revealed that microhomology (1 bp) is a frequently observed feature among Xq22-*PLP1* deletion rearrangements (Figure 2) and that large inverted repeat structures as well as non-B DNA motifs, specifically STRs, have a potential role in inciting genomic instability and stimulating Xq22 deletion formation near the *PLP1* locus (Figure 4). Additionally, the sequencing and alignment of deletion breakpoint junctions exposed experimental data implicating an apparent breakpoint hotspot that was not delineated from a computational approach alone (Figure 4).

Xq22 Contiguous-Gene Deletion Syndrome in Females

Disease manifestation in the four EONDT female patients that were clinically investigated in this study echoes the characteristic clinical features described by Yamamoto et al. (2014) in their cohort of three large heterozygous Xq22-*PLP1* deletion cases (Table 2 and Additional file 3), hence enabling us to examine for genotype/phenotype associations among both cohorts. Individuals from both cohorts were found to present with core clinical characteristics including neonatal hypotonia, severe DD/ID, behavioral abnormalities and strabismus. A photograph of the face is available in a published clinical report on one patient from our cohort (Brender et al., 2015) and it revealed dysmorphic features common with what was found in the other three published cases (Yamamoto et al., 2014) including a triangular face, large forehead, strabismus and prominent jaw. This suggests that facial dysmorphism may also emerge as a possible core recognizable feature of EONDT, although one of our patients (BAB12522) did not have dysmorphism.

Brain MRI was obtained at different ages in the two cohorts thus it is challenging to directly compare findings; however, there are common features in both studies including nonspecific increased white matter signal intensities, delayed myelination and thin corpus callosum. Increased white matter signal intensity on T2-weighted images of MRI and spasticity are features common to male subjects with SPG2 and subtle but similar features can be present in females with *PLP1* LoF (Inoue, 2005; Inoue et al., 2002; Matsufuji et al., 2013; Torisu et al., 2012) (Table 2). However, the other core features of EONDT are distinct in nature and are not parsimoniously explained by *PLP1* nullisomy on its own. Moreover, EONDT can be distinguished from PMD, an allelic disorder of SPG2, by the distinctive severity and early

onset of disease in EONDT females as compared to mostly non-symptomatic or late-onset-disease females from PMD families.

Variable expressivity in females carrying heterozygous Xq22-*PLP1* deletions is proposed to be related to inter-individual differences in XCI patterns as well as deletion size and gene content. Skewed XCI has been one molecular explanation for manifesting female carriers of Xq22-*PLP1* deletion CNV and indeed skewed XCI patterns are seen in the peripheral blood of some EODNT patients (Table 3); however, this skewing is not necessarily reflected in the brain due to possible intra-individual XCI variation. Hence, although it may have a role, we are unable to provide any evidence for consistent XCI skewing potentially contributing to neurological disease severity.

Yamamoto et al. (2014) proposed two SROs flanking *PLP1* to potentially contribute to the distinct features of disease in EONDT females with large Xq22-*PLP1* deletions (>3 Mb) (Figure 3A, grey boxes). They suggested *BEX3*, located in an ~700 kb SRO upstream of *PLP1* that contains 14 genes in total (chrX:102,233,526-102,957,289), as the most likely contributing candidate gene as it is preferentially expressed in the brain and it encodes the Nerve Growth Factor Receptor Associated Protein 1, NGFRAP1, also called p75NTR-associated cell death executor, NADE, disruption of which may lead to abnormal signal transduction and aberrant neuronal development (Calvo et al., 2015; Mukai et al., 2000). However, due to the limited number of EONDT female patients, other brain-expressed genes encompassed within the SRO containing *BEX3* could not be ruled out from possible contribution to disease either alone or as part of a mutational burden along with *BEX3*. Additionally, they proposed a second SRO (chrX:102,993,719-103,982,269), downstream of *PLP1* and containing a single gene, *ILIRAPL2*, as a possible candidate region underlying EONDT in females with large Xq22 deletions. Here, we investigated the two suggested SROs in similarly affected patients and found further evidence that supports the potential contribution of the SRO upstream of *PLP1* to distinct disease features of EONDT in females. We found a refined interval (chrX:102,615,641-102,957,288) of the proposed upstream SRO (Yamamoto et al., 2014) to be shared among deletions of all EONDT cases and absent from deletions of all SPG2 cases (Figure 3A, EONDT-SRO). This further emphasizes a possible role for the EONDT-SRO in contributing to disease potentially via haploinsufficiency of one, some or all of the encompassed genes or by disrupting an enhancer of a disease-causing gene in the region. Importantly, the refined interval contains one of the candidate genes proposed by Yamamoto et al. (2014) to be associated with the neurological components of this disease, *BEX3*. In addition to being encompassed by deletions of EONDT female cases, *BEX3* has the highest expression levels in the brain as compared to the other five genes within the EONDT-SRO (all expressed in the brain) (Figure 3C). Moreover, *BEX3* (among other contiguous centromeric genes and one telomeric gene, *RAB40A*) has been associated with severe neurological impairment and reduced cerebral volume in a male patient harboring a hemizygous deletion that is not inclusive of *PLP1* (Shirai, Higashi, Shimojima, & Yamamoto, 2017). Nevertheless, further studies of patients with *BEX3* intragenic deletions or point mutations or studies of a knockout mouse model may be required to support its involvement in neurological disease.

The EONDT-SRO contains *MORF4L2* (and its antisense transcript *MORF4L2-AS1*), a brain-expressed gene that is proposed to escape XCI making it prone to dosage sensitivity in 46, XX females (Y. Zhang et al., 2013). Moreover, Labonne et al. (2016) reported a female patient with some clinical features that overlap with features of EONDT and a heterozygous Xq22 deletion that extends over *GLRA4* and two of the EONDT-SRO genes, *MORF4L2* and *TCEAL1* (Figure 3, deletion framed in yellow). They maintained that *GLRA4* is likely the disease-causing gene leading to the phenotype of their patient. We found *GLRA4* to be outside of the EONDT-SRO and to be encompassed by an SPG2 family (affected male and female) from the literature with no evidence of any of the EONDT features in affected individuals (Inoue et al., 2002) (Figure 3A); this leads us to the interpretation that *GLRA4* does not lead to EONDT features when deleted in a heterozygous nor in a hemizygous state. It also suggests that one or both of the EONDT-SRO genes encompassed by the patient of Labonne et al. (2016), *MORF4L2* and *TCEAL1* (Figure 3A, DGDP084; deletion framed in yellow), are potentially contributing as disease-causing candidates.

We cannot exclude the potential contribution of any of the genes encompassed by the EONDT-SRO to some or all of the phenotypic features of EONDT. It is likely that genes within the EONDT-SRO are contributing, in addition to *PLP1*, to the overall characteristics of an emerging contiguous gene deletion syndrome observed as EONDT in female cases (Campbell et al., 2012; Schmickel, 1986). An intragenic deletion or a rare likely damaging SNV/Indel in any one of the EONDT-SRO genes could potentially provide critical evidence to further dissect this contiguous gene syndrome. It is possible that in females with large deletions, *PLP1* is leading to features common to SPG2 such as white matter MRI findings and spasticity an expected outcome from *PLP1* LoF, while other genes in the EONDT-SRO, are contributing to the additional distinct features including the severe DD/ID, the behavioral abnormalities, and the distinct facial dysmorphism. The proposed contribution of *PLP1* to SPG2 features of disease in EONDT patients is further evidenced by the lack of these features in the patient described by Labonne et al. (2016) whose Xq22 deletion does not encompass *PLP1*.

Mechanisms Mediating the Formation of Xq22-PLP1 Deletions

The rarity of Xq22-*PLP1* deletions has been proposed to be due to 1) intolerance to the loss of contiguous genes and 2) possibly different mechanisms leading to their formation as opposed to the more common mutational events in this locus such as Xq22-*PLP1* duplications and triplications (Inoue et al., 2002). Understanding genomic features of Xq22-*PLP1* deletion breakpoints is therefore essential for the inference of potential triggers inciting DNA damage at this locus, including intrinsic genome architecture causing genomic instability, and for inference of the rearrangement mechanisms leading to deletion CNV formation and potentially contributing to the prevalence of disease. Due to the limited number of cases previously identified and studied, a comprehensive analysis of genomic features of Xq22-*PLP1* deletions has not been possible. Earlier studies of five Xq22-*PLP1* deletions suggested microhomology as a prominent junction feature and non-homologous end joining (NHEJ) as a potential mechanism of formation (Inoue et al., 2002; Matsufuji et al., 2013; Torisu et al., 2012).

In line with previous observations and contentions, here we also find microhomology in the majority of junctions (75%) and find features of the end-joining mechanisms of double stranded break (DSB) repair, NHEJ and microhomology-mediated end joining (MMEJ), to be common among new Xq22-*PLP1* deletions. Nevertheless, features of break induced replication (BIR) mechanisms such as *Alu-Alu* mediated rearrangement (AAMR) and fork stalling template switching/microhomology-mediated break induced replication (FoSTeS/MMBIR) are also apparent (Beck et al., 2019; Carvalho & Lupski, 2016; Carvalho et al., 2013; Carvalho et al., 2011; Gu et al., 2015; Hsiao et al., 2015; Lee et al., 2007; Song et al., 2018; F. Zhang et al., 2009). In fact, the SNVs, microhomology, microhomeology and the chimeric *Alu* hybrid at the breakpoint-junction of BAB2614 (Figure 2G) are all features indicative of AAMR (Song et al., 2018). The breakpoint-junction in BAB12522 as well reveals features of BIR including microhomology, microhomeology, a pair of LINEs and an SNV. Moreover, the complexities observed in the junction sequence of BAB8120 (5-bp direct repeat flanking a 22-bp insertion as shown in Figure 2F) indicate the probable involvement of serial replication slippage (SRS) or FoSTeS/MMBIR (Supp. Figure S2) although further narrowing of the most parsimonious explanatory mechanism leading to the deletion formation is challenging to discern in this case given the presence of a pair of 5-bp direct repeats.

While taking microhomology into consideration only when 2 bp in size (as 1 bp apparent microhomology can potentially be due to NHEJ and end processing), we propose that NHEJ and MMEJ collectively mediate at least seven out of the 13 sequenced Xq22-*PLP1* deletions to date, FoSTeS/MMBIR may play a role in at least three where insertional complexities and iterative template switching are apparent and AAMR is implicated in the two cases where *Alu-Alu* junctions are observed. The diversity of mechanisms potentially underlying the formation of Xq22-*PLP1* deletions argues against a mechanistic limiting factor in their formation and supports the contention that other factors may be contributing to the rarity of these cases such as lethality in cases of large hemizygous deletions or milder disease that does not call for genetic testing in females with small non-EONDT deletions.

Hotspot of Genomic Instability

Although the observed Xq22-*PLP1* deletions (16 deletions) are highly variable in size, more than 30% of all breakpoints taken together were found clustered in an interval that is less than 90 kb in size indicating a genomic instability hotspot within this interval. One-sided grouping or clustering of nonrecurrent breakpoints has been observed in other loci including distal breakpoints of duplications at the *MECP2* locus which were clustered in an LCR-rich 215-kb genomic interval on Xq28 and the grouping of the triplication end site of Xq22-*PLP1* DUP-TRP/INV-DUP events near a pair of PMD-LCRs, PMD-LCRA1a and PMD-LCR-A1b (Beck et al., 2015; Carvalho et al., 2009). LCRs in the *MECP2* locus and the Xq22-*PLP1* DUP-TRP/INV-DUP events were proposed to stimulate replicative repair mechanisms leading to the formation of complex genomic rearrangement events. Here, we identified a GI-hotspot via the experimental approach of mapping deletion breakpoint to base pair level and computationally found two architectural features to be overlapping the GI-hotspot of Xq22-*PLP1* deletions, an inverted paralogous repeat, RepX-i1010, and significant enrichment of non-B DNA motifs, STR and Z-DNA motifs. RepX-i1010 was not detected

by any of the publicly available repeat datasets; however, it was functionally delineated in this study due to its frequent coincidence with base-pair-resolution mapped breakpoints and was further confirmed by our computational approach for identifying HSIRs. This illustrates the role of experimental data and of mapping CNV breakpoints from the personal genomes of individual patients in not only supplementing computational approaches but also in allowing the elucidation of clustering of breakpoint-junctions and the discovery of intrinsic genomic architecture causing instability that can be missed otherwise due to methodology limitations.

Inverted repeat structures are proposed to incite genomic instability (Dittwald et al., 2013) and to stimulate single ended DSB (seDSB) repair (Carvalho & Lupski, 2016) while STR and Z-DNA motifs are predicted to form non-B DNA structures rendering genomic instability, DNA fragility and consequently leading to genomic disorders (Bacolla, Tainer, Vasquez, & Cooper, 2016; X. Chen et al., 2013; Javadkar, Yadav, & Raghavan, 2018; Li et al., 1992; Nambiar, Srivastava, Gopalakrishnan, Sankaran, & Raghavan, 2013; Wojcik et al., 2012). Moreover, Z-DNA motifs have been specifically associated with genomic instability rendering the formation of large deletions (Wang, Christensen, & Vasquez, 2006). The coincidence of both these features in a small interval where non-recurrent breakpoints cluster parsimoniously explains the heterogeneity of mutational signatures seen in the rearrangement products. We propose that each of these architectural features likely incites a distinct environment of genomic instability (e.g. DSBs, fork stalling and/or DNA nicks) and potentially stimulates the employment and progression of different mechanisms of repair (e.g. NHEJ, MMEJ, DS-MMEJ, MMBIR/FoSTeS, and AAMR).

Conclusion

This study provides mechanistic insights into Xq22-*PLP1* deletions both at the upstream level of rearrangement mechanisms leading to deletion formation and the downstream level of genic mechanisms and gene(s) contributions to trait manifestations. It provides further evidence for an Xq22 contiguous-gene deletion syndrome in females and elaborates on the recognizable clinical features of this syndrome. It also emphasizes that patient studies can enable experimental and functional characterization of mutational events and uncover intrinsic properties of the human genome rendering genomic instability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all individuals, their families and the referring physicians who submitted samples for testing. No additional compensation was received for these contributions. The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at <https://gnomad.broadinstitute.org/about>.

Funding

This work was supported in part by the US National Human Genome Research Institute (NHGRI)/National Heart Lung and Blood Institute (NHLBI) UM1 HG006542 to the Baylor Hopkins Center for Mendelian Genomics

(BHCMG), by the US National Institute of Neurological Disorders and Stroke (NINDS) grants R01 NS058529 and R35 NS105078 to J.R. Lupski and R01 NS058978 to G.M. Hobson, by the National Institute of General Medical Sciences (NIGMS) grants, R01 GM106373 to J.R. Lupski and by the PMD Foundation to G.M. Hobson. G.M. Hobson and the Nemours Biomolecular Core Lab are supported by the COBRE grant, P30 GM114736. The Nemours Biomolecular Core Lab is also supported by the INBRE grant, P20 GM103446. The Genotype-Tissue Expression (GTEx) Project was supported by the [Common Fund](#) of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

Funding Information:

UM1 HG006542, R01 NS058529, R35 NS105078, R01 NS058978, RO1GM106373, P30GM114736 and P20GM103446.

References

- Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, ... Jasin M (1997). Palindrome resolution and recombination in the mammalian germ line. *Mol Cell Biol*, 17(9), 5559–5570. doi: 10.1128/mcb.17.9.5559 [PubMed: 9271431]
- Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, & Belmont JW (1992). Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet*, 51(6), 1229–1239. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/1281384> [PubMed: 1281384]
- Bacolla A, Tainer JA, Vasquez KM, & Cooper DN (2016). Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. *Nucleic Acids Res*, 44(12), 5673–5688. doi:10.1093/nar/gkw261 [PubMed: 27084947]
- Bacolla A, & Wells RD (2009). Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog*, 48(4), 273–285. doi:10.1002/mc.20507 [PubMed: 19306308]
- Bahrambeigi V, Song X, Sperle K, Beck CR, Hijazi H, Grochowski CM, ... Lupski JR (2019). Distinct patterns of complex rearrangements and a mutational signature of microhomeology are frequently observed in PLP1 copy number gain structural variants. Manuscript submitted for publication.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, ... Eichler EE (2002). Recent segmental duplications in the human genome. *Science*, 297(5583), 1003–1007. doi:10.1126/science.1072047 [PubMed: 12169732]
- Bailey JA, Yavor AM, Massa HF, Trask BJ, & Eichler EE (2001). Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res*, 11(6), 1005–1017. doi:10.1101/gr.gr-1871r [PubMed: 11381028]
- Beck CR, Carvalho CM, Baner L, Gambin T, Stubbolo D, Yuan B, ... Lupski JR (2015). Complex genomic rearrangements at the PLP1 locus include triplication and quadruplication. *PLoS Genet*, 11(3), e1005050. doi:10.1371/journal.pgen.1005050 [PubMed: 25749076]
- Beck CR, Carvalho CMB, Akdemir ZC, Sedlazeck FJ, Song X, Meng Q, ... Lupski JR (2019). Megabase Length Hypermutation Accompanies Human Structural Variation at 17p11.2. *Cell*, 176(6), 1310–1324 e1310. doi:10.1016/j.cell.2019.01.045 [PubMed: 30827684]
- Boyer AS, Grgurevic S, Cazaux C, & Hoffmann JS (2013). The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. *J Mol Biol*, 425(23), 4767–4781. doi:10.1016/j.jmb.2013.09.022 [PubMed: 24095858]
- Brender T, Wallerstein D, Sum J, & Wallerstein R (2015). Unusual presentation of pelizaeus-merzbacher disease: female patient with deletion of the proteolipid protein 1 gene. *Case Rep Genet*, 2015, 453105. doi:10.1155/2015/453105 [PubMed: 25789183]
- Calvo L, Anta B, Lopez-Benito S, Martin-Rodriguez C, Lee FS, Perez P, ... Arevalo JC (2015). Bex3 Dimerization Regulates NGF-Dependent Neuronal Survival and Differentiation by Enhancing trkA Gene Transcription. *J Neurosci*, 35(18), 7190–7202. doi:10.1523/JNEUROSCI.4646-14.2015 [PubMed: 25948268]
- Campbell IM, Yatsenko SA, Hixson P, Reimschisel T, Thomas M, Wilson W, ... Scaglia F (2012). Novel 9q34.11 gene deletions encompassing combinations of four Mendelian disease genes: STXBP1, SPTAN1, ENG, and TOR1A. *Genet Med*, 14(10), 868–876. doi:10.1038/gim.2012.65 [PubMed: 22722545]

- Carvalho CM, Bartnik M, Pehlivan D, Fang P, Shen J, & Lupski JR (2012). Evidence for disease penetrance relating to CNV size: Pelizaeus-Merzbacher disease and manifesting carriers with a familial 11 Mb duplication at Xq22. *Clin Genet*, 81(6), 532–541. doi:10.1111/j.1399-0004.2011.01716.x [PubMed: 21623770]
- Carvalho CM, & Lupski JR (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet*, 17(4), 224–238. doi:10.1038/nrg.2015.25 [PubMed: 26924765]
- Carvalho CM, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, ... Lupski JR (2013). Replicative mechanisms for CNV formation are error prone. *Nat Genet*, 45(11), 1319–1326. doi:10.1038/ng.2768 [PubMed: 24056715]
- Carvalho CM, Ramocki MB, Pehlivan D, Franco LM, Gonzaga-Jauregui C, Fang P, ... Lupski JR (2011). Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat Genet*, 43(11), 1074–1081. doi:10.1038/ng.944 [PubMed: 21964572]
- Carvalho CM, Zhang F, Liu P, Patel A, Sahoo T, Bacino CA, ... Lupski JR (2009). Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. *Hum Mol Genet*, 18(12), 2188–2203. doi:10.1093/hmg/ddp151 [PubMed: 19324899]
- Cer RZ, Donohue DE, Mudunuri US, Temiz NA, Loss MA, Starner NJ, ... Stephens RM (2013). Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res*, 41(Database issue), D94–D100. doi:10.1093/nar/gks955 [PubMed: 23125372]
- Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, & Patrinos GP (2010). Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*, 20(4), 222–233. doi:10.1016/j.semcancer.2010.05.007 [PubMed: 20541013]
- Chen X, Shen Y, Zhang F, Chiang C, Pillalamarri V, Blumenthal I, ... Gusella JF (2013). Molecular analysis of a deletion hotspot in the NRXN1 region reveals the involvement of short inverted repeats in deletion CNVs. *Am J Hum Genet*, 92(3), 375–386. doi:10.1016/j.ajhg.2013.02.006 [PubMed: 23472757]
- Chiaromonte F, Yap VB, & Miller W (2002). Scoring pairwise genomic sequence alignments. *Pac Symp Biocomput*, 115–126. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11928468> [PubMed: 11928468]
- Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, ... Hurles ME (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat Genet*, 42(5), 385–391. doi:10.1038/ng.564 [PubMed: 20364136]
- Dittwald P, Gambin T, Gonzaga-Jauregui C, Carvalho CM, Lupski JR, Stankiewicz P, & Gambin A (2013). Inverted low-copy repeats and genome instability--a genome-wide analysis. *Hum Mutat*, 34(1), 210–220. doi:10.1002/humu.22217 [PubMed: 22965494]
- Drier Y, Lawrence MS, Carter SL, Stewart C, Gabriel SB, Lander ES, ... Getz G (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res*, 23(2), 228–235. doi:10.1101/gr.141382.112 [PubMed: 23124520]
- Garbern JY, Cambi F, Tang XM, Sima AA, Vallat JM, Bosch EP, ... Kamholz J (1997). Proteolipid protein is necessary in peripheral as well as central myelin. *Neuron*, 19(1), 205–218. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9247276> [PubMed: 9247276]
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, ... Abecasis GR (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. doi:10.1038/nature15393 [PubMed: 26432245]
- Gu S, Yuan B, Campbell IM, Beck CR, Carvalho CM, Nagamani SC, ... Lupski JR (2015). Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Hum Mol Genet*, 24(14), 4061–4077. doi:10.1093/hmg/ddv146 [PubMed: 25908615]
- Hsiao MC, Piotrowski A, Callens T, Fu C, Wimmer K, Claes KB, & Messiaen L (2015). Decoding NF1 Intragenic Copy-Number Variations. *Am J Hum Genet*, 97(2), 238–249. doi:10.1016/j.ajhg.2015.06.002 [PubMed: 26189818]
- Inoue K (2005). PLP1-related inherited dysmyelinating disorders: Pelizaeus-Merzbacher disease and spastic paraplegia type 2. *Neurogenetics*, 6(1), 1–16. doi:10.1007/s10048-004-0207-y [PubMed: 15627202]

- Inoue K (2017). Cellular Pathology of Pelizaeus-Merzbacher Disease Involving Chaperones Associated with Endoplasmic Reticulum Stress. *Front Mol Biosci*, 4, 7. doi:10.3389/fmolb.2017.00007 [PubMed: 28286750]
- Inoue K, Osaka H, Thurston VC, Clarke JT, Yoneyama A, Rosenbarker L, ... Lupski JR (2002). Genomic rearrangements resulting in PLP1 deletion occur by nonhomologous end joining and cause different dysmyelinating phenotypes in males and females. *Am J Hum Genet*, 71(4), 838–853. doi:10.1086/342728 [PubMed: 12297985]
- Javadekar SM, Yadav R, & Raghavan SC (2018). DNA structural basis for fragility at peak III of BCL2 major breakpoint region associated with t(14;18) translocation. *Biochim Biophys Acta Gen Subj*, 1862(3), 649–659. doi:10.1016/j.bbagen.2017.12.003 [PubMed: 29246583]
- Jurka J (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet*, 16(9), 418–420. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10973072> [PubMed: 10973072]
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, ... MacArthur DG (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210. doi:10.1101/531210
- Kent WJ, Baertsch R, Hinrichs A, Miller W, & Haussler D (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A*, 100(20), 11484–11489. doi:10.1073/pnas.1932072100 [PubMed: 14500911]
- Labonne JD, Graves TD, Shen Y, Jones JR, Kong IK, Layman LC, & Kim HG (2016). A microdeletion at Xq22.2 implicates a glycine receptor GLRA4 involved in intellectual disability, behavioral problems and craniofacial anomalies. *BMC Neurol*, 16, 132. doi:10.1186/s12883-016-0642-z [PubMed: 27506666]
- Lee JA, Carvalho CM, & Lupski JR (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131(7), 1235–1247. doi:10.1016/j.cell.2007.11.037 [PubMed: 18160035]
- Lee JA, Inoue K, Cheung SW, Shaw CA, Stankiewicz P, & Lupski JR (2006). Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease. *Hum Mol Genet*, 15(14), 2250–2265. doi:10.1093/hmg/ddl150 [PubMed: 16774974]
- Lee JA, Madrid RE, Sperle K, Ritterson CM, Hobson GM, Garbern J, ... Inoue K (2006). Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann Neurol*, 59(2), 398–403. doi:10.1002/ana.20732 [PubMed: 16374829]
- Li XM, Yen PH, & Shapiro LJ (1992). Characterization of a low copy repetitive element S232 involved in the generation of frequent deletions of the distal short arm of the human X chromosome. *Nucleic Acids Res*, 20(5), 1117–1122. doi:10.1093/nar/20.5.1117 [PubMed: 1549475]
- Liu P, Erez A, Nagamani SC, Bi W, Carvalho CM, Simmons AD, ... Lupski JR (2011). Copy number gain at Xp22.31 includes complex duplication rearrangements and recurrent triplications. *Hum Mol Genet*, 20(10), 1975–1988. doi:10.1093/hmg/ddr078 [PubMed: 21355048]
- Liu P, Yuan B, Carvalho CMB, Wuster A, Walter K, Zhang L, ... Lupski JR (2017). An Organismal CNV Mutator Phenotype Restricted to Early Human Development. *Cell*, 168(5), 830–842 e837. doi:10.1016/j.cell.2017.01.037 [PubMed: 28235197]
- Madireddy A, & Gerhardt J (2017). Replication Through Repetitive DNA Elements and Their Role in Human Diseases. *Adv Exp Med Biol*, 1042, 549–581. doi:10.1007/978-981-10-6955-0_23 [PubMed: 29357073]
- Matsufuji M, Osaka H, Gotoh L, Shimbo H, Takashima S, & Inoue K (2013). Partial PLP1 deletion causing X-linked dominant spastic paraplegia type 2. *Pediatr Neurol*, 49(6), 477–481. doi:10.1016/j.pediatrneurol.2013.07.012 [PubMed: 24095575]
- Mukai J, Hachiya T, Shoji-Hoshino S, Kimura MT, Nadano D, Suvanto P, ... Sato TA (2000). NADE, a p75NTR-associated cell death executor, is involved in signal transduction mediated by the common neurotrophin receptor p75NTR. *J Biol Chem*, 275(23), 17566–17570. doi:10.1074/jbc.C000140200 [PubMed: 10764727]
- Nambiar M, Srivastava M, Gopalakrishnan V, Sankaran SK, & Raghavan SC (2013). G-quadruplex structures formed at the HOX11 breakpoint region contribute to its fragility during t(10;14)

- translocation in T-cell leukemia. *Mol Cell Biol*, 33(21), 4266–4281. doi:10.1128/MCB.00540-13 [PubMed: 24001773]
- Oren M, Barela Hudgell MA, D'Allura B, Agronin J, Gross A, Podini D, & Smith LC (2016). Short tandem repeats, segmental duplications, gene deletion, and genomic instability in a rapidly diversified immune gene family. *BMC Genomics*, 17(1), 900. doi:10.1186/s12864-016-3241-x [PubMed: 27829352]
- Osorio MJ, & Goldman SA (2018). Neurogenetics of Pelizaeus-Merzbacher disease. *Handb Clin Neurol*, 148, 701–722. doi:10.1016/B978-0-444-64076-5.00045-4 [PubMed: 29478609]
- Ottaviani D, LeCain M, & Sheer D (2014). The role of microhomology in genomic structural variation. *Trends Genet*, 30(3), 85–94. doi:10.1016/j.tig.2014.01.001 [PubMed: 24503142]
- Raskind WH, Williams CA, Hudson LD, & Bird TD (1991). Complete deletion of the proteolipid protein gene (PLP) in a family with X-linked Pelizaeus-Merzbacher disease. *Am J Hum Genet*, 49(6), 1355–1360. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/1720927> [PubMed: 1720927]
- Schmickel RD (1986). Contiguous gene syndromes: a component of recognizable syndromes. *J Pediatr*, 109(2), 231–241. doi:10.1016/s0022-3476(86)80377-8 [PubMed: 3016222]
- Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, ... Miller W (2003). Human-mouse alignments with BLASTZ. *Genome Res*, 13(1), 103–107. doi:10.1101/gr.809403 [PubMed: 12529312]
- Shirai K, Higashi Y, Shimojima K, & Yamamoto T (2017). An Xq22.1q22.2 nullisomy in a male patient with severe neurological impairment. *Am J Med Genet A*, 173(4), 1124–1127. doi: 10.1002/ajmg.a.38134 [PubMed: 28328133]
- Song X, Beck CR, Du R, Campbell IM, Coban-Akdemir Z, Gu S, ... Lupski JR (2018). Predicting human genes susceptible to genomic instability associated with Alu/Alu-mediated rearrangements. *Genome Res*, 28(8), 1228–1242. doi:10.1101/gr.229401.117 [PubMed: 29907612]
- Torisu H, Iwaki A, Takeshita K, Hiwatashi A, Sanefuji M, Fukumaki Y, & Hara T (2012). Clinical and genetic characterization of a 2-year-old boy with complete PLP1 deletion. *Brain Dev*, 34(10), 852–856. doi:10.1016/j.braindev.2012.02.006 [PubMed: 22401669]
- Wang G, Christensen LA, & Vasquez KM (2006). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc Natl Acad Sci U S A*, 103(8), 2677–2682. doi:10.1073/pnas.0511084103 [PubMed: 16473937]
- Wang G, & Vasquez KM (2014). Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst)*, 19, 143–151. doi:10.1016/j.dnarep.2014.03.017 [PubMed: 24767258]
- Wells RD (2007). Non-B DNA conformations, mutagenesis and disease. *Trends Biochem Sci*, 32(6), 271–278. doi:10.1016/j.tibs.2007.04.003 [PubMed: 17493823]
- Wojciechowska M, Napierala M, Larson JE, & Wells RD (2006). Non-B DNA conformations formed by long repeating tracts of myotonic dystrophy type 1, myotonic dystrophy type 2, and Friedreich's ataxia genes, not the sequences per se, promote mutagenesis in flanking regions. *J Biol Chem*, 281(34), 24531–24543. doi:10.1074/jbc.M603888200 [PubMed: 16793772]
- Wojcik EA, Brzostek A, Bacolla A, Mackiewicz P, Vasquez KM, Korycka-Machala M, ... Dziadek J (2012). Direct and inverted repeats elicit genetic instability by both exploiting and eluding DNA double-strand break repair systems in mycobacteria. *PLoS One*, 7(12), e51064. doi:10.1371/journal.pone.0051064 [PubMed: 23251422]
- Woodward KJ, Cundall M, Sperle K, Sistermans EA, Ross M, Howell G, ... Hobson GM (2005). Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination. *Am J Hum Genet*, 77(6), 966–987. doi: 10.1086/498048 [PubMed: 16380909]
- Yamamoto T, Wilsdon A, Joss S, Isidor B, Erlandsson A, Suri M, ... Stefanova M (2014). An emerging phenotype of Xq22 microdeletions in females with severe intellectual disability, hypotonia and behavioral abnormalities. *J Hum Genet*, 59(6), 300–306. doi:10.1038/jhg.2014.21 [PubMed: 24646727]

- Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, ... Park PJ (2013). Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell*, 153(4), 919–929. doi: 10.1016/j.cell.2013.04.010 [PubMed: 23663786]
- Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, & Lupski JR (2009). The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet*, 41(7), 849–853. doi:10.1038/ng.399 [PubMed: 19543269]
- Zhang L, Wang J, Zhang C, Li D, Carvalho CMB, Ji H, ... Jiang Y (2017). Efficient CNV breakpoint analysis reveals unexpected structural complexity and correlation of dosage-sensitive genes with clinical severity in genomic disorders. *Hum Mol Genet*, 26(10), 1927–1941. doi:10.1093/hmg/ddx102 [PubMed: 28334874]
- Zhang Y, Castillo-Morales A, Jiang M, Zhu Y, Hu L, Urrutia AO, ... Hurst LD (2013). Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Mol Biol Evol*, 30(12), 2588–2601. doi: 10.1093/molbev/mst148 [PubMed: 24023392]
- Zhao J, Bacolla A, Wang G, & Vasquez KM (2010). Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci*, 67(1), 43–62. doi:10.1007/s00018-009-0131-2 [PubMed: 19727556]

Figure 1: Pedigrees of families with available clinical information.

EOND in female probands who do not have similarly-affected male kindreds. Complicated SPG2 in male probands and in their less severely affected female kindreds.

Figure 2: Custom high-density array Comparative Genomic Hybridization and breakpoint-junction sequences of the eight Xq22-*PLP1* deletions.

Microhomology and large repeats are frequently observed at breakpoint-junctions.

Figure 3: Xq22-*PLP1* deletions size, extent and gene content.

A) Alignment of all Xq22-*PLP1* deletions known to date and magnification (green) of the narrowed down smallest region of overlap between EOND deletions (EOND-SRO) that is proposed to contribute to disease. Grey boxes indicate the SROs previously proposed by Yamamoto et al. (2014) to contribute to disease in females with large deletions. **B)** All partial *PLP1* deletions known to date. **C)** GTEx heatmap showing expression levels (as Transcripts per Million, TPM) of genes in the EOND-SRO in comparison to *PLP1* in different parts of the brain. All of the EOND-SRO genes appear to be expressed in the brain.

(F): female case. (M): male case. (M & F): family with male and female cases. EONDT-Like: some clinical features overlapping EONDT. Unclassified: unclassified phenotype. ?: clinical information not available. [‡]Yamamoto et al., 2014, [§]Inoue et al., 2002, [¶]Torisu et al., 2012, [‡]Matsufuji et al., 2013. [¥]Labonne et al., 2016. TPM: transcripts per million.

Figure 4: Non-B DNA motifs in RepX-i1010 and the nearby genomic instability hotspot.

A) The clustering of ~47% of proximal breakpoints and > 30% of all (i.e. proximal and distal) breakpoints of Xq22-*PLP1* deletions in a genomic instability hotspot that partially overlaps with RepX-i1010 and the distribution of non-B DNA motifs within RepX-i1010 **B)** The distribution of non-B DNA in regions flanking RepX-i1010, **C)** The distribution of non-B DNA motifs in regions flanking the 90-kb genomic instability hotspot.

[‡]Yamamoto et al., 2014, [§]Inoue et al., 2002, [¶]Toritsu et al., 2012, [¶]Matsufuji et al., 2013.

[¶]Labonne et al., 2016. ^{**a} Significant difference (P-value <0.05) in the distribution of a motif

between the proximal half and the distal half of RepX-i1010. *^b Significant enrichment (P-value <0.05) of a motif compared to both the proximal and the distal flanking regions. *^c Significant enrichment (P-value <0.05) of a motif compared to the distal flanking region. *^d Significant enrichment (P-value <0.05) of a motif compared to the proximal flanking region.

Table 1 –

Abbreviations Glossary

Acronym	Expansion
aCGH	array-comparative genomic hybridization
AAMR	<i>Alu-Alu</i> mediated rearrangement
APR	A-phased repeat
BIR	Break-induced replication
BAEP	Brain Auditory Evoked Potential
CC	Corpus Callosum
CNV	Copy-number variant
CP	Cerebral palsy
DSB	Double stranded break
DD/ID	Developmental delay/intellectual disability
DR	Direct repeat
EEG	Electroencephalogram
EONDT	Early-onset neurological disease trait
FoSTeS	Fork stalling template switching
GERD	Gastroesophageal reflux disease
GI	Genomic instability
GQ	G-quadruplex
HSIR	Highly similar intrachromosomal repeat
IR	Inverted repeat
LCR	Low-copy repeat
LINE	Long interspersed nuclear element
LoF	Loss of function
MAF	Minor Allele Frequency
MMBIR	Microhomology-mediated break-induced replication
MMEJ	Microhomology-mediated end joining
MR	Mirror repeat
MRI	Magnetic resonance imaging
NADE	p75NTR-associated cell death executor
NHEJ	Non-homologous end joining
PMD	Pelizaeus Merzbacher disease
PVL	Periventricular leukomalacia
seDSB	single ended double stranded break
SINE	Short interspersed nuclear element
SNV	Single nucleotide variant.
SPG2	Spastic paraplegia type 2
SRO	Smallest region of overlap
SRS	Serial replication slippage

Acronym	Expansion
STR	Short tandem repeat
SV	Structural variation
WM	White matter
XCI	X-chromosome inactivation
Xq22-PLP1-DEL	Xq22 deletion that encompasses any part of <i>PLP1</i>

Table 2 –Major clinical findings in new *PLP1* deletion cases

	BAB2595	BAB2650	BAB8120	BAB12522	BAB8201	BAB8204
Sex	F	F	F	F	M	M
DEL Size	2.5 Mb	693 kb	5.6 Mb	3.3 Mb	6.7 kb	71 kb
Gestation	32-33 weeks	34-35 weeks	41 weeks	39 weeks	Full term *	41 weeks
Age **	13 years	9 years	3.5 years	8 years	15 years	16 years
DD/ID	Severe	Severe	Severe	Severe	DD	DD
Behavioral abnormality	ASD	Anxiety	ASD	ASD	Perseverative	None
Eyes	Strabismus	Strabismus, Nystagmus	Strabismus	Strabismus, left amblyopia	Strabismus	Strabismus
Brain MRI/MRS	Posterior WM signal	Delayed myelination, thin CC, cerebral atrophy	Delayed myelination in parietal/periventricular regions	Diffuse hypomyelination, partially progressed myelination, thin CC, WM atrophy, decreased NAA	Diffuse hypomyelination, thin CC, brain atrophy	Periventricular WM change
Spasticity	Hypotonia followed by spasticity	Hypotonia mixed with spasticity	Hypotonia followed by dystonia	Hypotonia then spasticity	Spasticity, dystonia	Hypotonia, spasticity
GI	GERD, poor weight gain, constipation	GERD	GERD	GERD, constipation	Poor appetite, G-tube	GERD, poor weight gain
Other	NR	Facial dysmorphic features	Abnormal BAEP, dysmorphic features	Seizure, ventricular septal defect, decreased bone mineral density, hypothyroidism	Seizure, extended latency in BAEP, peripheral neuropathy	Abnormal EEG, initially diagnosed as CP with PVL

* Specific length not reported.

** Age at most recently reported clinical evaluation. GI: Gastrointestinal. NR: not reported, ASD: Autism Spectrum Disorder, WM: white matter, CC: corpus callosum, PVL: Periventricular leukomalacia, CP: cerebral palsy, GERD: gastroesophageal reflux disease, BAEP: Brain Auditory Evoked Potential, EEG: Electroencephalogram. Female cases are separated from male cases by a vertical double line.

Table 3 –Molecular features of all identified Xq22-*PLP1* deletions to date

Case ID	Sex	Inheritance	Coordinates	Size (Mb)	XCI (blood)	Junction Features
BAB2595	F	<i>de novo</i> ^a	chrX:100866604-103411980	2.5	Random 37:63	2 bp microhomology
BAB2650	F	<i>de novo</i>	chrX:102615641-103309503	0.693	Random 22:78	3 bp microhomology
BAB8120 (D: 268030)	F	<i>de novo</i>	chrX:101029649-106702784	5.6	Skewed >90	22 bp insertion 5 bp direct-repeat
BAB12522	F	<i>de novo</i>	chrX:102066350-105409822	3.3	Random 67:33	124 bp microhomology LINE-LINE
BAB2614	F	Unknown	chrX:102436725-105520605	3	Random 60:40	2 bp microhomology <i>Alu-Alu</i>
BAB8201	M	Maternal	chrX:103029773-103036548	0.0067	NA	None
BAB8204	M	Unknown	chrX:102967297-103038606	0.071	NA	1 bp microhomology
BAB2615	M	Unknown	chrX:102543473-103398234	0.854	NA	1 bp microhomology
Patient 1 [‡]	F	<i>de novo</i>	chrX:100659116-105523589 ^b	4.8	Skewed ^d 79:21	NA
Patient 2 ^{‡3}	F	<i>de novo</i>	chrX:101365862-105847036 ^b	4.4	Uninform-ative	NA
Patient 3 [‡]	F	<i>de novo</i>	chrX:100907884-103982269 ^b	3	Skewed ^c	NA
Patient 5 [‡]	F	<i>de novo</i>	chrX:102959459-103044544 ^b	0.085	Skewed 95:5	NA
BAB1379 [§]	M	Maternal	chrX:102993735-103510087	0.516	NA	18 bp microhomology <i>Alu-Alu</i>
BAB1684 [§]	M	Maternal	chrX:102957288-103314255	0.356	NA	12 bp insertion 3 bp direct-repeat
H152 ^{§5}	M	Maternal	chrX:103009829-103214881 ^e	0.205	NA	32 bp insertion 2 bp microhomology
Patient [¶]	M	Maternal	chrX:103018951-103092038	0.073	NA	1 bp microhomology
Pt. 1 (II-1) [‡]	M	Maternal	chrX:103033333-103066901	0.033	Random II-2: 67:33 I-1: 64:36	2 bp microhomology
Pt. 2 (II-2) [‡]	F					
Pt. 3 (I-1) [‡]	F					

XCI: X-chromosome inactivation. D: DECIPHER ID.

^aPaternal sample not available, but no clinical indication of disease reported for the father.^bCoordinates estimated using aCGH probes.^cRaw data unavailable.^dReported as skewed in the original publication, but does not meet the skewed cutoff in this study (80:20).^eNot all junctions resolved.[‡]Yamamoto et al., 2014[§]Inoue et al., 2002, ⁵Lee et al., 2007[¶]Toritsu et al., 2012[‡]Matsufuji et al., 2013, NA: Not applicable. Novel samples are separated from published samples by a horizontal double line.

Table 4 –

Wilcoxon Rank Sum test P-values of non-B DNA forming motifs in tested regions

			IR	STR	MR	DR	GQ	ZDNA	APR
Group 1	RepZ-within: Distal Vs. Proximal	P-Value (Two-sided)	0.01746 *	1	0.8973	0.6752	0.3031	0.1762	0.726
		P-Value (One-sided)	0.9939	0.5516	0.6018	0.7456	0.8784	0.08808	0.363
Group 2	RepZ Vs. Proximal Flank	P-Value (Two-sided)	0.2309	0.04161 *	0.4357	0.3676	0.5506	0.4886	0.2696
		P-Value (One-sided)	0.8933	0.02081 *	0.8715	0.1838	0.2753	0.7709	0.8756
	RepZ Vs. Distal Flank	P-Value (Two-sided)	1	0.01854 *	0.08597	0.04195 *	0.2896	0.5658	0.0606
		P-Value (One-sided)	0.5	0.00927 *	0.04298 *	0.02098 *	0.1448	0.7331	0.973
Group 3	Hotspot Vs. Proximal Flank	P-Value (Two-sided)	0.01465 *	0.5006	0.9292	0.9285	0.3295	0.05537	1
		P-Value (One-sided)	0.9943	0.2503	0.5706	0.5712	0.8572	0.02769 *	0.5375
	Hotspot Vs. Distal Flank	P-Value (Two-sided)	0.06151	0.04525 *	0.5049	0.5047	0.2892	0.3643	0.1811
		P-Value (One-sided)	0.9749	0.02262 *	0.2525	0.2524	0.1446	0.1821	0.9236

The three tested groups were: 1) The distal half of RepX-i1010 (70 kb) tested against the proximal half of RepX-i1010 (70 kb), 2) 140 kb of RepX-i1010 tested against 140 kb of proximal sequence and 140 kb of distal sequence, and 3) A 90 kb genomic instability hotspot tested against 90 kb of proximal sequence and 90 kb of distal sequence. Different groups are separated by a horizontal double line. APR: A-phased repeats, DR: Direct repeats, ZDNA: Z-DNA, GQ: G-quadruplexes, IR: Inverted repeats, MR: Mirror repeats, STR: Short tandem repeats.

* P-value <0.05.