

# *Alu*-specific microhomology-mediated deletion of the final exon of *SPAST* in three unrelated subjects with hereditary spastic paraplegia

Philip M. Boone, BS<sup>1</sup>, Pengfei Liu, BS<sup>1</sup>, Feng Zhang, PhD<sup>2,3</sup>, Claudia M. B. Carvalho, PhD<sup>1</sup>, Charles F. Towne, PhD<sup>4</sup>, Sat Dev Batish, PhD<sup>4</sup>, and James R. Lupski, MD, PhD<sup>1,5,6</sup>

**Purpose:** Autosomal dominant spastic paraplegia, type 4 (SPG4), a debilitating disorder of progressive spasticity and weakness of the lower limbs, results from heterozygous mutations in the *SPAST* gene. The full spectrum of *SPAST* mutations causing SPG4 and their mechanisms of formation remain to be determined. **Methods:** We used multiplex ligation-dependent probe amplification, locus-specific array comparative genomic hybridization, and breakpoint DNA sequencing to identify and describe genomic rearrangements in three patients with a clinical presentation of hereditary spastic paraplegia. **Results:** We describe three SPG4 patients with intragenic rearrangements in *SPAST*; all specifically delete the final exon, exon 17. Breakpoint sequence analyses provide evidence for *Alu*-specific microhomology-mediated deletion as the mechanism of exon loss; one complex rearrangement apparently occurred by multiple *Alu*-facilitated template switches. **Conclusion:** We hypothesize that the high concentration of *Alu* family members in the introns and flanking sequence of *SPAST* may predispose to intragenic rearrangements. Thus, *Alu*-specific microhomology-mediated intragenic rearrangements in *SPAST* may be a common cause of SPG4. Furthermore, we propose that genomic deletions encompassing the final exon of *SPAST* may affect expression of *SLC30A6*, the most proximal downstream locus and a gene that has been implicated in the pathogenesis of Alzheimer disease, potentially explaining recent reports of dementia in selected SPG4 patients. *Genet Med* 2011;13(6):582–592.

**Key Words:** *SPG4*, *spastin*, intragenic rearrangements, exon mutations, *FoSteS*

Hereditary spastic paraplegias (HSPs) are a group of nervous system disorders characterized by progressive spasticity and weakness of the lower extremities.<sup>1</sup> HSP subtypes are described either as pure or uncomplicated, with spasticity as the sole feature, or as complicated, in which other neurologic find-

ings, including dementia, deafness, ataxia, extrapyramidal disturbances, retinopathy, optic neuropathy, and ichthyosis, can occur either individually or in selective combinations. The HSPs vary not only in clinical manifestations but also in modes of inheritance (autosomal dominant [AD], autosomal recessive, and X linked) and genetic causes; to date, more than 40 spastic paraplegia gene loci (*SPG*) have been identified, defining an equal number of similarly named HSP subtypes.<sup>1,2</sup>

AD spastic paraplegia-4 (SPG4; OMIM# 182601), characterized by relatively pure spastic paraplegia, is the most common type of AD-HSP and the most common form of HSP overall.<sup>3</sup> SPG4 may present as either an inherited (AD) or sporadic condition that is caused by heterozygous mutations in the *SPAST* gene (OMIM# 604277; 2p22.3). *SPAST* point mutations, including missense, nonsense, and splice site mutations, were first identified in patients with AD-HSP by Hazan et al.<sup>4</sup> Since then, small insertions/deletions and gross deletions,<sup>5</sup> a gross duplication,<sup>6</sup> and a small complex rearrangement<sup>5</sup> in *SPAST* have all been identified in patients with AD-HSP, consistent with haploinsufficiency at this locus as the cause of SPG4.

*SPAST* codes for spastin, a member of the AAA (ATPases associated with diverse cellular activities) family of ATPases<sup>4</sup> with roles in microtubule dynamics<sup>7,8</sup> and membrane trafficking.<sup>9</sup> Spastin appears to interact with atlastin-1 (encoded by *SPG3A*) and receptor expression enhancing protein 1 (encoded by *REEP1*); haploinsufficiency of these proteins leads to SPG3A and SPG31, respectively, two other types of AD-HSP.<sup>10</sup> Although convincing genotype-phenotype correlations for *SPAST* mutations in SPG4 have been alternately reported<sup>11–13</sup> and discounted,<sup>14,15</sup> point mutations and insertions/deletions tend to cluster in the sequences coding for the AAA domain, microtubule interacting and trafficking domain, microtubule-binding domain, and a functionally undefined region (residues 228–269) of spastin.<sup>16</sup> Several sequence variants within *SPAST* (reviewed in Ref. 17) and one in heat shock 60 kDa protein 1 (chaperonin) (*HSPD1*; the gene associated with SPG13)<sup>18</sup> have been proposed to be modifiers of the SPG4 phenotype.

Intragenic copy-number variants (CNVs) in *SPAST*, most commonly identified by multiplex ligation-dependent probe amplification (MLPA), have been found to cause disease in a substantial proportion of mutation-negative (i.e., those lacking single-nucleotide variants identifiable by gene sequencing) patients with HSP.<sup>12</sup> Intragenic copy-number changes have been implicated in the pathogenesis of numerous disorders,<sup>19</sup> most often reasoned to result from gene disruption. Mechanisms of genomic copy-number change (reviewed by Zhang et al.<sup>20</sup>) include (1) nonallelic homologous recombination (NAHR), characterized by recurrent rearrangements mediated by flanking low-copy repeats. NAHR may also occur at highly homologous non low-copy repeats interspersed repetitive sequences, for example, LINES or *Alu* elements (i.e., *Alu-Alu* recombination<sup>21,22</sup>); (2) nonhomologous end joining, which generates non-

From the <sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; <sup>2</sup>State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University; <sup>3</sup>MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; <sup>4</sup>Athena Diagnostics, Inc., Worcester, Massachusetts; <sup>5</sup>Department of Pediatrics, Baylor College of Medicine; and <sup>6</sup>Texas Children's Hospital, Houston, Texas.

James R. Lupski, MD, PhD, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604B, Houston, TX 77030. E-mail: jlupski@bcm.edu.

Disclosure: See author disclosures in Acknowledgments.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.geneticsinmedicine.org).

Submitted for publication November 29, 2010.

Accepted for publication January 14, 2011.

Published online ahead of print March 30, 2011.

DOI: 10.1097/GIM.0b013e3182106775

recurrent rearrangements with variable breakpoint architecture; (3) fork stalling and template switching (FoSTeS) or microhomology-mediated break-induced replication (MMBIR), characterized by breakpoint microhomology and the possibility of complex rearrangements,<sup>23,24</sup> and (4) transposition of mobile elements throughout the genome.<sup>20</sup> FoSTeS and MMBIR, proposed, replication-based processes, have been implicated in the generation of CNV of a wide size range spanning from the genomic scale (megabases) to the exonic scale (~100 bp to kb in size).<sup>25</sup> Deletion CNVs have been recently described with breakpoints, the architecture of which suggests FoSTeS or MMBIR, that localize to *Alu* elements.<sup>26–30</sup> *Alu*-specific microhomology-mediated deletion is the mechanism proposed to have generated these CNVs, wherein moderate sequence similarity between breakpoint *Alus* (in contrast to the nearly complete sequence identity of breakpoint *Alus* required for minimal efficient processing segments in homologous *Alu-Alu* recombination) is sufficient to provide a substrate for the microhomology-mediated processes of FoSTeS/MMBIR.

Which of the above mechanisms play(s) a role in generating intragenic CNV in *SPAST* is unknown, because deletion or duplication breakpoints have only been sequenced for two patients.<sup>6,31</sup> We report three deletions of the final exon of *SPAST*, one of which is complex, all with breakpoints displaying microhomology and localized to *Alu* elements with moderate sequence similarity and in direct orientation but displaying  $\leq 89\%$  sequence identity. These findings suggest *Alu*-specific microhomology-mediated generation of intragenic CNV in *SPAST* as a recurrent, and potentially common, cause of HSP.

## MATERIALS AND METHODS

Anonymized genomic DNA samples from three patients (A37, A38, and A39), each with a suspected clinical diagnosis of HSP, were obtained from Athena Diagnostics (Worcester, MA). Each sample was analyzed by MLPA using the MRC-Holland (Amsterdam, the Netherlands) SALSA MLPA kit P165. This kit contains one probe pair complementary to each of exons 1–16 of *SPAST* and two probe pairs complementary to exon 17.

To determine the size, genomic context, and extent of each deletion, samples were analyzed using an  $8 \times 15$  K Agilent (Santa Clara, CA) oligonucleotide microarray, custom-designed using eArray (<https://earray.chem.agilent.com/earray/>). This array interrogated the entire *SPAST* gene and ~950,000 kb up-

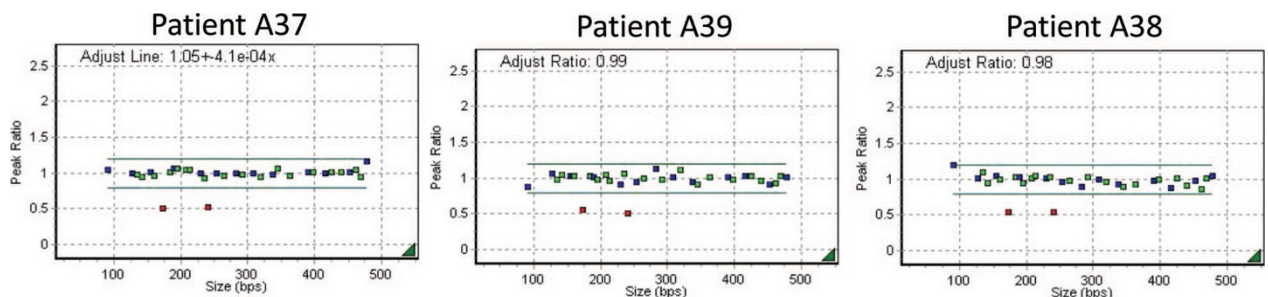
stream and downstream with 11,940 probes, for an average probe spacing of ~200 base pairs (bp). The remaining probes were control probes, spread throughout the genome. Array comparative genomic hybridization (aCGH) and subsequent bioinformatic analyses were performed as in the study by Gonzaga-Jauregui et al.<sup>32</sup>

For each of the three patients, estimated deletion breakpoints derived from aCGH were used to design primers to polymerase chain reaction (PCR) amplify by the breakpoint regions: Primer A37F: 5'-CATAAATACCAACAGAAGAA-AATTACA-3'; A37R: 5'-GGGAAAATTTTCACGCATCAT-3'; A39F: 5'-TCCAAAGGCAATTTAAAAGATCA-3'; A39R: 5'-GGGAAAATTTTCACGCATCAT-3'; A38F: 5'-TGTCTGC-CAGTGAGGTATAGTATTTT-3'; and A38R: 5'-CATCCCAAATGCTTAAGACCA-3'. A 23  $\mu$ L PCR mix consisting of 170  $\mu$ M each deoxynucleotide triphosphate, 0.43  $\mu$ M each primer, 75 ng DNA, 1.5 U HotStarTaq DNA polymerase (Qiagen, Valencia, CA), and 1X PCR buffer was subjected to the following "touchdown" PCR reaction in a thermal cycler: 95°C  $\times$  15 minutes (min); 10 cycles of 94°C  $\times$  30 seconds, 61°C  $\times$  15 seconds (this annealing temperature dropped by 0.5°C in each cycle following the first), and 72°C  $\times$  90 seconds; 32 cycles of 94°C  $\times$  30 seconds, 56°C  $\times$  15 seconds, and 72°C  $\times$  90 seconds; and 72°C  $\times$  10 minutes.

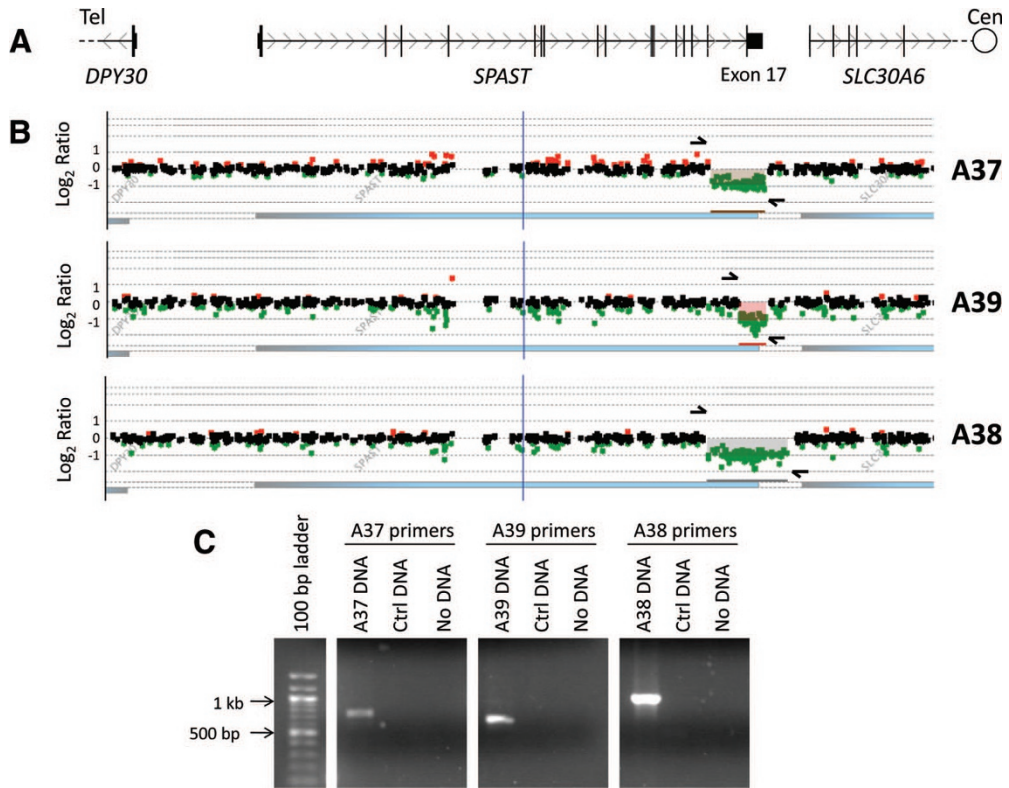
Sex-matched control DNA (NA10851 and NA15510) was derived from cell lines obtained from Coriell Cell Repositories (Camden, NJ). ExoSAP-IT (USB, Cleveland, OH) was used to clean PCR products, according to the manufacturer's instructions. DNA sequencing was performed by the Sanger di-deoxynucleotide method (Child Health Research Center Molecular Core Laboratory, Baylor College of Medicine, Houston, TX; and SeqWright, Houston, TX). The March 2006 assembly of the reference genome (NCBI36/hg18) was used to analyze both sequencing and array data. *SPAST* exon numbering is based on RefSeq transcript NM\_014946.3.

## RESULTS

MLPA revealed a heterozygous genomic deletion including exon 17, the final exon of the *SPAST* gene, in each of the three patients (Fig. 1 and Figs., Supplemental Digital Content 1, <http://links.lww.com/GIM/A161>, which show a prenormalized plots of MLPA data). This allowed confirmation of the clinically suspected spastic paraplegia and enabled a molecular diagnosis of SPG4. To determine the boundaries of each deletion, patient DNA was analyzed on a genomic



**Fig. 1.** MLPA reveals deletions in the final exon of *SPAST*. Dot plots depicting normalized MLPA values demonstrate that the genomic regions corresponding to two probe pairs—both localized to exon 17 of *SPAST*—are present in only a single copy in the three patients. This indicates a heterozygous genomic loss in exon 17 (the final exon) of *SPAST*. Green dots indicate probe pairs directed against exons of *SPAST*, blue dots indicate control probe pairs, and red dots indicate probe pairs in regions of copy-number change.



**Fig. 2.** Targeted aCGH and PCR assays confirm and define the range of deletions of *SPAST* exon 17. **A**, Local genomic map of the *SPAST* locus. **B**, Local view of aCGH results, aligned with the genomic map in (A). Regions of copy-number loss are indicated by shading and underlined in similar color. In each patient, the deletion spans exon 17, the final exon of *SPAST*. Arrows indicate the locations of PCR primers used to further define each deletion. As primer pairs span the predicted breakpoints, a deletion is expected to substantially reduce the amplicon size. This enables a PCR product to be amplified from the deleted allele in the three patients but not from control DNA. **C**, Results of PCR, followed by agarose gel electrophoresis, which for each patient further confirms a deletion. Cen, centromeric; tel, telomeric.

microarray, custom designed to interrogate the *SPAST* locus at high resolution. Array CGH confirmed each deletion and provided an approximation of genomic breakpoints (Fig. 2A, B). In each case, PCR further confirmed the genomic deletion (Fig. 2B, C) and provided substrate for subsequent breakpoint sequencing.

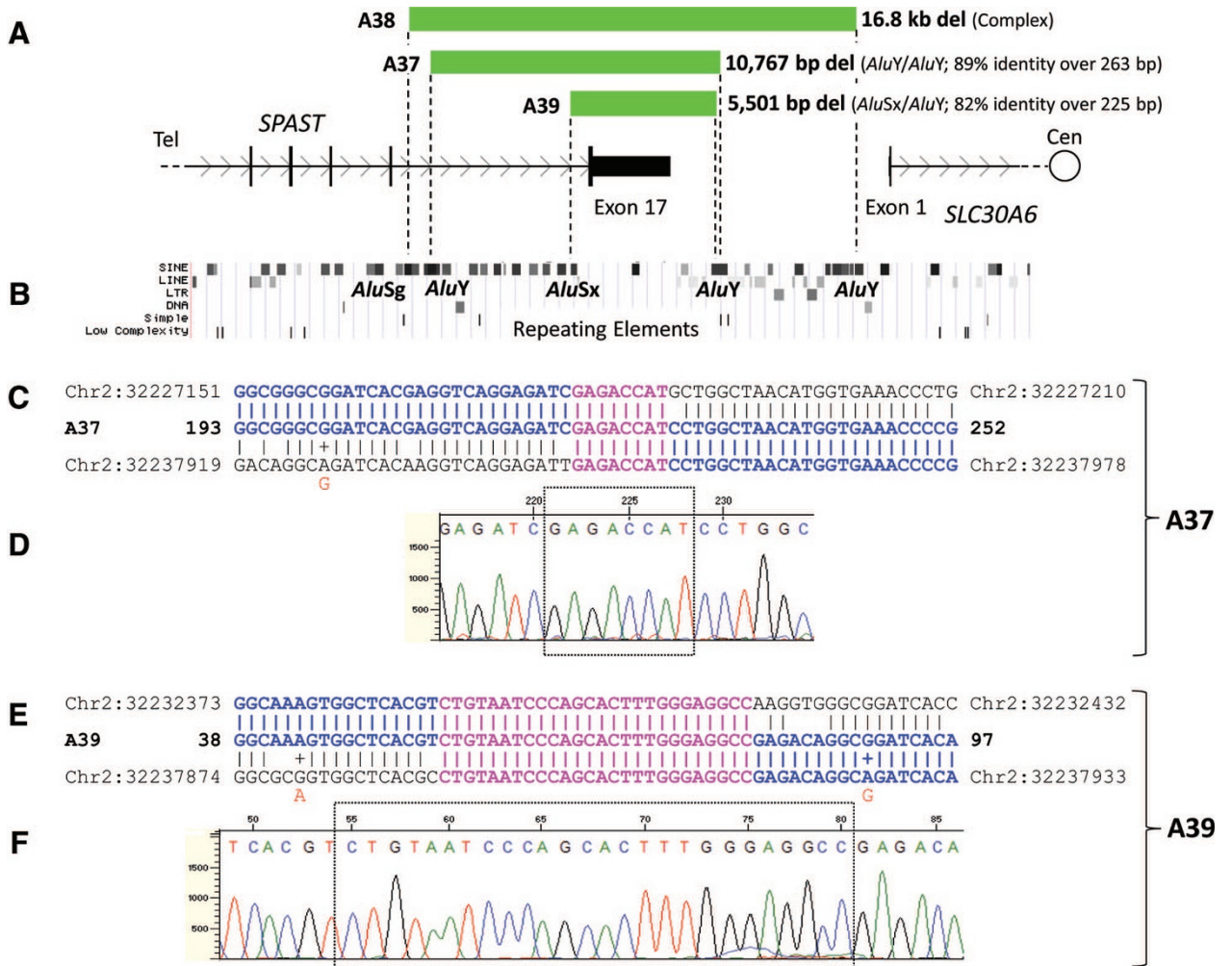
DNA sequencing revealed simple (i.e., noncomplex) deletions of 10,767 bp and 5,501 bp in patients A37 and A39, respectively, which solely encompassed exon 17 of *SPAST* (Fig. 3). Interestingly, both sets of deletion breakpoints exhibited microhomology of 8 and 26 bp, respectively (Fig. 3C, E), and localized to *Alu* elements in direct orientation (Fig. 3A, B). Although the centromeric breakpoints of these two cases localize to the same *AluY*, their exact rearrangement join-point locations differ (Fig. 3C–F).

Patient A38 was found by DNA sequencing to have a complex rearrangement involving a 16.8 kb genomic loss encompassing exon 17 of *SPAST* (Fig. 3A). The boundaries of the deletion map to *Alu* elements of differing families (*AluY* and *AluSg*) in direct orientation. Interestingly, in place of the deleted genomic region, >100 bp of inserted sequence was found (Figs. 4A and 5). This insertion consists of either one or two discrete regions that map to intronic/intergenic *Alu* elements within or near *SPAST*. This uncertainty exists owing to several possible

inferred series of events that could result in the recombinant, mutant sequence product derived from patient DNA (Figs. 4B–F and 5).

For each proposed mechanism, breakpoints are localized to *Alu* elements in direct orientation (Fig. 4C–F) and exhibit microhomology (Fig. 5), and all combinations of potential sequential events in the rearrangement process result in the experimentally identified breakpoint displayed in Figure 5, A and B (also illustrated in Fig. 4C–E). Moreover, all inferred avenues for deriving the recombinant chromosome involve a replication-based mechanism of rearrangement, for example, FoSTeS or MMBIR (see “Discussion”).

The proposed alternative mechanisms are each based on the presumed reference haploid human genome sequence and differ as follows: (1) Figures 4C and 5C illustrate the possibility that a single nontandem duplication of between 113 and 136 bp accounts for the inserted genomic material in the patient (a range of sizes must be proposed, as microhomology limits the resolution with which a breakpoint position may be known). This proposed sequence of events necessitates that a single-nucleotide polymorphism (SNP) exists at Chr2:32,243,095 (C>T). Although this SNP is not present either in dbSNP (build 130) or as one of the additional personal genome variants listed in the University of California, Santa Cruz genome browser (<http://genome.ucsc.edu>), this nucleotide position occurs in

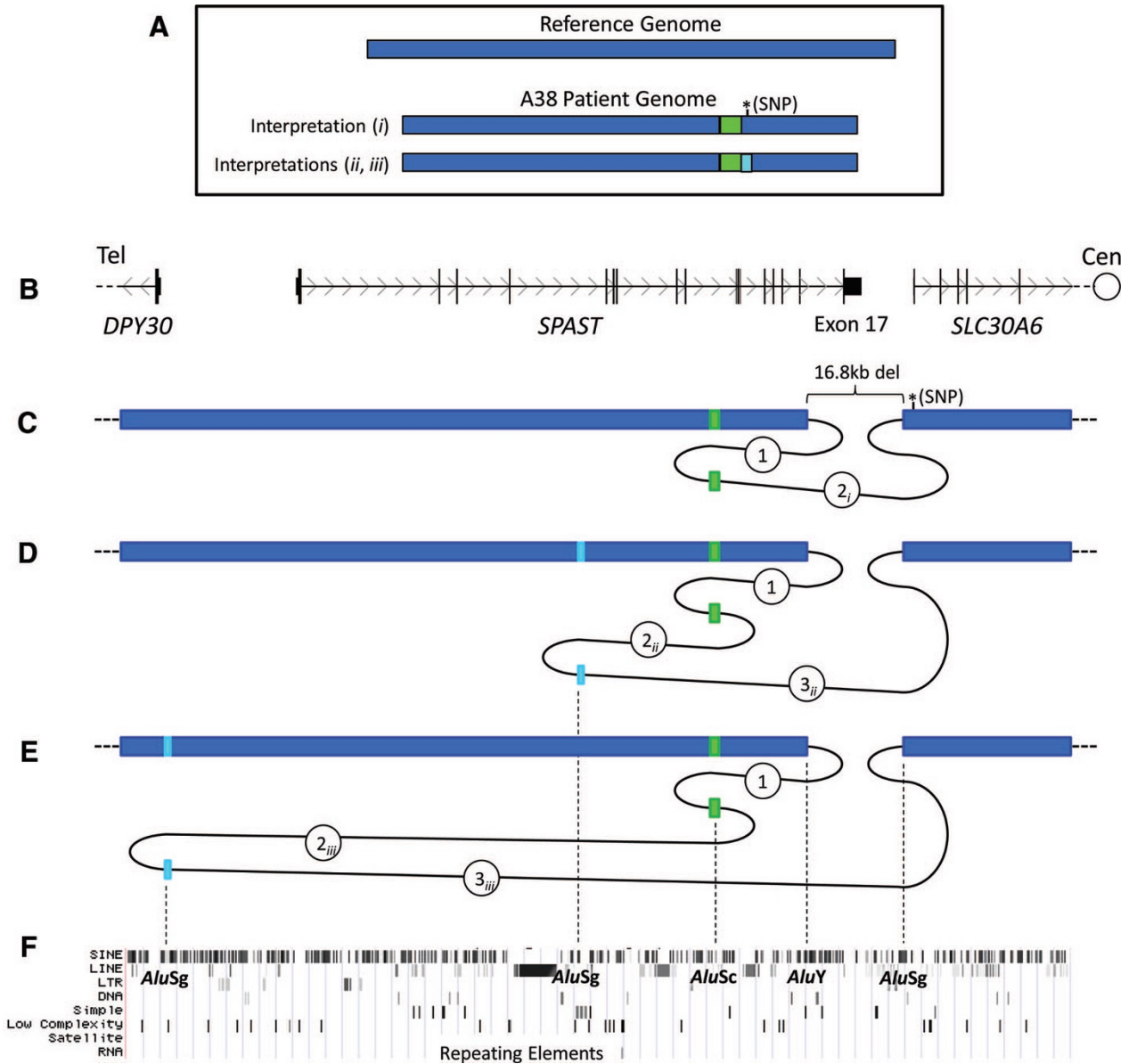


**Fig. 3.** Deletion breakpoints in patients A37 and A39 localize to *Alu* elements and display microhomology. **A**, Green bars represent the regions deleted in patients A37, A38, and A39, superimposed on a genomic map of the 3' end of *SPAST*. In each case, only exon 17 is deleted. **B**, A map of repeating elements, obtained from the University of California, Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/>) is aligned with (**A**). All deletion breakpoints (indicated by dotted lines in [**A**]) fall within *Alu* elements; each pair of *Alus* flanking a deletion are in direct orientation with each other. **C–F**, Sequencing of the breakpoint regions in patients A37 and A39 reveals microhomology. **C** and **E**, The DNA sequences of the deletion breakpoint regions of mutant patient chromosomes (“A37” and “A39”) are displayed between the telomeric (top) and centromeric (bottom) reference sequences. Perfect sequence identity with one of the reference sequences is indicated with bold blue text. Microhomology (sequence identity among all three sequences) is indicated with bold pink text. A “+” and red lettering indicate a known SNP, as listed by the UCSC genome browser, which matches the sequenced patient DNA. **D** and **F**, DNA sequencing traces. The regions of microhomology are boxed. Although the centromeric breakpoints of patients A37 and A39 fall within the same *AluY* element, their exact locations differ. Del, deletion; cen, centromeric; tel, telomeric.

noncoding sequence and lacks conservation even among primates, so a benign SNP at this location is quite plausible in the personal genome of this individual patient with *SPG4*. As this was an anonymized sample, parental DNA is not available for SNP genotyping; and (2) it is also possible that an additional (third) FoSTeS/MMBIR event occurred, producing a second, smaller duplication (Figs. 4D–E and 5D–E). This latter inference may more parsimoniously explain the observed data, as it eliminates the need for the aforementioned novel SNP to exist at Chr2:32,243,095. The locations depicted in Figures 4D–E and 5D–E for this second, smaller duplicated sequence are not

the only possibilities, as a sufficient length of identical sequence exists in multiple other copies throughout the genome; they are pictured on account of being the only such sequences within one megabase of the *SPAST* locus.

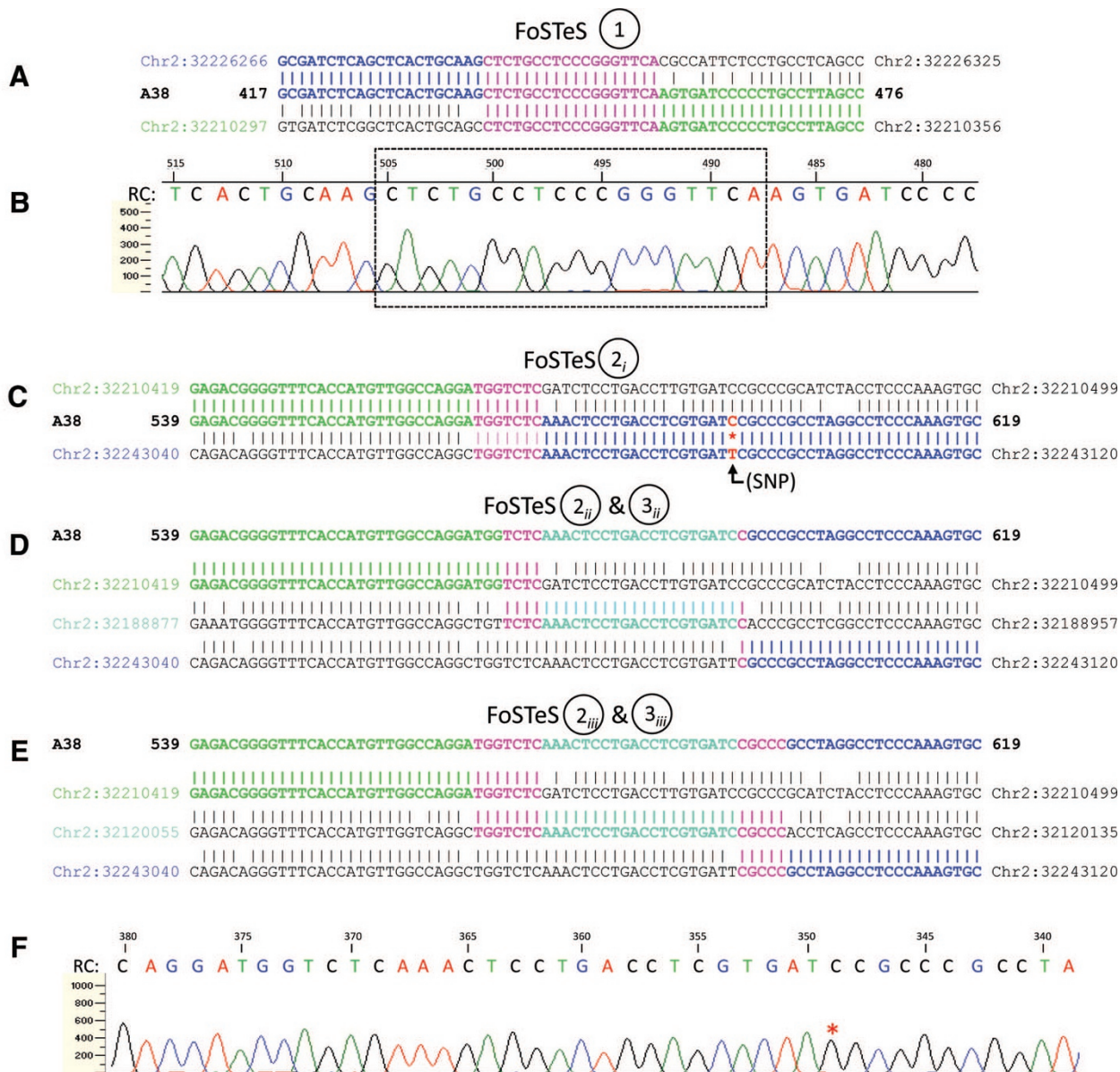
Given the overwhelming proclivity of our patients' deletion breakpoints to exist in *Alu* elements, we computationally examined the architecture of the *SPAST* locus to assess whether a local enhancement of *Alu* concentration may exist. We tabulated each *Alu* family member from 10 kb upstream of *SPAST* to 8,225 bp downstream, where the solute carrier family 30 (zinc transporter), member 6 gene (*SLC30A6*), the



**Fig. 4.** A complex rearrangement in patient A38 deletes the final exon of *SPAST* and probably occurs by multiple *Alu*-specific FoSTeS events. **A**, With respect to the reference genome, patient A38 has a heterozygous loss of 16.8 kb and in its place either one or two shorter, inserted sequences. This uncertainty exists owing to several possible inferred series of events that could result in the recombinant, mutant sequence product derived from patient DNA. **B**, Local genomic map of the *SPAST* locus, aligned with (C–F). **C–E**, Schematic diagrams of three potential series of FoSTeS events: (C) FoSTeS X 2, deleting 16.8 kb (including exon 17), and inserting (creating a nontandem duplication) between 113 and 136 bp. This mechanistic explanation requires a novel SNP to exist centromeric to breakpoint 2 (see Fig. 5C and text); (D) one possible FoSTeS X 3 event, deleting 16.8 kb, inserting/duplicating 113–136 bp, and inserting/duplicating 21–24 bp; and (E) another (of multiple) possible FoSTeS X 3 events, deleting 16.8 kb, inserting/duplicating 113–136 bp, and inserting/duplicating 21–31 bp. **F**, A map of repeating elements, obtained from the UCSC genome browser (<http://genome.ucsc.edu/>). All five potential deletion breakpoints (indicated by dotted lines) fall within *Alu* elements in direct orientation. Del, deletion; cen, centromeric; tel, telomeric.

most proximal downstream gene, begins. One hundred sixty-three *Alu* family members exist in this 112,252 bp region, totaling 40,639 bp, or 36%, of the genomic sequence. This significantly exceeds the ~11% genome-wide average ( $P < 0.0001$ ; Pearson's  $\chi^2$  test; see calculation, Supplemental Dig-

ital Content 1, <http://links.lww.com/GIM/A161>).<sup>33</sup> Furthermore, the sequences flanking exon 17, the exon deleted in our three patients, are particularly enriched for *Alu* family members (intron 16: 4,045/7,115 bp [57%]; 3' flanking intergenic region: 2,951/8,225 bp [36%]).



**Fig. 5.** Breakpoint sequencing in patient A38 reveals a complex rearrangement and suggests multiple *Alu*-specific FoSTeS events. Sequencing of the breakpoint region reveals that this is a complex rearrangement with multiple instances of microhomology. This suggests one of several multiple-FoSTeS schemes as having generated this rearrangement. The DNA sequence of the breakpoint regions of the mutant patient chromosome (“A38”) is displayed either in between or on top of relevant reference sequences. Perfect sequence identity with one of the reference sequences is indicated by bold colored text, corresponding with the coloring in Figure 4C–E. Microhomology (homology among three sequences) is indicated by bold pink text. As sequencing was performed in reverse orientation to the reference sequence, sequence traces have been flipped horizontally and a reverse complement (RC) sequence accompanies them. Circled numbers and subscripts signify a correspondence to similarly labeled FoSTeS events depicted in Figure 4C–E. As displayed in Figure 4, all breakpoints localize to directly oriented *Alu* elements. A, This breakpoint is shared by all possible rearrangement schemes. B, DNA sequencing trace, corresponding to (A), in which the 18 bp region of microhomology is boxed. C–E, Three of several possibilities for the remaining FoSTeS event(s): (C) The second of two breakpoints in this scheme, displaying 7 bp of microhomology and necessitating that the base at Chr2:32,243,095 (in red text with an accompanying asterisk) be a novel SNP on this allele in the patient (see text); (D) the second and third breakpoints of one possible FoSTeS X 3 event, featuring four and one bp of microhomology, respectively; and (E) the second and third breakpoints of another possible FoSTeS X 3 event, featuring 7 and 5 bp of microhomology, respectively. The light blue regions in this figure and in Figure 4 were found by searching for the sequence CAAACTCTGACCTCGTGATCC, the minimum sequence needed to complete this rearrangement with no requirement for a novel SNP. Although multiple copies of this short sequence exist on chromosome 2, those displayed in (D and E) and in Figure 4D and E are the copies located most near to the final exon of *SPAST* (all others are >1 Mb away). F, DNA sequencing trace, corresponding to (C–E) in which the base at Chr2:32,342,095 is marked with a red asterisk.

**Table 1** Published intragenic CNVs in *SPAST* of size >30 bp

No.	Exon(s) involved	Clinical diagnosis (no. different families if >1)	Method of detection	Coordinates <sup>a</sup>	References
<b>Deletions</b>					
1	1	AD-HSP	MLPA	—	15
2	1	AD-HSP <sup>b</sup> (3 families)	MLPA	—	12
3	1	HSP	MLPA	—	14
4	1	AD-HSP (1 or 2 families)	MLPA	—	17
5	1 (partial)	AD-HSP	Southern blot, seq	Chr2:32,142,293–32,144,598	31
6	1–3	AD-HSP	MLPA	—	15
7	1–3	HSP	MLPA	—	16
8	1–4	AD-HSP <sup>b</sup>	MLPA	—	12
9	1–7	AD-HSP	MLPA	—	15
10	1–17	AD-HSP <sup>b</sup> (4 families)	MLPA	—	12
11	2–5	AD-HSP	MLPA	—	15
12	2–9	Familial HSP	MLPA	—	16
13	2–16	AD-HSP	MLPA	—	34
14	2–16	Familial HSP	MLPA	—	16
15	2–17	HSP	MLPA	—	14
16	3–17	Spastic paraplegia	MLPA	—	35
17	4	HSP	MLPA	—	35
18	4–17	AD-HSP	MLPA	—	34
19	4–17	AD-HSP <sup>b</sup>	MLPA	—	12
20	5	HSP	MLPA	—	14
21	5–6	AD-HSP <sup>b</sup>	MLPA	—	12
22	5–7	AD-HSP	MLPA	—	15
23	5–7	AD-HSP <sup>b</sup>	MLPA	—	12
24	5–15	AD-HSP (2 families)	MLPA	—	15
25	6	AD-HSP <sup>b</sup>	MLPA	—	12
26	6–17	AD-HSP	MLPA	—	15
27	8	Familial HSP	MLPA	—	16
28	8–9	AD-HSP	MLPA	—	17
29	8–12	AD-HSP <sup>b</sup>	MLPA	—	12
30	8–17	AD-HSP	MLPA	—	34
31	8–17	AD-HSP <sup>b</sup> (2 families)	MLPA	—	12
32	9	AD-HSP <sup>b</sup>	MLPA	—	12
33	9–12	AD-HSP <sup>b</sup>	MLPA	—	12
34	9–17	Familial HSP	MLPA	—	16
35	10	AD-HSP	MLPA	—	15
36	10–12	AD-HSP	MLPA	—	15
37	10–12	AD-HSP	MLPA	—	34

(Continued)

Table 1 Continued

No.	Exon(s) involved	Clinical diagnosis (no. different families if >1)	Method of detection	Coordinates <sup>a</sup>	References
38	10–12	HSP	MLPA	—	14
39	10–12	AD-HSP	MLPA	—	36
40	10–16	AD-HSP <sup>b</sup>	MLPA	—	12
41	10–16	Familial HSP	MLPA	—	37
42	13	AD-HSP <sup>b</sup> (2 families)	MLPA	—	12
43	13–16	HSP	PCR and electrophoresis of cDNA	—	5
44	14–17	Spastic paraplegia	MLPA	—	35
45	16	AD-HSP	MLPA	—	34
46	16	AD-HSP <sup>b</sup> (2 families)	MLPA	—	12
47	16–17	AD-HSP	MLPA	—	15
48	16–17	AD-HSP <sup>b</sup>	MLPA	—	12
49	17	AD-HSP	MLPA	—	15
50	17	AD-HSP (2 families)	MLPA	—	34
51	17	AD-HSP <sup>b</sup>	MLPA	—	12
52	17	HSP	MLPA	—	16
53	17	Sporadic spastic paraplegia	MLPA	—	36
54	17	HSP <sup>c</sup>	MLPA, LS-aCGH, seq	See text and figures	Present investigation (patient A37)
55	17	HSP <sup>c</sup>	MLPA, LS-aCGH, seq	See text and figures	Present investigation (patient A39)
Duplication					
56	10–12	AD-HSP	MLPA, seq	Chr2:32,212,239–32,216,284	6
Complex					
57	17	HSP <sup>c</sup>	MLPA, LS-aCGH, seq	See text and figures	Present investigation (patient A38)

<sup>a</sup>March 2006 assembly of the reference genome (NCBI36/hg18).

<sup>b</sup>One of 121 families studied by Depienne et al.<sup>12</sup> displayed disease in two siblings, rather than strict AD inheritance. It is unclear whether this family is represented by any exonic CNV reported in the table and, if so, whether the parents were evaluated for mosaicism of the deletion CNV<sup>25</sup> or for subtle clinical features of HSP.

<sup>c</sup>Inheritance is not known for these anonymized samples.

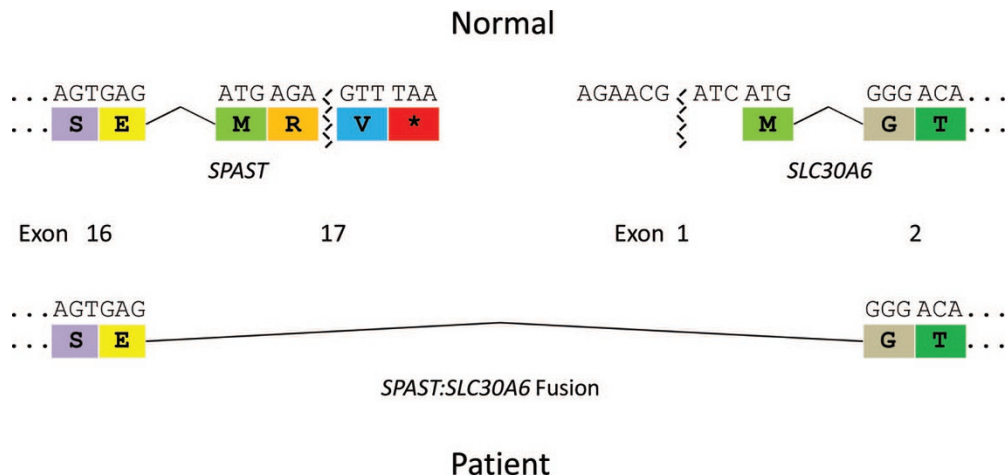
AD, autosomal dominant; HSP, hereditary spastic paraplegia; MLPA, multiplex ligation-dependent probe amplification; LS-aCGH, locus-specific aCGH; seq, DNA sequencing.

## DISCUSSION

We describe three SPG4 patients with intragenic rearrangements in *SPAST*, each deleting its final exon (exon 17). This finding confirms the importance of exonic CNV in *SPAST* as a cause of SPG4. Table 1 lists known intragenic rearrangements of >30 bp in *SPAST* and demonstrates that pathogenic exonic deletions can exist throughout the gene. The use of a custom genomic microarray and DNA sequencing allowed us to determine the exact breakpoint sequences of intragenic CNVs in *SPAST*. These techniques also have enabled us to identify a complex exonic rearrangement in this gene in one patient (A38) and have provided insight into the mechanism of origin of this and the other deletion CNVs we report.

In all three patients, it is the final exon of *SPAST* that is deleted, raising the possibility that expression of the most proximal downstream gene, *SLC30A6*, may be altered. The downstream (centromeric) breakpoint of each patient's CNV lies within 6.5–1.3 kb of the start of *SLC30A6*, potentially a proximity that could either disrupt promoter elements or convey a position effect.<sup>38,39</sup> A final-exon deletion in *SPAST* could also allow a fusion transcript to be made that joins some portions of *SPAST* and the directly oriented *SLC30A6*.<sup>40,41</sup> Conceptual transcription and translation suggest reading frame preservation for a theoretical fusion transcript resulting from splicing of the penultimate exon of *SPAST* (exon 16) to *SLC30A6* exon 2 (Fig. 6). Nervous system mRNA is not available from the patients presented in this study, precluding further experimental investigation of the above possibilities.





**Fig. 6.** DNA and protein sequences demonstrating that a (heterozygous) deletion of the final exon (exon 17) of *SPAST* may result in a fusion protein. Removal of the final exon of *SPAST* may allow in-frame splicing from the penultimate exon (exon 16) of *SPAST* to exon 2 of *SLC30A6* and a fusion protein product to be made.

*SLC30A6* encodes a zinc transport protein<sup>42</sup> that may play a role in the pathogenesis of Alzheimer disease.<sup>43–46</sup> Intriguingly, a subset of patients with SPG4 has been reported to have dementia and/or cognitive impairment (reviewed in Ref. 47). Murphy et al.<sup>47</sup> recently studied a family of patients with SPG4 with deletions of *SPAST* exon 17 and described a comorbid, age-dependent cognitive decline and dementia in patients carrying the deletion, although any potential effect on *SLC30A6* was not assessed. If altered expression of *SLC30A6* (or function of the zinc transporter it encodes, in the case of a fusion protein) is potentially associated with cognitive phenotypes in such individuals, then perhaps either single-nucleotide variants or CNVs in or near this gene that do not affect *SPAST* may either cause or predispose to isolated cognitive decline or dementia.

In all three patients, breakpoints exhibit microhomology and localize to *Alu* elements. This breakpoint architecture provides evidence for *Alu*-specific microhomology-mediated deletion as the mechanism of exon loss. Although *Alu*-mediated NAHR cannot be excluded as a possible CNV-generating mechanism in patients A37 and A39, the observed breakpoint *Alus* are either of different families or, in the case of the *AluYs* at the breakpoint junction of patient A37's deletion, share only 89% identity over 263 bp based on the reference haploid human genome. The experimentally suggested minimum requirement for favorable homologous recombination (i.e., the minimal efficient processing segment) for mammalian genomes is thought to be approximately 200–300 bp.<sup>29,48</sup> Thus, Holliday structures potentially formed by *Alu* members from unrelated families or with <95% identity may be broken down by the mismatch repair pathway. The CNV in patient A38 is complex and more parsimoniously explained by template switching than multiple NAHR events.

The concentration of *Alu* family members in the introns and flanking sequence of *SPAST* is more than three times greater than the genomic average. In intron 16, it is more than five times greater. A similar, though less pronounced, phenomenon is seen in the *CDKL5* gene, in which Erez et al.<sup>26</sup> have reported *Alu*-mediated intragenic deletions with breakpoints localizing to regions of *Alu* concentration greater than that of the gene as a whole—itsself enriched for *Alus* above the baseline genomic level (data not shown). We hypothesize that this genomic architecture may predispose to *Alu*-specific microhomology-

mediated intragenic rearrangements within *SPAST*, in particular involving exons near the highest *Alu* density (e.g., exon 17). Thus, we predict that this mechanism is not only a possible but also frequent cause of pathogenic CNV in this gene, leading to SPG4. Beetz et al.<sup>15</sup> noted that both the *Alu* density and the frequency of intragenic CNV in the *SPG3A* locus (in a small sample of patients with AD-HSP) were lower than those of *SPAST* and hypothesized a causal relationship including potentially *Alu*-mediated NAHR. We suggest that *Alu*-specific microhomology-mediated intragenic rearrangement may explain their observation.

Exonic deletions predominate among known, pathogenic, intragenic CNV in *SPAST* (Table 1); indeed, only a single exonic duplication has been reported in an SPG4 patient.<sup>6</sup> Duplication (or higher order copy-number gain) of one or several exons could be benign, whereas deletion of the same exon(s) would be pathogenic. Nonetheless, the number of reported exonic duplications seems to be lower than expected (see calculation, Supplemental Digital Content 1, <http://links.lww.com/GIM/A161>). This incongruity, first noted by Beetz et al.,<sup>15</sup> may derive from an unforeseen quality of the *SPAST* locus and/or an unexpected feature of the mutational mechanism(s) creating these intragenic CNV that favors deletions, although more likely from a bias of ascertainment; i.e., perhaps MPLA, the predominant CNV detection method in previous studies (Table 1), fails to detect copy-number gains with the same accuracy as losses. Using locus-specific aCGH (as in this study) or genome-wide aCGH with exonic coverage<sup>19</sup> may improve detection of duplication and higher order gain CNVs at the *SPAST* locus, although detecting copy-number gains using aCGH and the clinical interpretation of such CNV gains each possesses its own challenges.<sup>49,50</sup>

It is also of note that neither intronic CNVs in *SPAST* nor CNVs in flanking intergenic regions, including regulatory regions, have been reported in SPG4 patients. CNV in noncoding regions such as these have been shown to cause disease in an increasing number of conditions.<sup>39,51,52</sup> As MLPA, the diagnostic methodology by which the majority of intragenic CNVs in *SPAST* has been discovered (Table 1), is usually performed in a strictly exon-targeted manner, it is an insufficient method to detect such CNV, should they exist in SPG4 patients. In con-

trast, aCGH with sufficient coverage of noncoding regions of and near *SPAST* may be a more appropriate methodology if all SPG4-causing mutations are to be detected.

This is the first report of a complex rearrangement within *SPAST* that spans an exon. Given recent data suggesting the importance of replication-based mechanisms in generating copy-number change, for example, MMBIR/FoSTeS/*Alu*-specific microhomology-mediated copy-number change<sup>25,26</sup> and, specifically, the complex rearrangements they can generate,<sup>29</sup> one might expect that some number of the CNVs in Table 1 may in fact represent complex rearrangements. Such complexity, if it exists, has likely gone undetected, as MLPA enables detection of only the presence or absence of assayed exons.

In summary, local genome architecture, including distribution and frequency of repetitive elements such as *Alu*, may facilitate genomic rearrangements. Such rearrangements can have consequences not only for the involved gene but also for other genes in proximity to the breakpoints. We hypothesize that the high concentration of *Alu* family members in the introns and flanking sequence of *SPAST* may predispose to intragenic rearrangements in this gene, such as those described in this study. Thus, *Alu*-specific microhomology-mediated intragenic rearrangements in *SPAST* may be a common cause of SPG4. Furthermore, we speculate that genomic deletions encompassing the final exon of *SPAST* may affect expression of *SLC30A6*, the most proximal downstream locus, and a gene that has been implicated in the pathogenesis of Alzheimer disease, potentially explaining recent reports of dementia in SPG4 patients with this particular exon deletion.

#### ACKNOWLEDGMENTS

P.M.B. is supported by the Baylor College of Medicine Medical Scientist Training Program (T32GM007330-34). This work was supported in part by a National Institute of Neurological Disorders and Stroke (NINDS) Grant (R01NS058529) from the United States National Institutes of Health (J.R.L.). J.R.L. is a paid consultant for Athena Diagnostics and Ion Torrent Systems and is a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the chromosomal microarray analysis offered in the Medical Genetics Laboratory.

#### REFERENCES

- Salinas S, Proukakis C, Crosby A, Warner TT. Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. *Lancet Neurol* 2008;7:1127–1138.
- HUGO Gene Nomenclature Committee. HGNC Home Page, 2010. Available at: <http://www.genenames.org>. Accessed October 15, 2010.
- McDermott CJ, Burness CE, Kirby J, et al. Clinical features of hereditary spastic paraplegia due to spastin mutation. *Neurology* 2006;67:45–51.
- Hazan J, Fonknechten N, Mavel D, et al. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat Genet* 1999;23:296–303.
- Sauter S, Mitterski B, Klimpe S, et al. Mutation analysis of the *spastin* gene (*SPG4*) in patients in Germany with autosomal dominant hereditary spastic paraplegia. *Hum Mutat* 2002;20:127–132.
- Mitne-Neto M, Kok F, Beetz C, et al. A multi-exonic *SPG4* duplication underlies sex-dependent penetrance of hereditary spastic paraplegia in a large Brazilian pedigree. *Eur J Hum Genet* 2007;15:1276–1279.
- Salinas S, Carazo-Salas RE, Proukakis C, et al. Human spastin has multiple microtubule-related functions. *J Neurochem* 2005;95:1411–1420.
- Roll-Mecak A, Vale RD. Structural basis of microtubule severing by the hereditary spastic paraplegia protein spastin. *Nature* 2008;451:363–367.
- Connell JW, Lindon C, Luzzio JP, Reid E. Spastin couples microtubule severing to membrane traffic in completion of cytokinesis and secretion. *Traffic* 2009;10:42–56.
- Park SH, Zhu P-P, Parker RL, Blackstone C. Hereditary spastic paraplegia proteins REEP1, spastin, and atlastin-1 coordinate microtubule interactions with the tubular ER network. *J Clin Invest* 2010;120:1097–1110.
- Svenson IK, Ashley-Koch AE, Gaskell PC, et al. Identification and expression analysis of spastin gene mutations in hereditary spastic paraplegia. *Am J Hum Genet* 2001;68:1077–1085.
- Depienne C, Fedirko E, Forlani S, et al. Exon deletions of *SPG4* are a frequent cause of hereditary spastic paraplegia. *J Med Genet* 2007;44:281–284.
- Bönsch D, Schwindt A, Navratil P, et al. Motor system abnormalities in hereditary spastic paraparesis type 4 (SPG4) depend on the type of mutation in the spastin gene. *J Neurol Neurosurg Psychiatry* 2003;74:1109–1112.
- Erichsen AK, Inderhaug E, Matningsdal M, Eiklid K, Tallaksen CM. Seven novel mutations and four exon deletions in a collection of Norwegian patients with SPG4 hereditary spastic paraplegia. *Eur J Neurol* 2007;14:809–814.
- Beetz C, Nygren AOH, Schickel J, et al. High frequency of partial *SPAST* deletions in autosomal dominant hereditary spastic paraplegia. *Neurology* 2006;67:1926–1930.
- Shoukier M, Neesen J, Sauter SM, et al. Expansion of mutation spectrum, determination of mutation cluster regions and predictive structural classification of *SPAST* mutations in hereditary spastic paraplegia. *Eur J Hum Genet* 2009;17:187–194.
- Svenstrup K, Bross P, Koefoed P, et al. Sequence variants in *SPAST*, *SPG3A* and *HSPD1* in hereditary spastic paraplegia. *J Neurol Sci* 2009;284:90–95.
- Hewamadduma CAA, Kirby J, Kershaw C, et al. HSP60 is a rare cause of hereditary spastic paraparesis, but may act as a genetic modifier. *Neurology* 2008;70:1717–1718.
- Boone PM, Bacino CA, Shaw CA, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010;31:1326–1342.
- Zhang F, Gu W, Hurler ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 2009;10:451–481.
- Deininger PL, Batzer MA. Alu repeats and human disease. *Mol Genet Metab* 1999;67:183–193.
- Lehman MA, Russell DW, Goldstein JL, Brown MS. Alu-Alu recombination deletes splice acceptor sites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. *J Biol Chem* 1987;262:3354–3361.
- Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131:1235–1247.
- Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet* 2009;5:e1000327.
- Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, gene and exonic complex rearrangements in humans. *Nat Genet* 2009;41:849–853.
- Erez A, Patel AJ, Wang X, et al. *Alu*-specific microhomology-mediated deletions in *CDKL5* in females with early-onset seizure disorder. *Neurogenetics* 2009;10:363–369.
- Visser LELM, Bhatt SS, Janssen IM, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 2009;18:3579–3593.
- Stankiewicz P, Sen P, Bhatt SS, et al. Genomic and gene deletions of the FOX gene cluster on 16q24.1 and inactivating mutations of *FOXF1* cause alveolar capillary dysplasia and other malformations. *Am J Hum Genet* 2009;84:780–791. Erratum 85:537.
- Zhang F, Carvalho CMB, Lupski JR. Complex human chromosomal and genomic rearrangements. *Trends Genet* 2009;25:298–307.
- Bi W, Sapir T, Shchelochkov OA, et al. Increased LIS1 expression affects human and mouse brain development. *Nat Genet* 2009;41:168–177.
- Iwanaga H, Tsujino A, Shirabe S, et al. Large deletion involving the 5'-UTR in the spastin gene caused mild phenotype of autosomal dominant hereditary spastic paraplegia. *Am J Med Genet A* 2005;133A:13–17.
- Gonzaga-Jauregui C, Zhang F, Towne CF, Batish SD, Lupski JR. GJB1/Connexin 32 whole gene deletions in patients with X-linked Charcot-Marie-Tooth disease. *Neurogenetics* 2010;11:465–470.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
- Beetz C, Zuchner S, Ashley-Koch A, et al. Linkage to a known gene but no mutation identified: comprehensive reanalysis of *SPG4* HSP pedigrees reveals large deletions as the sole cause. *Hum Mutat* 2007;28:739–740.
- de Bot ST, van den Elzen RTM, Mensenkamp AR, et al. Hereditary spastic paraplegia due to SPAST mutations in 151 Dutch patients: new clinical aspects and 27 novel mutations. *J Neurol Neurosurg Psychiatry* 2010;81:1073–1078.
- Magariello A, Muglia M, Patitucci A, et al. Mutation analysis of the *SPG4* gene in Italian patients with pure and complicated forms of spastic paraplegia. *J Neurol Sci* 2010;288:96–100.
- Battini R, Fogli A, Borghetti D, et al. Clinical and genetic findings in a series of Italian children with pure hereditary spastic paraplegia. *Eur J Neurol* 2011;18:150–157.

38. McCarroll SA, Huett A, Kuballa P, et al. Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nat Genet* 2008;40:1107–1112.
39. Zhang F, Seeman P, Liu P, et al. Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am J Hum Genet* 2010;86:892–903.
40. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008;320:539–543.
41. Lifton RP, Dluhy RG, Powers M, et al. A chimaeric 11  $\beta$ -hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* 1992;355:262–265.
42. Huang L, Kirschke CP, Gitschier J. Functional characterization of a novel mammalian zinc transporter, ZnT6. *J Biol Chem* 2002;277:26389–26395.
43. Lyubartseva G, Smith JL, Markesbery WR, Lovell MA. Alterations of zinc transporter proteins ZnT-1, ZnT-4 and ZnT-6 in preclinical Alzheimer's disease brain. *Brain Pathol* 2010;20:343–350.
44. Zhang L-H, Wang X, Stoltenberg M, Danscher G, Huang L, Wang Z-Y. Abundant expression of zinc transporters in the amyloid plaques of Alzheimer's disease brain. *Brain Res Bull* 2008;77:55–60.
45. Lovell MA, Smith JL, Markesbery WR. Elevated zinc transporter-6 in mild cognitive impairment, Alzheimer disease, and pick disease. *J Neuropathol Exp Neurol* 2006;65:489–498.
46. Smith JL, Xiong S, Markesbery WR, Lovell MA. Altered expression of zinc transporters-4 and -6 in mild cognitive impairment, early and late Alzheimer's disease brain. *Neuroscience* 2006;140:879–888.
47. Murphy S, Gorman G, Beetz C, et al. Dementia in SPG4 hereditary spastic paraplegia: clinical, genetic, and neuropathologic evidence. *Neurology* 2009;73:378–384.
48. Metzberg AB, Wurzer G, Huisman THJ, Smithies O. Homology requirements for unequal crossing over in humans. *Genetics* 1991;128:143–161.
49. Stankiewicz P, Pursley AN, Cheung SW. Challenges in clinical interpretation of microduplications detected by array CGH analysis. *Am J Med Genet A* 2010;152A:1089–1100.
50. Bacino CA, Cheung S-W. Introductory comments on special section—Genomic microduplications: when adding may equal subtracting. *Am J Med Genet A* 2010;152A:1063–1065.
51. Weterman MAJ, van Ruisven F, de Wissel M, et al. Copy number variation upstream of *PMP22* in Charcot-Marie-Tooth disease. *Eur J Hum Genet* 2010;18:421–428.
52. Wang LL, Worley K, Gannavarapu A, Chintagumpala MM, Levy ML, Plon SE. Intron-size constraint as a mutational mechanism in Rothmund-Thomson syndrome. *Am J Hum Genet* 2002;71:165–167.