

A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders

Jennifer A. Lee,¹ Claudia M.B. Carvalho,¹ and James R. Lupski^{1,2,3,*}

¹Department of Molecular and Human Genetics

²Department of Pediatrics

Baylor College of Medicine, Houston, TX, 77030, USA

³Texas Children's Hospital, Houston, TX, 77030, USA

*Correspondence: jlupski@bcm.tmc.edu

DOI 10.1016/j.cell.2007.11.037

SUMMARY

The prevailing mechanism for recurrent and some nonrecurrent rearrangements causing genomic disorders is nonallelic homologous recombination (NAHR) between region-specific low-copy repeats (LCRs). For other nonrecurrent rearrangements, nonhomologous end joining (NHEJ) is implicated. Pelizaeus-Merzbacher disease (PMD) is an X-linked dysmyelinating disorder caused most frequently (60%–70%) by nonrecurrent duplication of the dosage-sensitive proteolipid protein 1 (*PLP1*) gene but also by nonrecurrent deletion or point mutations. Many *PLP1* duplication junctions are refractory to breakpoint sequence analysis, an observation inconsistent with a simple recombination mechanism. Our current analysis of junction sequences in PMD patients confirms the occurrence of simple tandem *PLP1* duplications but also uncovers evidence for sequence complexity at some junctions. These data are consistent with a replication-based mechanism that we term FoSTeS, for replication Fork Stalling and Template Switching. We propose that complex duplication and deletion rearrangements associated with PMD, and potentially other nonrecurrent rearrangements, may be explained by this replication-based mechanism.

INTRODUCTION

Structural variation of the human genome can be associated with disease traits or represent benign copy-number variation (CNV) (Lee and Lupski, 2006; Lupski, 2007). Genomic disorders are a group of human genetic diseases caused by DNA rearrangements that result in the gain, loss, or disruption of a gene(s) for which dosage is critical

(Lupski, 1998; Lupski and Stankiewicz, 2005, 2006). These disorders are often characterized by similar genomic features in the rearrangement-susceptible regions but represent a wide spectrum of unrelated clinical entities.

The recombination mechanisms nonallelic homologous recombination (NAHR) and nonhomologous end joining (NHEJ) have been shown to underlie the rearrangements causing genomic disorders (Shaw and Lupski, 2004), with the former accounting for the majority. NAHR can use either region-specific low-copy repeats (LCRs, or segmental duplications) or sometimes repetitive sequences (e.g., *Alu*) as homologous recombination substrates, yielding recurrent events with clustered breakpoints. Much less is known about the mechanism(s) for nonrecurrent rearrangements. NHEJ has been implicated for several such disorders (Inoue et al., 2002; Padiath et al., 2006; Shaw and Lupski, 2005; Toffolatti et al., 2002). However, a number of disease-associated rearrangements are not explained readily by either the NAHR or simple NHEJ recombinational mechanisms.

To elucidate the molecular mechanism responsible for nonrecurrent rearrangements, we focused on the duplications associated with the genomic disorder Pelizaeus-Merzbacher disease (PMD). PMD is an X-linked recessive dysmyelinating disorder caused most frequently (~60%–70%) by nonrecurrent duplication of the dosage-sensitive proteolipid protein 1 (*PLP1*) gene but also by nonrecurrent *PLP1* deletion or point mutations (Inoue et al., 1999; Sistermans et al., 1998). *PLP1* rearrangement breakpoint junction analyses reveal microhomology or a lack of homology (Inoue et al., 2002; Woodward et al., 2005) characteristic of nonrecurrent rearrangement junctions; as reported in the literature, such products of recombination are consistent with NHEJ repair. However, other genomic rearrangement mechanisms requiring microhomology are also possible.

We performed high-resolution oligonucleotide (oligo) array comparative genomic hybridization (CGH) and breakpoint sequence analyses of PMD-associated *PLP1* nonrecurrent duplications. To our knowledge, there have been no such combined analyses performed to date on

any genomic disorder due to nonrecurrent duplications. From our analyses, we identified the extent of genomic gain/loss and determined junction sequences in PMD patients with different-sized (~200 kb to ~7 Mb) genomic rearrangements. To our surprise, we uncovered interspersed stretches of DNA of normal copy number or triplicated but contained within duplicated sequence and sequence complexity at the junctions. These findings suggest parsimoniously an alternative mechanism involving errors of DNA replication, instead of the more generally considered meiotic recombination mechanisms (NAHR and NHEJ). Our model of replication Fork Stalling and Template Switching (FoSTeS) can explain the complex duplication and deletion rearrangements associated with PMD and potentially other nonrecurrent rearrangements of the human genome.

RESULTS

Semiquantitative PCR Delimits Apparent Duplication Junction Regions

By semiquantitative multiplex PCR analyses, using our previously published BAC array comparative genomic hybridization (CGH) data (Lee et al., 2006a) as a starting point, we delineated *PLP1* duplication boundaries effectively for all patients studied (Table 1). From the positions of the furthest-apart primers that detected a DNA duplication (i.e., primers closest to the duplication boundaries and contained within the duplicated region), we calculated the approximate size of the duplicated region(s) for each patient (Table 1) with greater resolution than possible by BAC array CGH analysis. The precise locations of those breakpoints were narrowed to within ~160 bp to 64 kb (Table 1).

Novel Duplication Junctions Identified by Long-Range PCR and DNA Sequencing

Long-range PCR was attempted initially, under the assumption of direct or tandem duplication (Inoue et al., 1999), on eight selected patient samples for which there was no evidence of complex rearrangement by previous methods (Lee et al., 2006a) (Table 1). The furthest-apart semiquantitative multiplex PCR primers (outward-facing relative to reference sequence) that detected a duplicated genomic sequence were used to amplify duplication junctions (Table S1 available online). Junction sequence was successfully amplified and sequenced for two patients, BAB1258 and 1264. No significant homology was found within 2 kb to either side of these junctions in the reference human genome sequence by BLAST2 analysis.

Southern Blot Analyses Detect Junction Fragments in the Remaining Patients

For the remaining six patients (BAB1707, 2448, 1301, 1290, 1275, and 1327) for whom no duplication junction amplification product was detected, we performed genomic Southern blot analysis (Table S2 and Figure S1) to confirm junction location. Changes in the sizes of the restriction digest fragments relative to those predicted

by the reference sequence were observed, reflecting (1) the addition or deletion of sequence at a novel junction, (2) the relative position of the breakpoint to the nearest restriction site, or (3) polymorphisms (i.e., RFLPs) at the restriction sites.

Enrichment of novel restriction fragments followed by inverse, TOPO linker-mediated, VISA, and DOP PCR methodologies (See Supplemental Experimental Procedures) did not allow us to identify the breakpoints for these remaining six patients.

Oligonucleotide Array CGH Enables Higher-Resolution Breakpoint Junction Mapping

We designed a custom Agilent array consisting of 44,000 (44K) oligos tiling an ~10 Mb genomic region including *PLP1*, at a resolution of ~1 interrogating oligo every 500 bp. By this analysis we determined that the majority of our patients may have undergone complex rearrangements. We found evidence for the expected simple tandem duplications as well as interrupted duplications in which stretches of duplicated DNA were punctuated by stretches of DNA with no copy-number alteration (Figure 1). Further, as observed by other groups (Wolf et al., 2005), copy-number changes suggestive of triplicated or quadruplicated sequences contained within duplicated segments, for which the \log_2 (Cy3/Cy5) ratios approached 1.6 and 2.0, respectively, rather than the expected 1.0 for duplicated regions, were identified (Figure S2). These data are summarized in Figure 2. Because interrogating oligos were designed to exclude repeat-masked DNA, some coverage gaps exist for which no copy-number information could be obtained. Nevertheless, the genome resolution enabled by oligo array CGH analysis was significantly greater than that achieved previously by BAC array CGH (Lee et al., 2006a) and semiquantitative multiplex PCR methodologies. This analysis revealed rearrangements more complex than simple tandem duplications in at least 11 of 17 (65%) patients tested.

However, once oligo array CGH was performed and long-range PCR was attempted based on more refined copy-number boundary delineation, we experienced similar difficulty in amplifying across duplication junctions, except for patient BAB2448 (discussed below). Despite our high-resolution analysis, for the majority of patients amplification of the rearrangement junctions was not achieved, making simple tandem duplications improbable and suggesting complex rearrangements in these patients. Overall, our success rate (3/17 = 18%) of effectively sequencing novel *PLP1* duplication junctions is similar to that of other groups (Woodward et al., 2005).

Junction Sequences Reveal Complex Rearrangements

From our junction analysis, we conclude that the rearrangement in patient BAB1258 may be a simple tandem duplication (Figures 3A and 3B), as has been identified for most sequenced *PLP1* duplications (Inoue et al., 1999; Woodward et al., 2005). Further, two base pairs of

Table 1. Junction Analysis of Selected *PLP1* Duplications from Patients with PMD

Patient (BAB#)	Rearrang. Span ^a	Apparent Junction Position(s) (Build 35)	No. of Junctions ^b	Nature of Rearrang. ^a
1264 ^c	~180 kb	102883770::102370246; 102730277::102726973; 102727187::102712464	3	Complex
1258 ^c	~190 kb	102665716::102855124	1	Noncomplex
1707 ^{c,d}	~400 kb	102640944-102647753; 103049282-103049439	≥ 1	Noncomplex
2448 ^{c,d}	~420 kb	103005282::102579802; 102782479-102783321; 102790832-102791889; 102950000-102963000; 102981000-102982000	≥ 3	Complex
H152 ^c	~430 kb ^e	102817314::103369461; 103369494::103246266; 103096667-103099990; 103110000-103138000; 103009450-103021026; 103029644-103049223	≥ 4	Complex
1334 ^f	~460 kb	102623208-102624396; 103029776-103049223; 103075656-103079184; 103108655-103145524	≥ 2	Complex
1482 ^f	~500 kb	102587535-102589735; 103029644-103049223; 103071867-103073186; 103110848-103143639	≥ 2	Complex
1301 ^{c,d}	~560 kb	102464125-102465370; 102636000-102637000; 102647000-102648000; 102892309-102893118; 102893904-102894913; 103029807-103049439	≥ 3	Complex
1261 ^f	~580 kb	102526363-102528665; 103109525-103146062	≥ 1	Complex
1305 ^f	~600 kb	102462505-102464066; 102636862-102638227; 102646604-102647086; 103016000-103049223	≥ 2	Complex
1282 ^f	~660 kb	102471795-102480245; 103080000-103143639	≥ 1	Complex
1290 ^{c,d}	~800 kb	102333875-102335628; 102917881-102942038; 103029644-103049223; 103103514-103109476	≥ 3	Complex
1275 ^{c,d}	~970 kb	102164207-102167232; 102231681-102235383; 102256404-102258043; 102782849-102783912; 102791320-102792583; 103103514-103108463	≥ 3	Complex
1420 ^f	~1 Mb	102524657-102527322; 103559197-103560962	≥ 1	Noncomplex
2389 ^f	~4 Mb	99191746-99210640; 102880716-102886748; 102940000-102959000; 102985000-102990000; 103029244-103063198; 103107915-103143639	≥ 4	Complex
1327 ^{c,d}	~7.3 Mb	99252780-99254843; 106569573-106570347	≥ 1	Noncomplex
2396 ^f	7.8+ Mb ^g	100196711-100201345; 103028718-103071407; 103107915-103143639	≥ 3	Complex

Rearrang., rearrangement; ::, exact breakpoint junction positions as determined by sequence analysis; other values are junction regions narrowed by array CGH analysis.

^a As determined by semiquantitative multiplex PCR, array CGH, previous PFGE analyses, and/or DNA sequencing.

^b Number of junctions sequenced/predicted based on array data.

^c Selected for breakpoint delineation by semiquantitative multiplex PCR.

^d Selected for breakpoint analysis by inverse, TOPO-linker mediated, VISA, and DOP PCR.

^e Harbors a deletion of ~190 kb including *PLP1*.

^f Analyzed only by array CGH and long-range PCR.

^g Duplication extends beyond region interrogated by array.

microhomology were identified at this junction, consistent with NHEJ repair as has been proposed previously for *PLP1* rearrangements (Inoue et al., 2002; Woodward et al., 2005).

By oligo array CGH analysis, we determined that there is an interrupted duplication that occurred in patient BAB2448, as there are three duplicated segments that

flank smaller segments with no copy-number alteration. Centromeric to telomeric, we observed a ~200 kb duplicated segment, followed by ~9 kb with no copy-number change, an ~170 kb duplicated segment, an ~15 kb segment with no copy-number change, and finally a 20 kb duplicated segment. Overall, the duplications span ~420 kb and appear to be complex. One junction was identified for

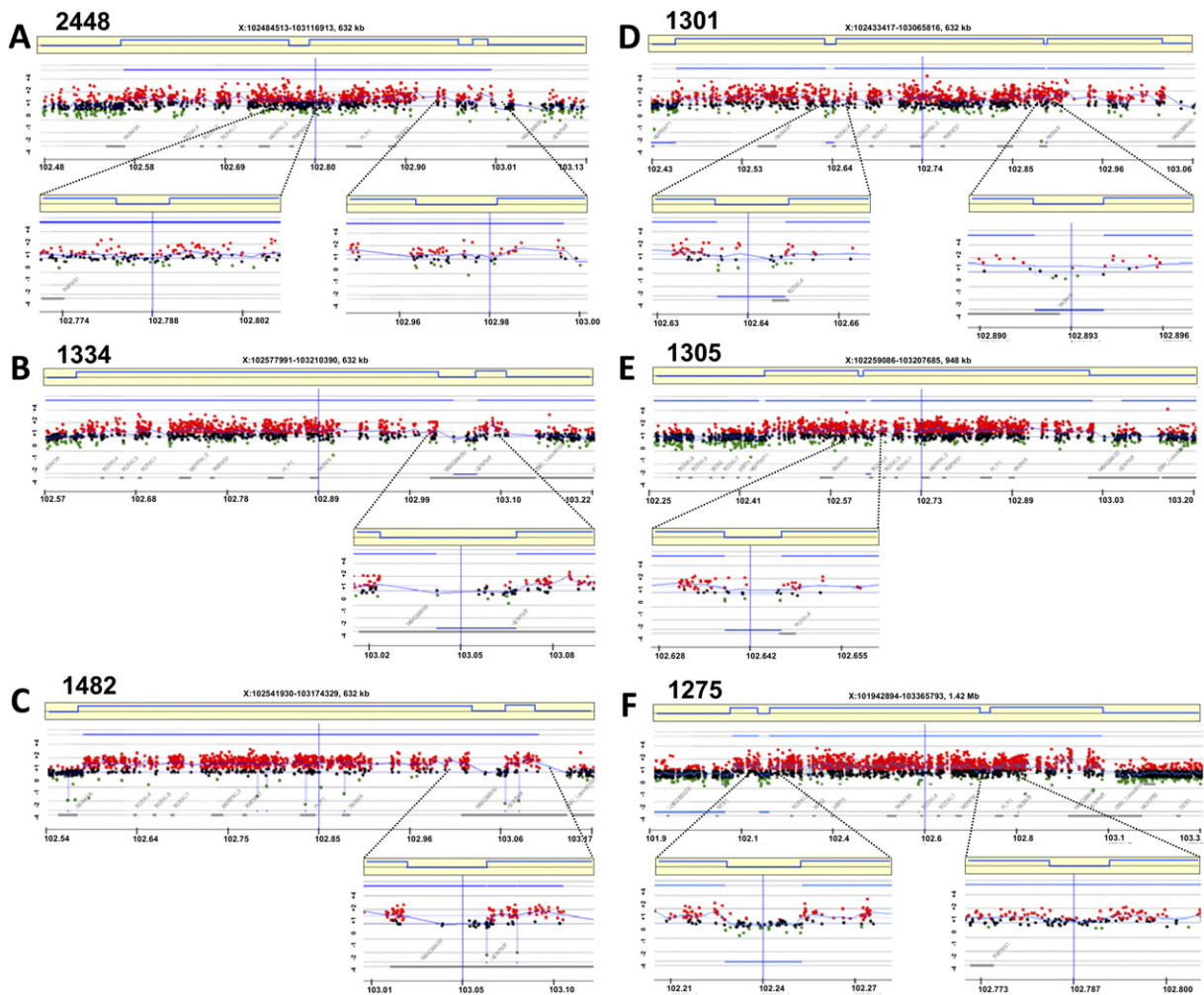


Figure 1. Complex Interrupted Duplications Revealed by Oligonucleotide Array CGH

Copy-number alteration data for patients: (A) 2448, (B) 1334, (C) 1482, (D) 1301, (E) 1305, and (F) 1275 are shown. Array data for patients 1264 and 1258 (Figure S3), 1707, 1261, 1282, and 1420 did not reveal obvious complex rearrangements (not shown). Displayed are copy-number changes observed, with enlargements of each area depicting regions of no copy-number change flanked by duplicated sequence below. Red data points indicate copy-number gains, green data points losses, and black data points no copy-number change. In each horizontal yellow box above, blue lines represent an average of the data points; \log_2 (Cy3/Cy5) values are on the y axis; genomic position in megabases (Mb); NCBI build 35) is on the x axis.

this patient, with five base pairs of microhomology (Table S3A; Figures 3C and 3D). Visually, it appears that a DNA replication fork “skipped backward” and then “skipped forward” along the chromosome (Figure 3E), similar to the proposed mechanism for the rearrangement in patient BAB1264 (see below).

The duplication rearrangement that occurred in patient BAB1264 also appears to be complex, as additional short sequences, or “junctional” sequences, were found at the duplication junction; a total of three junctions were identified for this patient (Figure 4). Interestingly, these junctional sequences (32 bp and 215 bp) were found to have 100% shared identity to regions contained within the duplicated segment. More interesting are the order and direct orientation, with respect to the positive strand, of these two junc-

tional sequences, as graphically illustrated in Figure 4C. Visually, it appears that a DNA replication fork skipped backward along the chromosome. Two or three base pairs of microhomology were identified at these junctions.

The striking assembly of unexpected additional sequences from discreet genomic positions at the duplication junction in patient BAB1264 and the complex rearrangement in patient BAB2448 are difficult to explain by either an NAHR or simple NHEJ recombinational mechanism.

A Complex Deletion Junction Is Inconsistent with a Simple Recombination Model

We reanalyzed the rearrangement of one previously published patient with a complex *PLP1* deletion (Inoue et al.,

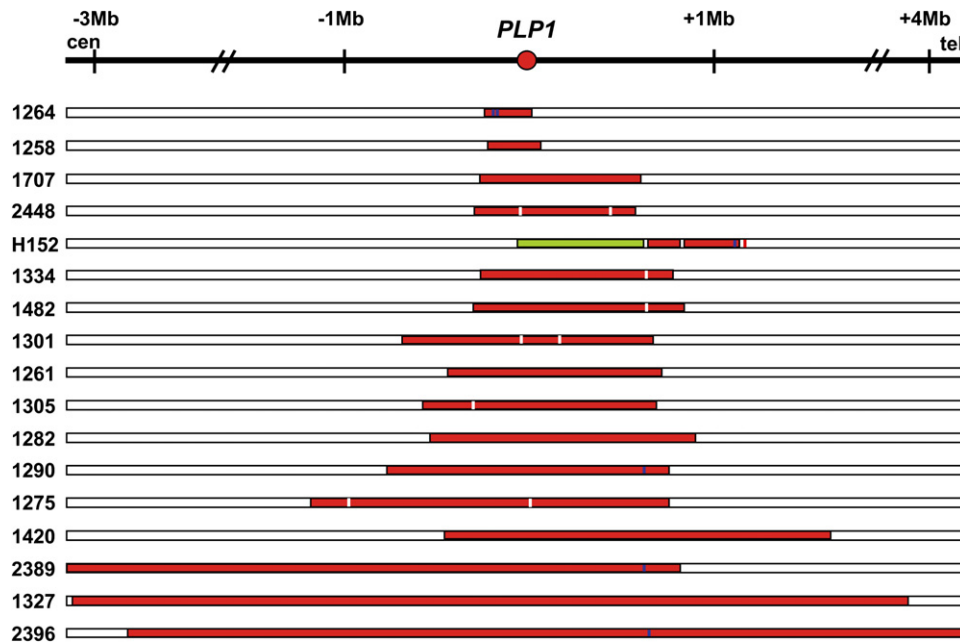


Figure 2. Comparison of Complex Genomic Rearrangements Involving *PLP1*

Summary of X chromosome rearrangements including *PLP1* (red circle); left, patient numbers. Duplications are shown as red bars, deletions in green, triplications in dark blue, and no copy-number change in white. Centromeric (cen) and telomeric (tel) positions are shown. Figure not drawn to scale; approximate positions given in megabases (Mb) relative to *PLP1*.

2002). For this patient, two junctions were previously sequenced, revealing two base pairs of microhomology at each with some additional unexpected sequences in an inverted orientation (discussed below). From these data, we predicted that there was at least one remaining junction to be sequenced, as the distal boundary of the deletion had not been identified yet at the sequence level. Thus, a higher-resolution analysis of this genomic region was applied.

Oligo array CGH analysis was performed on the DNA of this patient's carrier mother (H150), who harbors an identical genomic rearrangement to her son. Such an analysis yields the same copy-number results, as the mother's DNA is normalized to a female control. In this family, we confirmed the deletion of ~ 190 kb and also identified that it is followed by a stretch of ~ 9 kb with no copy-number change and duplications (~ 80 kb and ~ 110 kb) interrupted by a short stretch of DNA for which there is no copy-number change (Figures 5A and 5B). Furthermore, we were able to reinterpret the junctions previously reported based on our new findings. Previously, the proximal (centromeric) breakpoint at the start of the ~ 190 kb deletion was found to be joined to a 34 bp segment that normally maps ~ 760 kb distal (telomeric) to *PLP1*, but in an inverted orientation; this was followed by another junction with a stretch of DNA of at least 5 kb in length that normally maps ~ 640 kb distal to *PLP1*, also in an inverted orientation. Based on NCBI build 35, a more recent version of the draft sequence than that available when an analysis of patient H152 was originally published, we

found that this ≥ 5 kb stretch of DNA currently maps ~ 393 kb distal to *PLP1* and aligns to the distal end of the duplication identified by our oligo array CGH analysis. In other words, this ≥ 5 kb DNA segment is actually part of an ~ 110 kb duplication. A graphical illustration of the rearrangement for this family is found in Figures 5C and 5D.

Rearrangement Breakpoint Junctions and Genome Architecture

The genomic region surrounding *PLP1* contains complex genomic architecture, such as abundant LCRs of different sizes (~ 1 – 122 kb) with high sequence identity and in various orientations (Inoue et al., 1999; Lee et al., 2006a; Woodward et al., 2005), which may render susceptibility to genomic rearrangement. Accordingly, we sought to determine whether such complex architecture may play a role in replication fork stalling and template switching. We examined the genomic positions where the proposed template switching has occurred in our patient cohort and superimposed these locations on the regional genome architecture. Most occur in proximity to LCR-PMDs (i.e., LCR-PMDA, B, C, D, and E2), where a replisome may encounter complex architecture (Figure 6A). It appears that the proposed FoStEs events may preferentially occur within or adjacent to complex genomic architecture that has symmetrical features that may enable cruciform structures to form, suggesting that these events may be stimulated by LCRs. We suggest that the unusual genome architecture surrounding *PLP1* may confuse the DNA

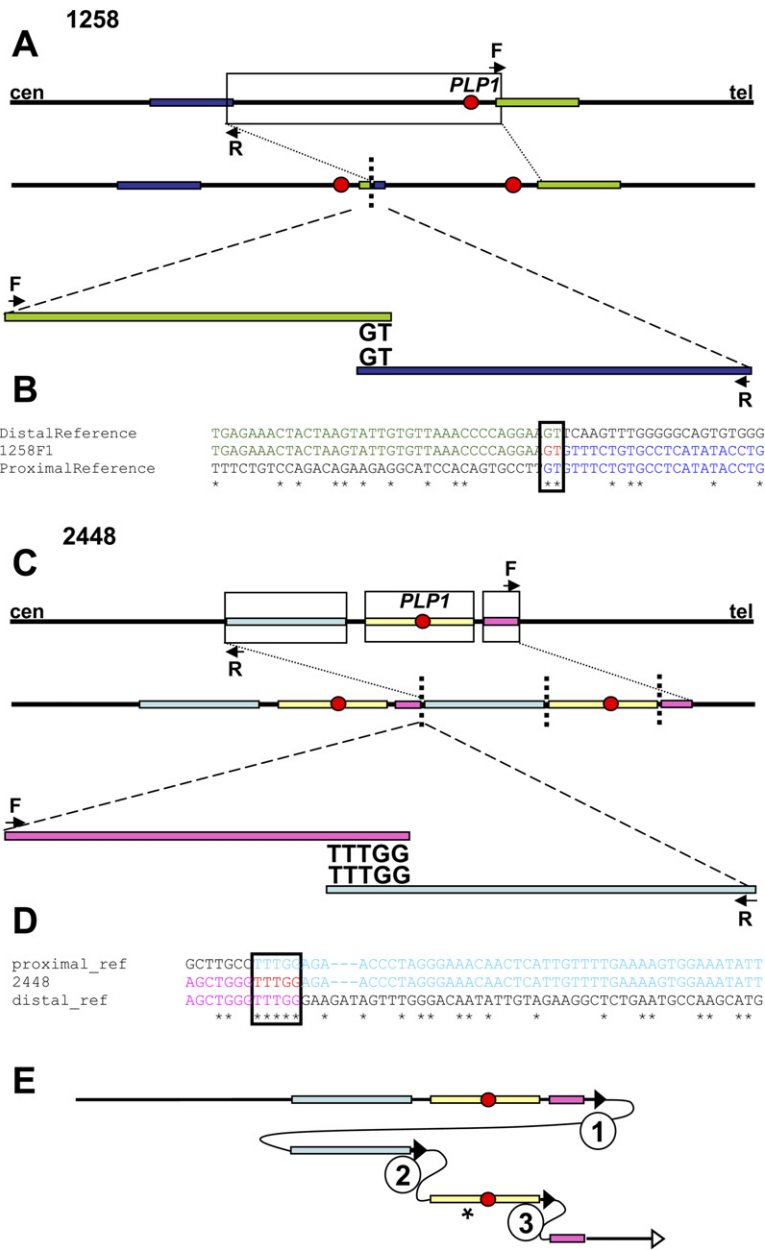


Figure 3. Junction Analysis for Patients BAB1258 and 2448

(A and C) Duplication junction(s) (vertical dotted line(s)) for patients BAB1258 and BAB2448, respectively, are displayed relative to reference sequence. Duplicated region is boxed. Two base pairs (BAB1258) and five base pairs (BAB2448) of microhomology were found at the breakpoint junctions after amplification with outward-facing primers (F and R). (B and D) Sequence data for duplication junctions; microhomologies (red letters) are boxed. Proximal (blue for patient BAB1258, light blue for patient BAB2448) and distal (green for patient BAB1258, pink for patient BAB2448) reference sequences are marked. (E) Illustration of the predicted order, origins, and relative orientations of duplicated sequences (light blue, yellow, and pink) for patient BAB2448. Arrowheads show direction of DNA relative to the positive strand; filled arrowheads with circled numbers below represent a FoSTeS event; open arrowhead signifies resumption of replication on the original template. Asterisk (*) indicates an unknown orientation of the ~170 kb segment. Proximal (centromeric) and distal (telomeric) are in relation to PLP1 (red circle).

replication machinery, causing one or multiple replication fork stalling and switching events before resuming replication on the original DNA template.

In the vicinity of these LCR-PMDs there are also various reported CNVs (Database of Genomic Variants, <http://projects.tcag.ca/variation/>) including duplications and inversions (Lee et al., 2006b). As the genomic boundaries of these CNVs do not coincide with those of the copy-number changes and FoSTeS event locations identified for the patients in our cohort (Figure 6A), and as the rearrangements were ascertained by virtue of the conveyed PMD phenotype, these patient-associated complex rearrangements cannot be explained by benign CNVs.

DISCUSSION

Junction Sequence Suggests Replication Mechanism

We propose that the rearrangements in patients BAB1264 and 2448 occurred via a replication-based mechanism, which we term FoSTeS, and not via a recombination-based mechanism (Figures 6B–6E). This mechanism is modified from a model suggested for *Escherichia coli* gene amplification (Slack et al., 2006) and similar to that suggested for smaller rearrangements mediated by short direct or inverted repeats that potentially form non-B DNA conformations in various organisms (Chen

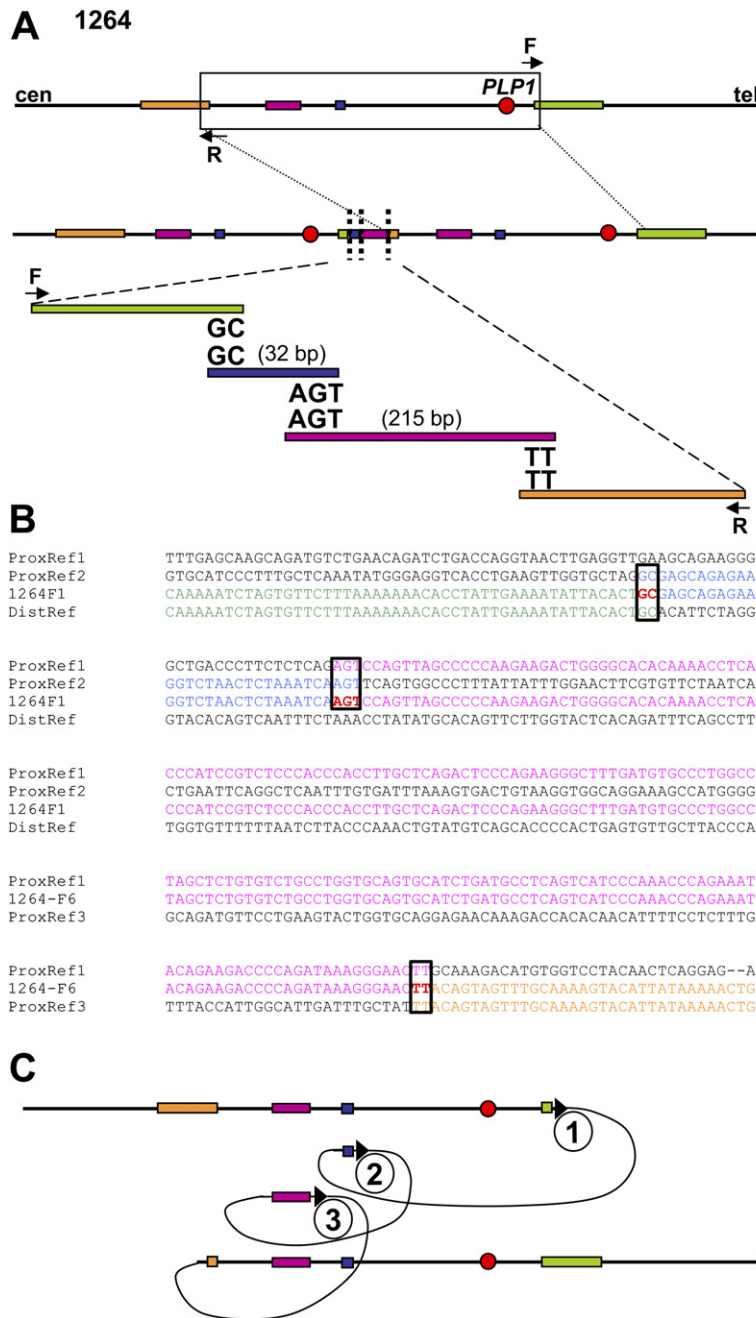


Figure 4. Junction Analysis for Patient BAB1264

(A) Duplication junctions (vertical dotted lines) for patient BAB1264 are displayed relative to reference sequence, with the duplicated region boxed. Two or three base pairs of microhomology were found at the breakpoint junctions after amplification with outward-facing primers (F and R).

(B) Sequence data for the duplication junction; microhomologies (red letters) are boxed. Proximal reference sequence 1 is marked in pink, proximal reference sequence 2 in blue, proximal reference sequence 3 in orange, and the distal reference sequence in green.

(C) Illustration of the order, origins, and relative orientations of junctional (pink and blue) and boundary reference sequences (orange and green) for patient BAB1264. Arrowheads show direction of DNA relative to the positive strand; filled arrowheads with circled numbers below represent a FoSTeS event; open arrowhead marks resumption of replication on the original template. Proximal (centromeric) and distal (telomeric) are in relation to *PLP1* (red circle).

et al., 2005; Gordon and Halliday, 1995; Ohshima et al., 1992; Trinh and Sinden, 1991; Wojciechowska et al., 2005), including the generation of inversions accompanied by duplications and deletions (Gordon and Halliday, 1995), and for spontaneous duplication events in eukaryotic genome evolution (Koszul et al., 2004). In the current model, we propose that during DNA replication, the replication fork stalls or pauses at DNA lesions resulting from the genomic instability at/near regional LCRs. After the replication fork stalls or pauses, the lagging strand serially disengages and switches to another nearby template at

another active replication fork, which could be advancing in either direction (5' to 3' or 3' to 5' with respect to the leading strand); this switch would require only microhomology. DNA would be copied at this second (or third, etc.) sequence, and the nascent strand might disengage again after a short time. The forks may be in physical proximity but separated by sizeable linear distances, even megabases away, thus enabling the template-driven joining of different sequences from discrete genomic positions. Eventually, the replication fork would proceed normally.

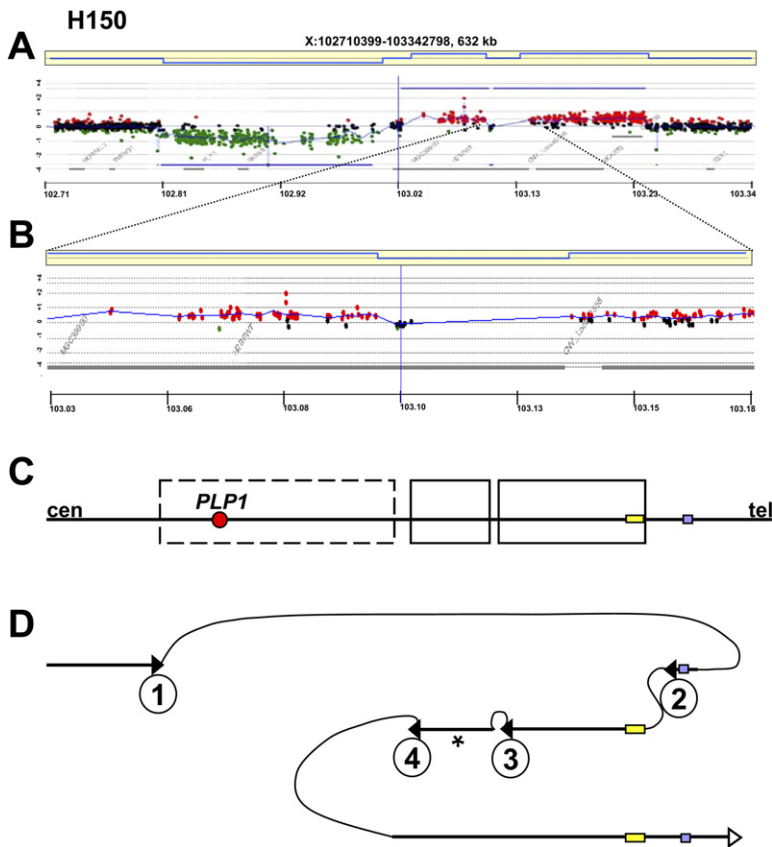


Figure 5. Complex Rearrangement Detected by Oligonucleotide Array CGH Analysis of a Previously Reported Patient

Shown are data for the carrier mother of patient H152 (Inoue et al., 2002).

(A) An ~190 kb deletion is followed by an ~9 kb segment with no copy-number change and an interrupted ~190 kb duplication.

(B) Enlargement of the ~190 kb interrupted duplication. In each horizontal yellow box above, blue lines represent an average of the data points. Red data points indicate copy-number gains, green data points losses, and black data points no copy-number change. Log₂ (Cy3/Cy5) values are on the y axis; genomic position in megabases (Mb; NCBI build 35) is on the x axis.

(C) Relative to reference sequence, junctional sequences are highlighted; boxed are deleted (dashed line) and duplicated (solid line) regions.

(D) Illustration of the order, origins, and relative orientations of junctional sequences for patient H152. Asterisk (*) indicates an unknown orientation of the ~80 kb segment. Arrowheads show direction of DNA relative to the positive strand: filled arrowheads with circled numbers below represent a FoSTeS event; open arrowhead marks resumption of replication on the original template. *PLP1* is indicated by the red circle.

For patient BAB1264, we propose that fork stalling or pausing, disengaging, and switching of templates may have occurred twice, resulting in two junctional sequences, 32 bp and 215 bp, before resuming replication on the original template (a total of three FoSTeS events, or FoSTeS × 3; Figure 4C). For patient BAB2448, we propose that this phenomenon may have also occurred a total of three times (FoSTeS × 3; Figure 3E); first the proximal ~200 kb duplicated segment is copied, next the ~170 kb duplicated segment, and then the ~20 kb segment before resuming replication on the original template. For patient BAB1258, although the rearrangement may be explained by simple tandem duplication via NHEJ, it is feasible that a simplified FoSTeS event may have occurred (FoSTeS × 1).

Physically, a close interaction between the participating DNA molecules would be necessary, potentially aided by the presence of highly homologous regional LCRs, cruciforms, or other non-B DNA structures that may form (Bacolla and Wells, 2004; Lee et al., 2006a). This mechanism may explain some of the more complex *PLP1* duplication and deletion rearrangements that have been reported in the literature (Inoue et al., 2002; Wolf et al., 2005; Woodward et al., 2005), for which (1) additional duplications nearby the duplicated segment containing *PLP1* were present, (2) multiple rearrangement junctions were identified in a single patient, (3) sequences found

at the junctions were different in origin from those predicted by dosage analysis, (4) junctional sequences were different in origin and/or orientation from those expected from simple tandem duplication, and (5) evidence for three or more copies of *PLP1* was obtained in association with a more severe PMD phenotype.

Judging from the properties of the junctional sequences and microhomologies identified, the rearrangement in patient H152 (and his carrier mother) might have occurred via an NHEJ repair mechanism, as originally suggested (Inoue et al., 2002); however, this rearrangement is also consistent with what may occur by a FoSTeS mechanism. Visually, a skipping forward along the chromosome is evident, followed by a skipping backward three times, for a total of four FoSTeS events (FoSTeS × 4; Figure 5D). Furthermore, we confirmed by PCR and DNA sequencing that the 34 bp junctional sequence (Figure 5C, light blue segment) is present in two copies, one copy at the junction and the other copy in its normal genomic position (data not shown). Thus, not only do the junctional sequences found for this family share identity with regions distal to the deleted interval in an inverted orientation, but they are also found to be duplicated. This finding is consistent with the rearrangement in this patient occurring via a replication-based mechanism (Figure 5D).

Such complex rearrangement products are consistent with a mechanism of serial disengaging of a nascent

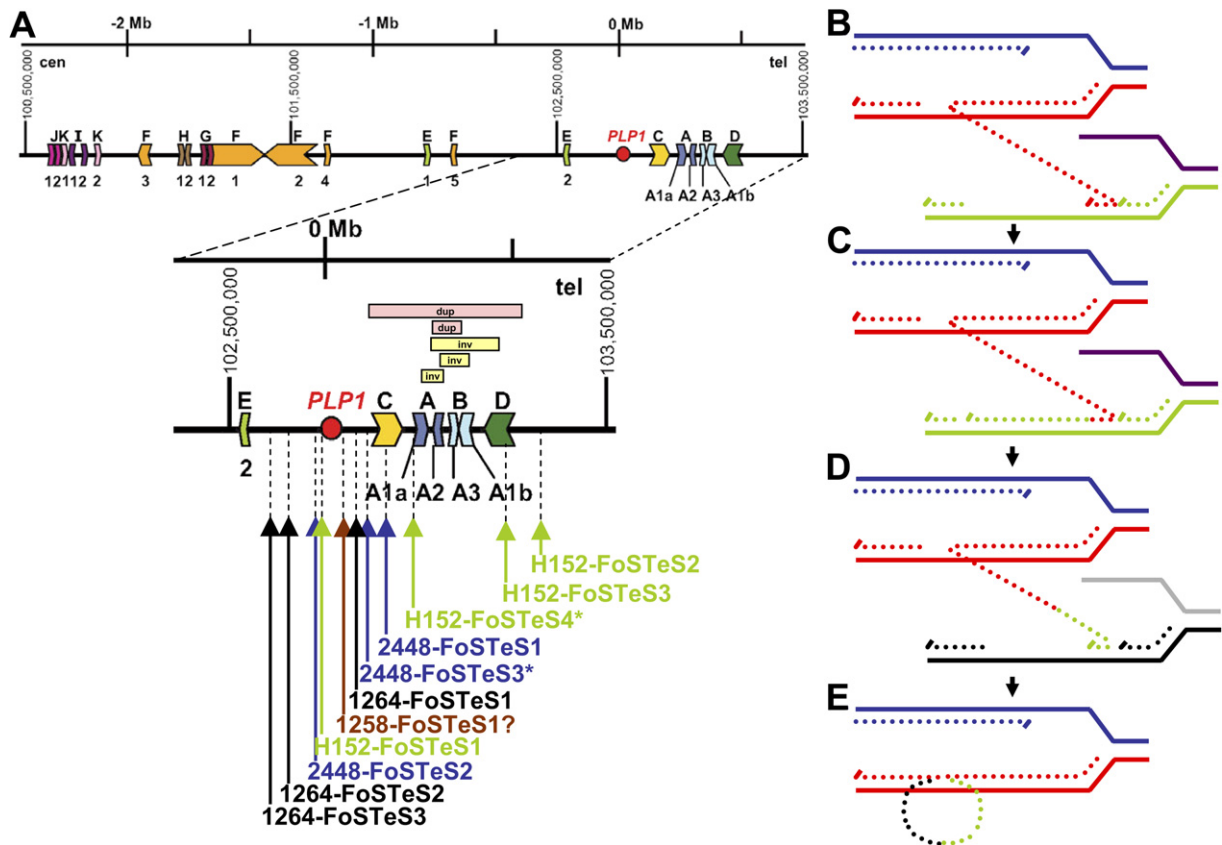


Figure 6. Alignment of Proposed FoSTeS Events with Complex Genomic Architecture and Proposed FoSTeS Model

(A) Shown are genomic locations of each FoSTeS event in the context of regional LCRs (Lee et al., 2006a). Two of eleven FoSTeS events (2448-FoSTeS1 and H152-FoSTeS3) occurred within an LCR; one additional event (H152-FoSTeS4) may have occurred either just centromeric to or just within LCR-PMDC. Positions of CNVs found in the Database of Genomic Variants are shown as pink (duplication) and yellow (inversion) horizontal bars. Asterisk (*) denotes that FoSTeS event position was determined under the assumption of directly ordered duplications. After encountering a DNA lesion (B), one replication fork (dark blue and red, solid lines) with a lagging strand (red, dotted line) would invade a second fork (purple and green, solid lines), followed by (C) DNA synthesis (green, dotted line). After the fork disengages (D), the original fork (dark blue and red, solid lines) with its lagging strand (red and green, dotted lines) could invade a third fork (gray and black, solid lines). Dotted lines represent newly synthesized DNA. Serial replication fork disengaging and lagging strand invasion could occur several times before (E) resumption of replication on the original template.

strand and reinitiation of DNA synthesis before the completion of DNA synthesis on the original template; this mechanism may yield two or more copies of junctional sequence that could be in either orientation. In contrast, if we had identified segments of DNA that only appear at the junctions and not in the normal genomic position, this would have suggested a recombination-based model. We propose that finding segments of DNA both at the junctions and at the normal genomic position is more suggestive of a template-driven event or replication-based model. On the basis of this model, we predict that the additional fragments at the junctions (32 bp and 215 bp) identified for patient BAB1264 will be found in three copies (Figures 4A and 4C).

Obtaining further junction sequence information is challenged by the complexity of the mechanism and the difficulty of predicting the template structure from which to

attempt amplification for experimental determination of junction sequence.

Replication Model for Genomic Rearrangements

Compared to the general mechanism we have proposed for the various *PLP1* rearrangements causing PMD (Lee et al., 2006a), our data herein suggest a single-strand DNA lesion as the initiating damage that causes replication fork stalling, rather than a double-strand break (DSB) that may be a more likely initiating substrate for a recombination-based mechanism. Repair of the DNA molecule after successive template-switching events would occur via DNA Ligase I to join Okazaki fragments (Prigent et al., 1994). Alternatively, an initiating DSB that causes replication fork collapse followed by repair via break-induced replication (BIR) with multiple rounds of strand invasions (Narayanan et al., 2006; Smith et al., 2007),

errors of replication fork repair involving a template switch (Goldfless et al., 2006), and template switches involving leading strands (Goldfless et al., 2006) cannot be ruled out. As the nonrecurrent rearrangements causing PMD have been reported to occur during spermatogenesis (Inoue et al., 1999; Mimault et al., 1999), we suggest that the replicative error may have occurred during mitotic or meiotic DNA synthesis during the early stages (Blanco-Rodríguez, 2002a, 2002b), in contrast to a recombination model that presupposes that the causative event occurs during the meiotic pairing of chromosomes.

Complex Rearrangements Evade Junction Sequence Determination

From our oligo array CGH analysis, we detected complex rearrangements in the majority (65%) of patients tested. However, as the complex rearrangement identified by PCR and sequence analyses in patient BAB1264 was not detectable by oligo array CGH methodology (Figure S3), it remains possible that other patients who appear to have simple tandem duplications by oligo array CGH analysis might actually have experienced more complex cryptic rearrangements. This may explain why breakpoint junctions were not successfully identified for the majority of patients in our cohort even when they appear by oligo array CGH analysis to have simple tandem duplications. Array CGH enables detection of genomic regions of gains and losses but provides neither genome position nor orientation information.

When long-range amplification across a duplication junction fails to yield a PCR product, there are many potential explanations. First, it is possible that the duplication is interrupted (i.e., the duplicated segment is inserted near the original sequence) (Lee et al., 2006a). Second, the duplication could be inverted in orientation; amplification with outward-facing primers across a duplication junction presumes simple tandem duplication. Third, complex duplications might have ends/boundaries that are not in the orientation or location anticipated based on a simple recombination model (Inoue et al., 2002; Woodward et al., 2005). Thus, although duplication (or deletion) boundaries may be delimited successfully by a variety of methods, it is possible that the junction(s) may not be amplified successfully. While our semiquantitative multiplex PCR assay successfully predicted the relative locations of junction sequences in all patients tested (i.e., Southern analysis detected junction fragments at predicted locations), because the duplication junctions were not successfully obtained by numerous standard methods despite the enrichment of DNA templates for junction sequence, we suggest that these rearrangements are not simple tandem duplications but rather are complex in nature.

The challenge we experienced in obtaining junction sequence for the majority of our cohort is not surprising given the regional genomic complexity involving LCRs surrounding *PLP1* (Lee et al., 2006a). This regional complexity limited the resolution of our assays and potentially contributed to a high level of genomic instability, possibly

causing a predisposition of the *PLP1* region to complex nonrecurrent rearrangements. This challenge in obtaining junction sequence is also reflected in the comparable success rates reported by other groups.

However, from the novel junction sequences we obtained and from our further experimentation and re-analysis of a previously reported PMD case with a complex rearrangement, we were able to (1) confirm the occurrence of simple tandem duplications of *PLP1* and (2) propose a replication-based template switching mechanism for human genomic rearrangements (Figures 6B–6E) to elucidate the etiology of some of the more complex, and potentially simple, *PLP1* rearrangements that have been reported for a subset of PMD patients.

Conclusions

Our proposed FoSTeS DNA replication model may offer insight into the mechanism by which the DNA rearrangements causing PMD, and potentially many other genomic disorders due to nonrecurrent rearrangements, may occur. These include: *MECP2* duplications (del Gaudio et al., 2006; Meins et al., 2005; Van Esch et al., 2005) and triplications (del Gaudio et al., 2006) (Figure S4) causing mental retardation and seizures in males, *APP* duplications associated with Alzheimer's disease (Cabrejo et al., 2006; Rovelet-Lecrux et al., 2006), *SNCA* duplications (Chartier-Harlin et al., 2004; Ibáñez et al., 2004; Nishioka et al., 2006) and triplications (Farrer et al., 2004; Polymeropoulos et al., 1997; Singleton et al., 2003) for Parkinson's disease, and a selected small fraction of complex rearrangements of chromosome 17 causing congenital anomalies and Potocki-Lupski syndrome (Potocki et al., 2007; Vissers et al., 2007) (Figure S4).

Further, we suggest that the higher-order genomic architecture, which may contain unusual symmetry, present particularly in regions of nonrecurrent rearrangement, may facilitate such faulty replication by confusing the replication machinery (causing replication fork stalling or pausing). Likewise, the switching of templates, for which only a microhomology is required, may be facilitated by the same unusual complex genomic architecture due to the resulting close proximity of highly similar DNA that would normally be situated much further apart. Template switching may be exacerbated by the presence of cruciform or other non-B DNA structures, resulting in both the complex and noncomplex nonrecurrent rearrangements that we and others (Kosmider and Wells, 2007) observe.

The concept of a replication-based as opposed to a recombination-based mechanism revolutionizes our thinking about nonrecurrent DNA rearrangements and how they may physically occur. The FoSTeS replication-based mechanism may be responsible for other disease-causing genomic rearrangements, gene duplications affecting immediate and long-term genome evolutionary changes (Dumas et al., 2007; Ohno, 1970) including the emergence of novel gene function, the generation of advantageous and/or benign CNVs (Redon et al., 2006) in response to environmental factors, as well as the generation of

complex segmental duplications or regional LCRs. Thus, such “errors of replication” may provide a mechanism for the maintenance of genome plasticity and, conceivably over longer periods of time, genome evolution.

EXPERIMENTAL PROCEDURES

Patients

We analyzed 17 unrelated individuals (Table 1) from our cohort of male patients with PMD, excluding any family members who might give redundant junction sequence data, to perform extensive duplication junction mapping analyses using several independent experimental approaches; eight of these males showed no evidence of a complex *PLP1* duplication rearrangement by various methods (Lee et al., 2006a). The majority of patients has been described previously (Inoue et al., 1999, 2002; Raskind et al., 1991); newly acquired patient samples were obtained with informed consent approved by the Institutional Review Board for Human Subject Research at Baylor College of Medicine. Patient DNA was isolated either directly from peripheral blood using a Puregene kit (Gentra) or from transformed lymphoblastoid cell lines by standard phenol/chloroform extraction.

Semiquantitative PCR Analysis

To map the extent of genomic duplication finely in eight selected patients (Table 1), we performed semiquantitative PCR analysis using a series of primer pairs multiplexed with a control primer set (Table S3B) amplifying the dystrophin gene using a Multiplex PCR Kit (QIAGEN) and following the manufacturer's instructions, except that the PCR reactions were set to amplify for only 25 cycles with an annealing temperature of 60°C. To determine dosage of *PLP1* and also of various other interrogated flanking regions, amplicon intensities for each test primer pair were first normalized to the intensity of the dystrophin product and then to the average amplicon intensity for three male controls. PCR reactions were electrophoresed on 4% NuSieve 3:1 agarose gels (Cambrex Bioscience), and amplicon intensities were measured using an Alpha Innotech Chemilmager 5500 imaging system with accompanying AlphaEase FC molecular imaging software (Alpha Innotech Corporation).

Long-Range PCR Amplification using Primers Facing outward

Junction regions were delineated to within a genomic interval considered feasible for long-range PCR (i.e., ~2–16 kb). Assuming tandem duplications (Inoue et al., 1999), and using outward-facing primers (with respect to reference sequence), long-range PCR was performed using Phusion high-fidelity polymerase with GC buffer (Finnzymes Oy), with and without DMSO (see Results and Table S1). The maximum distance between outward-facing primers was calculated as delineated by semiquantitative multiplex PCR and was used to determine primer extension time. The reaction conditions were as follows: 98°C for 30 s, followed by 35 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s per kb, followed by 72°C for 10 min. Amplification products were electrophoresed on 0.8%–1% agarose gels and DNA from any unique bands were analyzed by DNA sequencing.

Probe Design and Preparation for Southern Blot Analysis

For Southern blot analysis for each patient, unique primers were designed to construct DNA probes that would hybridize to the most distal and/or most proximal ends of the duplication, when possible given the availability of nonrepetitive and nonrepeat sequence. Primer names and sequences, and amplicon sizes and positions, are listed in Table S2. PCR products were purified by gel extraction using a QIAEX II Gel Extraction Kit (QIAGEN).

Southern Blot Analysis

Overnight single digests of 5 µg of patient genomic DNA with several restriction enzymes (Invitrogen; New England Biolabs; Table S2)

were electrophoresed on 0.8% agarose gels at 100 V for 5 hr. Size-fractionated DNA was transferred to a positively charged Sure Blot Nylon Membrane (Intergen Company) using alkaline transfer buffer (0.4 N NaOH/1 M NaCl) overnight. Agarose gels were dephosphorylated briefly (0.2 N HCl) if target bands were >20 kb. Membranes were neutralized (0.5 M Tris-Cl [pH 7.2]/1 M NaCl) for 15 min, and the transferred DNA was fixed to wet membranes using a UV Stratilinker 2400 (Stratagene).

Approximately 25 ng of probe DNA were labeled with ³²P-dCTP (MP Biomedicals) for 1 hr at 37°C using a Rediprime II Random Prime Labeling System (Amersham Biosciences) according to the manufacturer's instructions. Labeled probes were purified using MicroSpin G-50 columns (Amersham Biosciences). Membranes were prehybridized at 65°C in a prewarmed solution consisting of 19 ml dH₂O, 2 ml 10% milk, 4 ml 10% SDS, 3 ml 20× SSPE, and 0.5 ml salmon sperm DNA (denatured by boiling for 10 min) for at least 1 hr and hybridized overnight at 65°C in a pre-warmed solution consisting of 11.4 ml dH₂O, 4 ml 50% dextran sulfate, 1 ml 10% milk, 2 ml 10% SDS, and 1.5 ml 20× SSPE. Membranes were washed in 2× SSC/0.1% SDS at room temperature, 45°C, and 65°C. Hybridized blots were analyzed by autoradiography for the presence of bands of expected size and also for bands of varying size, potentially representing junction fragments. Blots were stripped and rehybridized as necessary.

Long-Range PCR Amplification using Primers Facing Either One Direction or inward

To include the possibility of small inversions at breakpoint junctions, PCR amplification was attempted using gene-specific primers (GSPs) or the furthest primers that detected a DNA duplication result for each patient as defined by semiquantitative multiplex PCR, both (1) singly or paired facing only one direction and (2) paired facing inward. Phusion high-fidelity polymerase with HF buffer (Finnzymes Oy) was used to amplify template DNA enriched for junction sequence under the following reaction conditions: 98°C for 30 s, followed by 10 cycles of 98°C for 10 s, 65°C for 30 s (minus 1°C per cycle), 72°C for 30 s per kb, followed by 30 cycles of 98°C for 10 s, 55°C for 30 s, and 72°C for 30 s per kb, followed by 72°C for 10 min. Reactions were electrophoresed on 0.8%–1% agarose gels and DNA from any unique bands were analyzed by DNA sequencing.

High-Resolution Genome Analysis by Oligonucleotide Array Comparative Genomic Hybridization

To obtain higher resolution of duplication breakpoint junctions than afforded by semiquantitative multiplex PCR and BAC array CGH, we applied oligo array CGH using an Agilent Technologies custom microarray in 4×44K format (#G4426A) to our larger cohort of patients (Table 1). Using the Agilent eArray tool, we selected 24,643 60-mer probes spanning ~10 Mb including *PLP1* (chrX:97,835,000–107,855,000, NCBI build 35). Probe labeling and hybridization were performed following the manufacturer's protocol (Agilent Oligonucleotide Array-based CGH for Genomic DNA analysis, version 4.0 plus the 4×44K complementary protocol with modifications unique to the four-pack format).

Briefly, 1.5 µg of genomic male/female reference and patient DNA were digested with AluI (5 U) and RsaI (5 U) (Promega) for 2 hr at 37°C. Digestions were verified by agarose gel electrophoresis. Labeling reactions with Cy5-dUTP for patient DNA and Cy3-dUTP for male/female reference DNA were performed according to the manufacturer's instructions (Agilent Genomic DNA Labeling Kit Plus, #5188-5309). Individual dye-labeled reference and patient samples were purified using Microcon Ym-30 filters (Millipore Corporation). DNA yield was determined using a NanoDrop ND-1000 UV-VIS spectrophotometer. Each dye-labeled patient and gender-matched reference DNA was combined with 5 µg human Cot-1 DNA (Invitrogen Corporation), 1× Agilent Blocking Agent, and 1× Agilent hybridization buffer (#5188-5220). These mixtures were denatured at 95°C for 3 min,

preincubated at 37°C for 30 min, and hybridized to the array in a hybridization chamber (Agilent Technologies) for 40 hr at 65°C in a rotating oven (Agilent Technologies). Array slides were washed using Agilent Wash solutions 1 and 2 (#5188-5226), Acetonitrile (Sigma-Aldrich), and Stabilization and Drying Solution (#5185-5979), according to the manufacturer's instructions. Slides were scanned on a GenePix 4000B Microarray Scanner (Axon Instruments). Images were analyzed and data were extracted, background subtracted, and normalized using Agilent Feature Extraction Software A.7.5.1. These data were subsequently imported into array CGH analytics software v3.1.28 (Agilent Technologies).

Using these high-resolution data, a new set of primers was designed at the apparent boundaries of each duplicated segment(s) (as determined by a transition from normal copy number to gain in copy number) for all patients except those for whom breakpoint junctions had already been sequenced (i.e., BAB1258 and 1264) and used for long-range PCR amplification as described above (primers facing outward, inward, and in the same direction). When more than one duplicated segment was identified for a given patient, primers facing outward, inward, and the same direction at the boundaries of each segment were paired in all various predicted combinations.

DNA Sequencing Analysis

PCR amplicons that potentially contained junction sequences were purified either with a QIAquick PCR Purification Kit (QIAGEN) or a QIAEX II Gel Extraction Kit (QIAGEN), following the manufacturer's instructions. Purified amplicons and PCR amplification primers were submitted to SeqWright DNA Technology Services or Lone Star Labs (Houston, TX), where DNA sequencing was performed by standard Sanger dideoxy and automated fluorescence methods. Patient DNA sequences were analyzed by comparing to reference sequences, with the use of the UCSC Genome Browser (<http://genome.ucsc.edu/>; Kent et al., 2002) and BLAST (nucleotide-nucleotide BLAST and BLAST for short, nearly exact matches).

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, three tables, and three figures and can be found with this article online at <http://www.cell.com/cgi/content/full/131/7/1235/DC1/>.

ACKNOWLEDGMENTS

We thank the patients and families for their participation; Dr. Chad Shaw for assistance with computational analysis; Drs. Albino Bacolla, Phil Hastings, Matt Hurler, Ken Inoue, Pawel Stankiewicz, and Bob Wells for their critical reviews; and Dr. Phil Hastings and an anonymous reviewer for suggesting the break-induced replication mechanism. This work was supported in part by the Charcot Marie Tooth Association, the National Institute of Child Health and Development (P01 HD39422), and The Baylor College of Medicine Mental Retardation and Developmental Disabilities Research Center (MRDDRC HD024064).

Received: August 17, 2007

Revised: October 10, 2007

Accepted: November 14, 2007

Published: December 27, 2007

REFERENCES

- Bacolla, A., and Wells, R.D. (2004). Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **279**, 47411–47414.
- Blanco-Rodríguez, J. (2002a). Deoxyribonucleic acid replication and germ cell apoptosis during spermatogenesis in the rabbit. *J. Androl.* **23**, 182–187.
- Blanco-Rodríguez, J. (2002b). DNA replication and germ cell apoptosis during spermatogenesis in the cat. *J. Androl.* **23**, 484–490.
- Cabrejo, L., Guyant-Maréchal, L., Laquerrière, A., Vercelletto, M., De La Fourmière, F., Thomas-Antérion, C., Verny, C., Letournel, F., Pasquier, F., Vital, A., et al. (2006). Phenotype associated with *APP* duplication in five families. *Brain* **129**, 2966–2976.
- Chartier-Harlin, M.-C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M., et al. (2004). α -Synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364**, 1167–1169.
- Chen, J.-M., Chuzhanova, N., Stenson, P.D., Ferec, C., and Cooper, D.N. (2005). Intrachromosomal serial replication slippage in *trans* gives rise to diverse genomic rearrangements involving inversions. *Hum. Mutat.* **26**, 362–373.
- del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M., et al. (2006). Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8**, 784–792.
- Dumas, L., Kim, Y.H., Karimpour-Fard, A., Cox, M., Hopkins, J., Pollack, J.R., and Sikela, J.M. (2007). Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277.
- Farrer, M., Kachergus, J., Forno, L., Lincoln, S., Wang, D.S., Hulihan, M., Maraganore, D., Gwinn-Hardy, K., Wszolek, Z., Dickson, D., and Langston, J.W. (2004). Comparison of kindreds with parkinsonism and α -synuclein genomic multiplications. *Ann. Neurol.* **55**, 174–179.
- Goldfless, S.J., Morag, A.S., Belisle, K.A., Sutura, V.A., Jr., and Lovett, S.T. (2006). DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Mol. Cell* **21**, 595–604.
- Gordon, A.J.E., and Halliday, J.A. (1995). Inversions with deletions and duplications. *Genetics* **140**, 411–414.
- Ibáñez, P., Bonnet, A.M., Débarges, B., Lohmann, E., Tison, F., Pollak, P., Agid, Y., Dürr, A., and Brice, A. (2004). Causal relation between α -synuclein gene duplication and familial Parkinson's disease. French Parkinson's Disease Genetics Study Group. *Lancet* **364**, 1169–1171.
- Inoue, K., Osaka, H., Imaizumi, K., Nezu, A., Takahashi, J.-i., Arai, J., Murayama, K., Ono, J., Kikawa, Y., Mito, T., et al. (1999). Proteolipid protein gene duplications causing Pelizaeus-Merzbacher disease: molecular mechanism and phenotypic manifestations. *Ann. Neurol.* **45**, 624–632.
- Inoue, K., Osaka, H., Thurston, V.C., Clarke, J.T.R., Yoneyama, A., Rosenbarker, L., Bird, T.D., Hodes, M.E., Shaffer, L.G., and Lupski, J.R. (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**, 838–853.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
- Kosmider, B., and Wells, R.D. (2007). Fragile X repeats are potent inducers of complex, multiple site rearrangements in flanking sequences in *Escherichia coli*. *DNA Repair (Amst.)* **6**, 1850–1863.
- Kozul, R., Caburet, S., Dujon, B., and Fischer, G. (2004). Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* **23**, 234–243.
- Lee, J.A., and Lupski, J.R. (2006). Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52**, 103–121.
- Lee, J.A., Inoue, K., Cheung, S.W., Shaw, C.A., Stankiewicz, P., and Lupski, J.R. (2006a). Role of genomic architecture in *PLP1* duplication causing Pelizaeus-Merzbacher disease. *Hum. Mol. Genet.* **15**, 2250–2265.
- Lee, J.A., Madrid, R.E., Sperle, K., Ritterson, C.M., Hobson, G.M., Garbern, J., Lupski, J.R., and Inoue, K. (2006b). Spastic paraplegia

- type 2 associated with axonal neuropathy and apparent *PLP1* position effect. *Ann. Neurol.* 59, 398–403.
- Lupski, J.R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* 14, 417–422.
- Lupski, J.R. (2007). Structural variation in the human genome. *N. Engl. J. Med.* 356, 1169–1171.
- Lupski, J.R., and Stankiewicz, P. (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* 1, 627–633.
- Lupski, J.R. and Stankiewicz, P., eds. (2006). *Genomic Disorders: The Genomic Basis of Disease* (Totowa, New Jersey: Humana Press Inc.).
- Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K., and Epplen, J.T. (2005). Submicroscopic duplication in Xq28 causes increased expression of the *MECP2* gene in a boy with severe mental retardation and features of Rett syndrome. *J. Med. Genet.* 42, e12.
- Mimault, C., Giraud, G., Courtois, V., Cailloux, F., Boire, J.Y., Dastugue, B., and Boespflug-Tanguy, O. (1999). Proteolipoprotein gene analysis in 82 patients with sporadic Pelizaeus-Merzbacher disease: duplications, the major cause of the disease, originate more frequently in male germ cells, but point mutations do not. The Clinical European Network on Brain Demyelinating Disease. *Am. J. Hum. Genet.* 65, 360–369.
- Narayanan, V., Mieczkowski, P.A., Kim, H.-M., Petes, T.D., and Lobachev, K.S. (2006). The pattern of gene amplification is determined by the chromosomal location of hairpin-capped breaks. *Cell* 125, 1283–1296.
- Nishioka, K., Hayashi, S., Farrer, M.J., Singleton, A.B., Yoshino, H., Imai, H., Kitami, T., Sato, K., Kuroda, R., Tomiyama, H., et al. (2006). Clinical heterogeneity of α -synuclein gene duplication in Parkinson's disease. *Ann. Neurol.* 59, 298–309.
- Ohno, S. (1970). *Evolution by gene duplication* (New York, NY: Springer Verlag).
- Ohshima, A., Inouye, S., and Inouye, M. (1992). In vivo duplication of genetic elements by the formation of stem-loop DNA without an RNA intermediate. *Proc. Natl. Acad. Sci. USA* 89, 1016–1020.
- Padiath, Q.S., Saigoh, K., Schiffman, R., Asahara, H., Yamada, T., Koepfen, A., Hogan, K., Ptáček, L.J., and Fu, Y.H. (2006). Lamin B1 duplications cause autosomal dominant leukodystrophy. *Nat. Genet.* 38, 1114–1123.
- Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., et al. (1997). Mutation in the α -synuclein gene identified in families with Parkinson's disease. *Science* 276, 2045–2047.
- Potocki, L., Bi, W., Treadwell-Deering, D., Carvalho, C.M.B., Eifert, A., Friedman, E.M., Glaze, D., Krull, K., Lee, J.A., Lewis, R.A., et al. (2007). Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am. J. Hum. Genet.* 80, 633–649.
- Prigent, C., Satoh, M.S., Daly, G., Barnes, D.E., and Lindahl, T. (1994). Aberrant DNA repair and DNA replication due to an inherited enzymatic defect in human DNA ligase I. *Mol. Cell. Biol.* 14, 310–317.
- Raskind, W.H., Williams, C.A., Hudson, L.D., and Bird, T.D. (1991). Complete deletion of the proteolipoprotein gene (PLP) in a family with X-linked Pelizaeus-Merzbacher disease. *Am. J. Hum. Genet.* 49, 1355–1360.
- Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454.
- Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., et al. (2006). *APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* 38, 24–26.
- Shaw, C.J., and Lupski, J.R. (2004). Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* 13, R57–R64.
- Shaw, C.J., and Lupski, J.R. (2005). Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum. Genet.* 116, 1–7.
- Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). α -Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841.
- Sistermans, E.A., de Coo, R.F.M., De Wijs, I.J., and Van Oost, B.A. (1998). Duplication of the proteolipoprotein gene is the major cause of Pelizaeus-Merzbacher disease. *Neurology* 50, 1749–1754.
- Slack, A., Thornton, P.C., Magner, D.B., Rosenberg, S.M., and Hastings, P.J. (2006). On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.* 2, e48.
- Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. *Nature* 447, 102–105.
- Toffolatti, L., Cardazzo, B., Nobile, C., Danielli, G.A., Gualandi, F., Muntoni, F., Abbs, S., Zanetti, P., Angelini, C., Ferlini, A., et al. (2002). Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics* 80, 523–530.
- Trinh, T.Q., and Sinden, R.R. (1991). Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* 352, 544–547.
- Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Géczy, J., et al. (2005). Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* 77, 442–453.
- Vissers, L.E.L.M., Stankiewicz, P., Yatsenko, S.A., Crawford, E., Creswick, H., Proud, V.K., de Vries, B.B.A., Pfundt, R., Marcelis, C.L.M., Zackowski, J., et al. (2007). Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies. *Hum. Genet.* 121, 697–709.
- Wojciechowska, M., Bacolla, A., Larson, J.E., and Wells, R.D. (2005). The myotonic dystrophy type 1 triplet repeat sequence induces gross deletions and inversions. *J. Biol. Chem.* 280, 941–952.
- Wolf, N.I., Sistermans, E.A., Cundall, M., Hobson, G.M., Davis-Williams, A.P., Palmer, R., Stubbs, P., Davies, S., Endziniene, M., Wu, Y., et al. (2005). Three or more copies of the proteolipoprotein gene *PLP1* cause severe Pelizaeus-Merzbacher disease. *Brain* 128, 743–751.
- Woodward, K.J., Cundall, M., Sperle, K., Sistermans, E.A., Ross, M., Howell, G., Gribble, S.M., Burford, D.C., Carter, N.P., Hobson, D.L., et al. (2005). Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination. *Am. J. Hum. Genet.* 77, 966–987.