OXFORD

Human Molecular Genetics, 2015, Vol. 24, No. 14 4061–4077

doi: 10.1093/hmg/ddv146
Advance Access Publication Date: 23 April 2015
Original Article

ORIGINAL ARTICLE

Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3

Shen Gu[1], Bo Yuan[1], Ian M. Campbell[1], Christine R. Beck[1], Claudia M.B. Carvalho[1], Sandesh C.S. Nagamani[1,4], Ayelet Erez[1,5], Ankita Patel[1], Carlos A. Bacino[1,4], Chad A. Shaw[1], Paweł Stankiewicz[1], Sau Wai Cheung[1], Weimin Bi[1] and James R. Lupski[1,2,3,4,*]

[1]Department of Molecular & Human Genetics, [2]Department of Pediatrics and [3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA, [4]Texas Children's Hospital, Houston, TX 77030, USA and [5]Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

*To whom correspondence should be addressed. Tel: +1 7137986530; Fax: +1 7137985073; Email: jlupski@bcm.edu

Abstract

Alu repetitive elements are known to be major contributors to genome instability by generating *Alu*-mediated copy-number variants (CNVs). Most of the reported *Alu*-mediated CNVs are simple deletions and duplications, and the mechanism underlying *Alu*–*Alu*-mediated rearrangement has been attributed to non-allelic homologous recombination (NAHR). Chromosome 17 at the p13.3 genomic region lacks extensive low-copy repeat architecture; however, it is highly enriched for *Alu* repetitive elements, with a fraction of 30% of total sequence annotated in the human reference genome, compared with the 10% genome-wide and 18% on chromosome 17. We conducted mechanistic studies of the 17p13.3 CNVs by performing high-density oligonucleotide array comparative genomic hybridization, specifically interrogating the 17p13.3 region with ~150 bp per probe density; CNV breakpoint junctions were mapped to nucleotide resolution by polymerase chain reaction and Sanger sequencing. Studied rearrangements include 5 interstitial deletions, 14 tandem duplications, 7 terminal deletions and 13 complex genomic rearrangements (CGRs). Within the 17p13.3 region, *Alu*–*Alu*-mediated rearrangements were identified in 80% of the interstitial deletions, 46% of the tandem duplications and 50% of the CGRs, indicating that this mechanism was a major contributor for formation of breakpoint junctions. Our studies suggest that *Alu* repetitive elements facilitate formation of non-recurrent CNVs, CGRs and other structural aberrations of chromosome 17 at p13.3. The common observation of *Alu*-mediated rearrangement in CGRs and breakpoint junction sequences analysis further demonstrates that this type of mechanism is unlikely attributed to NAHR, but rather may be due to a recombination-coupled DNA replicative repair process.

Introduction

Primate-specific *Alu* repetitive elements, although short in length (~300 bp), represent one of the most successful transposable elements in the human genome. As a member of the retrotransposon family, *Alu* elements use a reverse-transcribed RNA intermediate to insert into different positions in the genome through a 'copy-and-paste' mechanism (1). Overall, these elements comprise about 10% of the human genome with more than 1 million copies (2).

Through two different approaches, *Alu* elements are major contributors to genomic diversity and/or genome instability.

Received: January 13, 2015. **Revised:** April 13, 2015. **Accepted:** April 20, 2015

The insertional activity of *Alu* elements can lead to disruption of a gene either in its coding or regulatory regions, causing various types of diseases (3,4). In addition to insertional mutagenesis, the sequence homology between *Alu* elements (average 71%) (5) provides substrates to potentially facilitate unequal crossover between genomic segments and generates *Alu*–*Alu*-mediated copy-number variants (CNVs). *Alu* dimorphism within a personal genome, the presence of only one *Alu* at an allelic position of a diploid genome, is the most frequently observed deletion in the allelic spectrum of CNV < 1000 bp in size (6). The resulting gain or loss of a certain genomic region due to rearrangements can lead to many human disease states (7).

Alu-mediated CNVs were found in both germ-line genomic disorders (8–12) and somatic mutations in malignancies (13,14). This mode of mutagenesis was estimated to account for 0.3% of human genetic diseases, but this incidence was believed to be underestimated (7). Among the observed *Alu*-mediated germline CNVs, the majority of those characterized were deletions, with some duplications presented. Although, for some cases, there may be additional complexities at the breakpoint(s) (15,16), almost all these CNVs appeared to be simple interstitial deletions or tandem duplications. The *Alu*–*Alu* mechanism for generation of CNVs has seldom been described in association with complex genomic rearrangements (CGRs) having multiple copy-number transitions. Only a few examples were described, including one *Alu*-mediated breakpoint junction in a patient with triplication at 17p11.2 (17), one of the two breakpoint junctions in a patient with deletion–normal–duplication at 17p12 (18), one identical breakpoint junction shared in three patients with duplication–normal–duplication at Xp22 (19) and one junction mapped in a patient with duplication–normal–duplication–normal–duplication at Xq26 (20).

The mechanism underlying *Alu*–*Alu*-mediated rearrangement, which is defined by the formation of a new chimeric *Alu* hybrid, has been attributed to non-allelic homologous recombination (NAHR) (21,22). However, the low degree of similarity between *Alu* pairs mediating rearrangements is inconsistent with the hypothesis that this is purely mediated by homologous recombination (16). Replicative mechanisms, such as fork stalling and template switching (FoSTeS)/microhomology-mediated break-induced replication (MMBIR), that are mediated by short stretches of microhomology could potentially explain *Alu*–*Alu*-mediated rearrangements (23,24). Likewise, FoSTeS/MMBIR also provides a parsimonious explanation for template switching, contributing to one or multiple iterations thereof, and to *Alu*-induced CGRs other than simple deletions and tandem duplications (17,18).

The gene-rich 17p13.3 region has been associated with various deletion and duplication CNV syndromes or genomic disorders, including Miller–Dieker syndrome (MDS; MIM #247200), 17p13.3 duplication syndrome (MIM #613215), 17p13.3 microdeletion syndrome (25,26) and split-hand/foot malformation with long bone deficiency 3 (SHFLD3; MIM #612576). The majority of previously reported 17p13.3 CNVs were interstitial deletions/duplications and terminal deletions, with only five CGRs (25,27,28). Moreover, CNVs with breakpoint junctions mapped to nucleotide resolution were primarily deletions (10,26).

We conducted comprehensive breakpoint mapping and junctional analyses to gain insights into mutational mechanisms in the 17p13.3 region. We investigated 17p13.3 CNVs identified in 39 unrelated individuals from 33 800 patient samples that were sent for clinical diagnostic testing at the Medical Genetics Laboratories of Baylor College of Medicine. Human chromosome 17 at p13.3 is highly enriched for *Alu* repetitive elements and indeed, breakpoint sequencing of these CNVs revealed that a remarkable proportion of the junctions were mediated by an *Alu*–*Alu* mechanism, including 50% of the breakpoint junctions in CGRs with evidence suggesting multiple template switches. Our studies reveal a prominent role for *Alu*–*Alu*-mediated genomic rearrangement events and further support the contention that replicative mechanisms play an important role in formation of CGRs.

Results

Interstitial deletions and duplications of 17p13.3 vary in size and genomic position

Five of the 39 rearrangements interrogated from individual personal genomes had simple interstitial deletions, ranging in size from 0.545 to 0.746 Mb (Fig. 1, Supplementary Material, Fig. S1 and Table 1). Deletions in two individuals (19 and 10) encompassed *YWHAE*. Four of the five deletions (in individuals 19, 10, 30 and 15) had breakpoint junctions localized to directly oriented *Alu* pairs with a chimeric *Alu* hybrid being formed (Fig. 1, Supplementary Material, Fig. S1 and Table 1). The remaining deletion in individual BAB3106 had 6 bp of microhomology between the proximal and distal reference sequences at the breakpoint junction (Supplementary Material, Fig. S1C).

Within the 17p13.3 region, 13 of the CNVs were tandem duplications, ranging from 0.018 to 0.921 Mb in size (Fig. 2, Supplementary Material, Fig. S2 and Table 1). In addition to variation in size, these duplications that varied in both genomic position and gene content demonstrated no evidence for recurrent events. The majority of the duplications (11 out of 13) involved *YWHAE* only, whereas one duplication (in individual 14) encompassed *PAFAH1B1*. Two duplications (in individuals 3 and 21) involved *RPA1*, increased dosage of which may result in the increased genomic instability associated with 17p13.3 duplication syndrome (29). Six of the 13 tandem duplications (in individuals 8, 1, 22, BAB3054, 26 and 14) had breakpoint junctions located in the directly oriented pairs of *Alu* repetitive elements, generating chimeric *Alu* upon tandem duplication (Fig. 2, Supplementary Material, Fig. S2 and Table 1). The remaining duplications (in individuals 29, 3, 9, 36, 7, 31 and 21) had 2–6 bp microhomology at breakpoint junctions. Notably, duplication in individual 9 may involve a replication slippage during the rearrangement process (Supplementary Material, Fig. S2A). In contrast to *Alu*–*Alu*-mediated CNVs, the two *Alu* repeats located at each side of the breakpoint junction of DNA from individual 29 lie in opposite orientation in the haploid human reference (Table 1), which is unlikely to result in an *Alu*–*Alu*-mediated tandem duplication (Supplementary Material, Fig. S2D). A large interstitial duplication case was found in individual 5, spanning 15.16 Mb from 17p12 to 17p13.3. This rearrangement appeared to be generated by LINE–LINE-mediated recombination (LINE, long interspersed nuclear element. Supplementary Material, Fig. S2M) (22,30).

Complex rearrangements of 17p13.3: duplication–normal–duplication

Among the 13 individuals with complex rearrangements (individuals 2, 4, 6, 17, 18, 20, 23, 24, 27, 28, 33, K2 and BAB3886; see Table 1), 4 had array comparative genomic hybridization (aCGH) results consistent with two duplications separated by a copy-number neutral region (individuals 2, 4, 6 and BAB3886; Fig. 3A). Breakpoint junction analysis further revealed that DNA from individual 2 had a duplication–normal–duplication (DUP–NML–DUP)

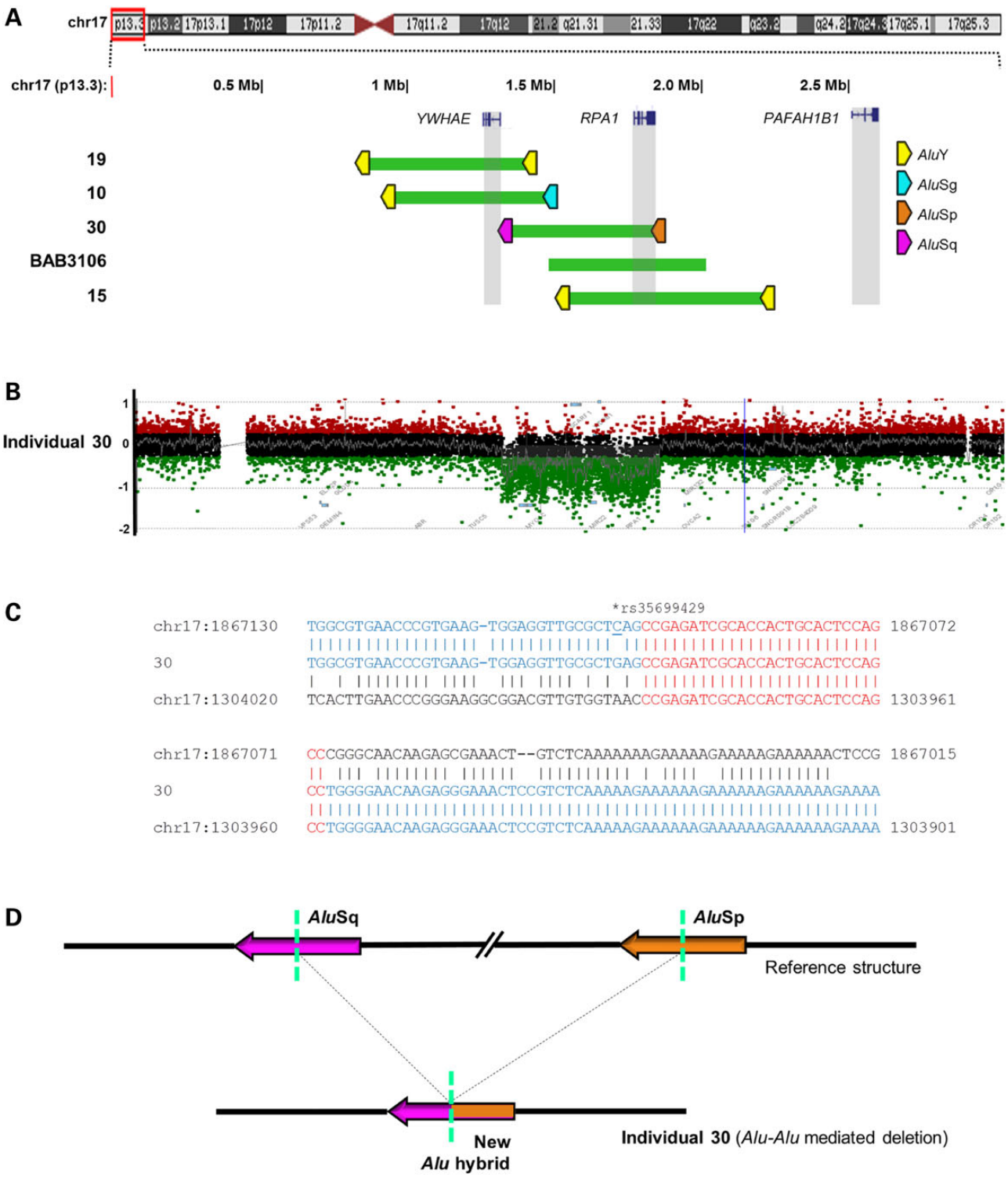


Figure 1. Interstitial deletions of 17p13.3. (**A**) Deletions were plotted by exact genomic coordinates according to breakpoint junction sequencing results using the UCSC genome browser custom track function (GRCh37/hg19 assembly). Green blocks represent deletions. Upper panel shows an ideogram of human chromosome 17 with the 17p13.3 region involved highlighted in the red box. Three critical genes in 17p13.3 deletion and duplication syndromes, *YWHAE*, *RPA1* and *PAFAH1B1*, are highlighted in gray vertical boxes. *Alu* pairs that mediated the deletions are denoted by left-facing (– strand) pentagons. Whether an *Alu* element presents on the plus or minus strand is defined according to the UCSC genome browser (GRCh37/hg19 assembly). Pentagons filled with different colors indicate different *Alu* families. (**B**, **C**) Representative aCGH $\log_2$ plot and breakpoint junction sequence of individual 30 with an interstitial deletion. Breakpoint junction sequence is aligned to the proximal and distal genomic references and color-matched. Microhomology at the breakpoint is indicated in red. (**D**) Proposed rearrangement of *Alu*–*Alu*-mediated deletion in individual 30. Array plots and breakpoint sequences of remaining interstitial deletions are shown in Supplementary Material, Figure S1.

Table 1. Features of the CNVs studied in 39 subjects with 17p13.3 rearrangements

Individual[a]	CNV type	CNV size (Mb)	Distal repeat	Proximal repeat	Microhomology (bp)[b]	Notes[c]
10	DEL	0.595	*Alu*Sg/–	*Alu*Yk4/–	17	*Alu–Alu* (82.2%)
15	DEL	0.746	*Alu*Yk4/–	*Alu*Y/–	43	*Alu–Alu* (90.6%)
19	DEL	0.611	*Alu*Y/–	*Alu*Yb8/ –	32	*Alu–Alu* (88.6%)
30	DEL	0.563	*Alu*Sq/–	*Alu*Sp/–	27	*Alu–Alu* (86.24%)
BAB3106	DEL	0.545	–	–	6	NHEJ/MMEJ
1	DUP	0.512	*Alu*Sp/+	*Alu*Sp/+	52	*Alu–Alu* (89.94%)
3	DUP	0.742	–	L2/+	4	MMEJ/FoSTeS/MMBIR
5	DUP	15.161	L1PA5/–	L1PA5/–	17	LINE/LINE
7	DUP	0.37	*Alu*Sx1/+	–	6	MMEJ/FoSTeS/MMBIR
8	DUP	0.455	*Alu*Jr/+	*Alu*Sx1/+	6	*Alu–Alu* (78.2%)
9	DUP	0.206	–	–	N	MMEJ/FoSTeS/MMBIR
14	DUP	0.921	*Alu*Sx/–	*Alu*Sz/–	3	*Alu–Alu* (84.1%)
21	DUP	0.146	*Alu*Sx/+	–	2	MMEJ/FoSTeS/MMBIR
22	DUP	0.341	*Alu*Y/–	*Alu*Yk4/–	46	*Alu–Alu* (90.16%)
26	DUP	0.018	*Alu*Sq/+	*Alu*Y/+	33	*Alu–Alu* (87.37%)
29	DUP	0.879	*Alu*Sq2/–	*Alu*Sx3/+	2	MMEJ/FoSTeS/MMBIR
31	DUP	0.028	*Alu*Sq2/+	LIMB4/+	2	MMEJ/FoSTeS/MMBIR
36	DUP	0.413	–	*Alu*Sx/+	3	MMEJ/FoSTeS/MMBIR
BAB3054	DUP	0.092	*Alu*Sg/–	*Alu*Y/–	19	*Alu–Alu* (84.21%)
2_1	DUP–NML–DUP	Complex	*Alu*Y/+	*Alu*Sc8/+	29	*Alu–Alu* (91.5%)
2_2	DUP–NML–DUP	Complex	*Alu*Sc/+	*Alu*Sc/+	9	*Alu–Alu* (91.5%)
4_1	DUP–NML–INV/DUP	Complex	*Alu*Y/–	*Alu*Sx/+	6	FoSTeS/MMBIR
4_2	DUP–NML–INV/DUP	Complex	*Alu*Sg/–	*Alu*Sc/+	34	*Alu–Alu* (89.13%)
6_1	DUP–NML–INV/DUP	Complex	*Alu*Y/–	*Alu*Y/+	39	*Alu–Alu* (89.6%)
6_2	DUP–NML–INV/DUP	Complex	*Alu*Y/–	L2b/+	N	FoSTeS/MMBIR
BAB3886_1	DUP–NML–INV/DUP	Complex	*Alu*Y/–	*Alu*Sc5/+	29	*Alu–Alu* (87.91%)
BAB3886_2	DUP–NML–INV/DUP	Complex	–	L2/–	N	FoSTeS/MMBIR
27_1	DUP-TRP/INV-DUP	Complex	*Alu*Jb/+	–	N	FoSTeS/MMBIR
27_2	DUP-TRP/INV-DUP	Complex	LTR8/+	L1MC2/+	2	FoSTeS/MMBIR
28_1	DUP-TRP/INV-DUP	Complex	–	–	2	FoSTeS/MMBIR
28_2	DUP-TRP/INV-DUP	Complex	*Alu*Sg/–	*Alu*Sg/+	21	*Alu–Alu* (84.69%)
33_1	DUP-TRP/INV-DUP	Complex	*Alu*Sx3/–	*Alu*Sz/+	25	*Alu–Alu* (85.14%)
33_2	DUP-TRP/INV-DUP	Complex	–	(TGTG)n/+	2	FoSTeS/MMBIR
K2_1	DUP-TRP-DUP	Complex	*Alu*Sz/–	*Alu*Y/–	5	*Alu–Alu* (76.32%)
K2_2	DUP-TRP-DUP	Complex	*Alu*Sx/ –	*Alu*Sq/–	25	*Alu–Alu* (83.16%)
18	TRP	0.011	*Alu*Sq/–	*Alu*Sg/–	30	*Alu–Alu* (82.01%)
20	TRP	0.811	*Alu*Sx1/+	–	1	FoSTeS/MMBIR
23_1	Complex	Complex	MER103C/+	AluSg4/–	N	FoSTeS/MMBIR
23_2	Complex	Complex	–	–	2	FoSTeS/MMBIR
23_3	Complex	Complex	–	LIMC4a/–	N	FoSTeS/MMBIR
23_4	Complex	Complex	–	–	N	FoSTeS/MMBIR
23_5	Complex	Complex	AluSx/+	LIPREC2/+	N	FoSTeS/MMBIR
23_6	Complex	Complex	MARNA/+	LIME4a/+	N	FoSTeS/MMBIR
24	Complex	Complex	*Alu*Y/–	*Alu*Sq2/–	4	Parental pericentric inversion
17_1	DEL-NML-DEL	0.545	–	–	6	FoSTeS/MMBIR
17_2	DEL-NML-DEL	0.007	*Alu*Y/–	–	N	FoSTeS/MMBIR
11	Terminal DEL	2.92	N/A	L1MA4	0	Telomeric healing
12	Terminal DEL	5.06	N/A	AluSc/–	3	Telomeric healing
25[d]	Terminal DEL	2.574	N/A	N/A	N/A	
35	Terminal DEL	1.964	N/A	–	4	Telomeric healing
38	Terminal DEL	1.379	N/A	–	2	Telomeric healing
BAB3291	Terminal DEL	2.46	N/A	*Alu*Jo/+	2	Telomeric healing
K1	Terminal DEL	2.46	N/A	*Alu*Sp/+	1	Telomeric healing

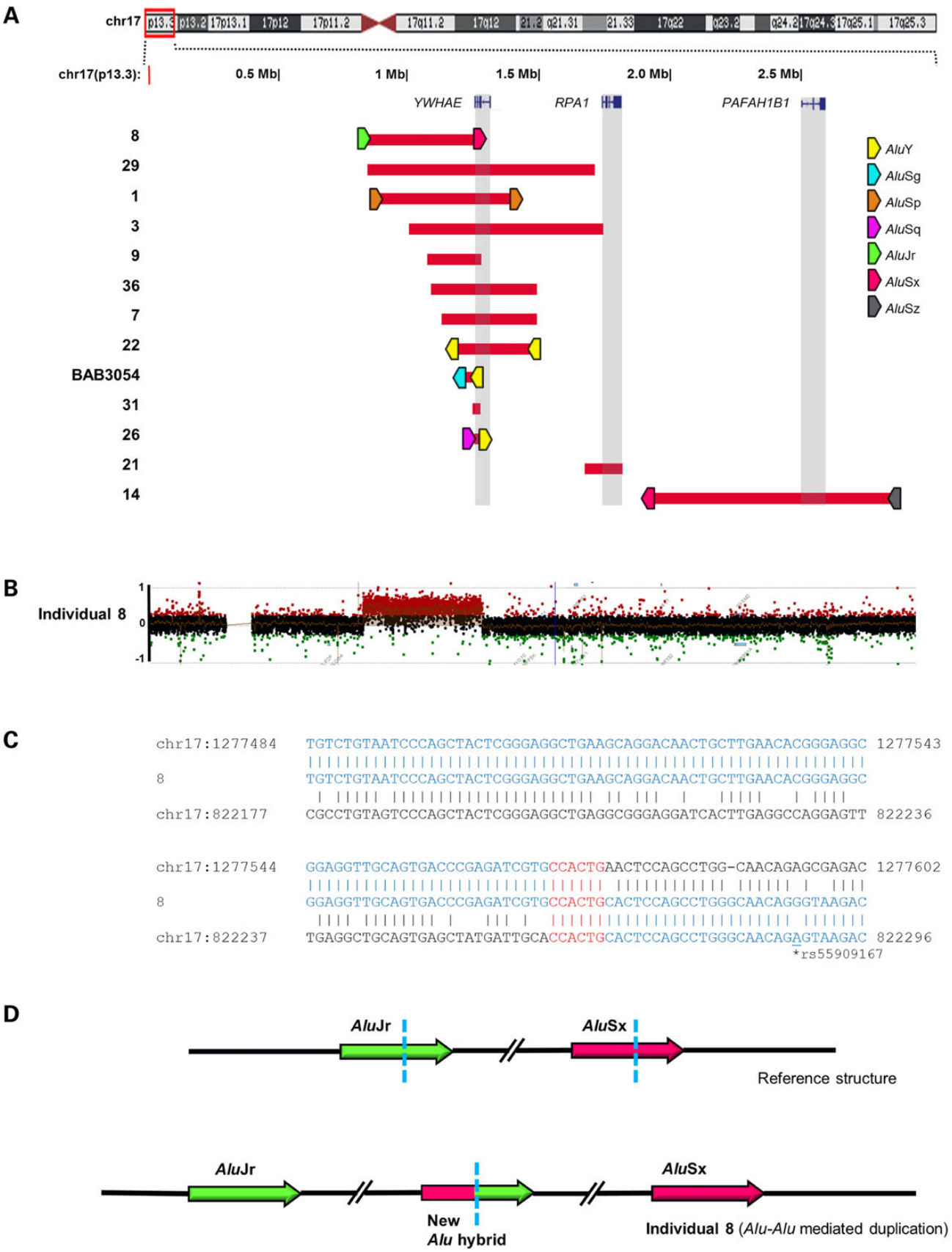
Abbreviations: /+, repeat locates on plus strand; /–, repeat locates on negative strand; *Alu–Alu*, *Alu–Alu*-mediated rearrangement; DEL, deletion; DUP, duplication; MMBIR, microhomology-mediated break-induced replication; FoSTeS, fork stalling and template switching; NHEJ, non-homologous end joining; MMEJ, microhomology-mediated end joining; N/A, not applicable.

[a]For individuals with complex CNVs, there may be more than one breakpoint junctions aligned.

[b]Number of bp of microhomology sequence not shown due to multiple template switches (indicated by N). For detailed breakpoint junction sequences alignment, see supplementary figures.

[c]*Alu–Alu*, *Alu–Alu*-mediated rearrangement as defined by the generation of a chimeric *Alu* hybrid; Similarity of the sequence between *Alu* pairs that generate the chimeric *Alu* is also listed (also see Fig. 7).

[d]We could not map the breakpoint of individual 25 with a terminal deletion, and therefore its breakpoint is unknown, as indicated by N/A.


A
chr17
chr17(p13.3):
0.5 Mb
1 Mb
1.5 Mb
2.0 Mb
2.5 Mb
YWHAE
RPA1
PAFAH1B1
8
29
1
3
9
36
7
22
BAB3054
31
26
21
14
AluY
AluSg
AluSp
AluSq
AluJr
AluSx
AluSz
B
Individual 8
C
chr17:1277484 TGTCTGTAATCCCAGCTACTCGGGAGGCTGAAGCAGGACAACTGCTTGAACACGGGAGGC 1277543
8 TGTCTGTAATCCCAGCTACTCGGGAGGCTGAAGCAGGACAACTGCTTGAACACGGGAGGC
chr17:822177 CGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCGGGAGGATCACTTGAGGCCAGGAGTT 822236
chr17:1277544 GGAGGTTGCAGTGACCCGAGATCGTGCCACTGAACTCCAGCCTGG-CAACAGAGCGAGAC 1277602
8 GGAGGTTGCAGTGACCCGAGATCGTGCCACTGCACTCCAGCCTGGGCAACAGGGTAAGAC
chr17:822237 TGAGGCTGCAGTGAGCTATGATTGCACCACTGCACTCCAGCCTGGGCAACAGAGTAAGAC 822296
*rs55909167
D
AluJr
AluSx
Reference structure
AluJr
AluSx
New
Alu hybrid
Individual 8 (Alu-Alu mediated duplication)

pattern, whereas the remaining three had one inverted duplication (DUP–NML–INV/DUP) (Fig. 3).

There are two potential explanations for the CNVs in individual 2 based on the breakpoint junction analysis. The first mechanism requires a combination of two alleles: the first allele is a tandem duplication of chromosome 17, whereas the second is a deletion of the homologous chromosome spanning the gap between the segments with increased copy number (Supplementary Material, Fig. S3A). The duplication or deletion could be inherited, however this hypothesis cannot be tested due to the absence of parental DNA. A second and simpler explanation of the mechanism requires only one allele generated by two template switches (Fig. 3B). Both the breakpoint junctions in individual 2 were situated in directly oriented pairs of *Alu* repetitive elements, generating chimeric *Alu* structures (Supplementary Material, Fig. S3A and Table 1).

Breakpoint junction analysis demonstrated a DUP–NML–INV/DUP pattern in individuals 4, 6 and BAB3886 (Fig. 3C, Supplementary Material, Figs S3 and S4). Based on the breakpoint junctions, the centromeric end of the first duplication was connected to that of the second duplication (junction 1 in Fig. 3C and Supplementary Material, Fig. S4A), whereas the telomeric ends of the two duplications were also connected (junction 2 in Fig. 3C and Supplementary Material, Fig. S4A), resulting in one of the duplications being inverted (Supplementary Material, Figs S3 and S4) with respect to the genomic reference. There are two possible mechanisms to generate this type of DUP–NML–INV/DUP rearrangement. The first mechanism is a one-step event and requires two iterative template switches that occur on the reference genome haplotype (Fig. 3C). During replication, a template switch occurs to the opposite strand, creating breakpoint junction 1 and an inversion, followed by a switch back to the original strand to avoid the generation of a potential acentric chromosome (Supplementary Material, Fig. S4A) (31,32). The second potential mechanism for DUP–NML–INV/DUP requires the postulation of a tandem duplication on an inversion haplotype and reflects a potential 'artifact' of the appearance of a DUP–NML–DUP structure from aCGH given the haploid human genome reference and the absence of its incorporation of copy-number neutral structural variations (SVs), specifically inversion haplotypes (17). This latter model necessitates that the subject inherits an inversion haplotype and then acquires a simple tandem duplication of part of the inverted segment (Supplementary Material, Fig. S4A). According to this mechanism, one of the two parents would also harbor the inversion and breakpoint junction 1 (Supplementary Material, Fig. S4A). However, long-range polymerase chain reaction (PCR) analysis of BAB3887 and BAB3888 (the mother and father of BAB3886, respectively) failed to detect such a postulated junction (Supplementary Material, Fig. S4B). Additionally, if such an inversion were to have occurred, a third potential breakpoint junction would be formed (breakpoint junction 3 in Supplementary Material, Fig. S4A). Multiple PCR reactions failed to amplify such a breakpoint junction in any of the DUP–NML–INV/DUP individuals (data not shown). Therefore, the first one-step mechanism is the most likely explanation for the DUP–NML–INV/DUP pattern of rearrangement, however other potential rearrangement mechanisms may also be possible. Breakpoint junction analysis demonstrated that breakpoint junction 2 in individual 4 and breakpoint junction 1 in individuals 6 and BAB3886 appeared to be *Alu–Alu*-mediated. The remaining breakpoints showed microhomology (6 bp microhomology of breakpoint junction 1 in individual 4) or templated insertion from nearby sequences (breakpoint junctions 2 in individuals 6 and BAB3886) (Fig. 3D, Supplementary Material, Fig. S3B–D and Table 1).

Triplication generated through *Alu–Alu*-mediated events

Among the 13 individuals with complex rearrangements, three had simple triplications (individuals 20, K2 and 18), whereas three showed DUP–TRP–DUP patterns (individuals 27, 28 and 33) by aCGH (Fig. 4A and Supplementary Material, Fig. S5). The largest triplication found in individual 20 (0.81 Mb) involved *YWHAE* and part of *RPA1*. The triplication in K2 (0.32 Mb) involved only *YWHAE*, and the smallest triplication (in individual 18, 11 kb) harbored sequence of neither of the two genes critical for causing MDS. As generation of a triplication potentially requires two breakpoint junctions, both the breakpoint junctions of K2 were mapped to nucleotide resolution and both appeared to be *Alu–Alu*-mediated with directly oriented *Alu* pairs (Fig. 4B and Supplementary Material, Fig. S5B). Upon careful inspection of the two breakpoints in K2, both sides of the triplicated region contained short duplicated sequences (2413 bp on the distal side and 792 bp on the proximal side), therefore it was further classified as a DUP–TRP–DUP CGR (Supplementary Material, Fig. S5B). For the remaining two individuals with triplications, we were able to resolve one out of the two predicted breakpoint junctions for each. The mapped breakpoint junction in individual 18 indicated an *Alu–Alu*-mediated event due to the generation of a new *Alu* hybrid, whereas the mapped breakpoint junction in individual 20 showed 1 bp microhomology (Supplementary Material, Fig. S5A and C). Similar to K2 junctional sequences, the mapped breakpoints in individuals 20 and 18 involved distal and proximal references on the same strand, suggesting that triplicated segments were in the same orientation embedded in duplicated segments.

Based on the aCGH results, DNA from individuals 27, 28 and 33 contained triplication flanked by duplications (Fig. 4C). Breakpoint junction analysis revealed a consistent pattern for all three rearrangements—breakpoint junction 1 connects the centromeric ends of both the large duplication rearrangement and the triplication, whereas junction 2 joins the telomeric ends of these two, indicating that the triplicated region in the middle was inverted, resulting in a DUP-TRP/INV-DUP pattern (33). Breakpoint junction 2 in individual 28 and breakpoint 1 in individual 33 appeared to be mediated by opposite-oriented *Alu* repeats (or inverted repeats) that led to the inverted triplication (Supplementary Material, Fig. S5D–F). A 2 bp microhomology was found at breakpoint junction 2 in individual 27, breakpoint

Figure 2. Interstitial duplications of 17p13.3. (**A**) Duplications were plotted by exact genomic coordinates according to breakpoint junction sequencing results using the UCSC genome browser custom track function. Red blocks represent duplications. Upper panel shows an ideogram of human chromosome 17 with the 17p13.3 region involved highlighted in the red box. Three dosage-sensitive genes in 17p13.3 duplication syndromes, *YWHAE*, *RPA1* and *PAFAH1B1*, are highlighted in gray vertical boxes. *Alu* pairs that mediated the duplications are denoted by right-facing (+ strand) or left-facing (– strand) pentagons. Pentagons filled with different colors indicate different *Alu* families. (**B**, **C**) Representative aCGH $\log_2$ plot and breakpoint junction sequence of individual 8 with tandem duplication. Breakpoint junction sequence is aligned to the proximal and distal genomic references and color-matched. Microhomology at the breakpoint is indicated in red. (**D**) Proposed rearrangement of *Alu–Alu*-mediated duplication in individual 8. Array plots and breakpoint sequences of remaining tandem duplications are shown in Supplementary Material, Figure S2. Note that individual 5 with a tandem duplication spanning 15.16 Mb from 17p12 to 17p13.3 is not listed here due to its much larger size and one breakpoint mapping in a different cytogenetic and genomic region. Array plot and breakpoint sequences of individual 5 are also shown in Supplementary Material, Figure S2.

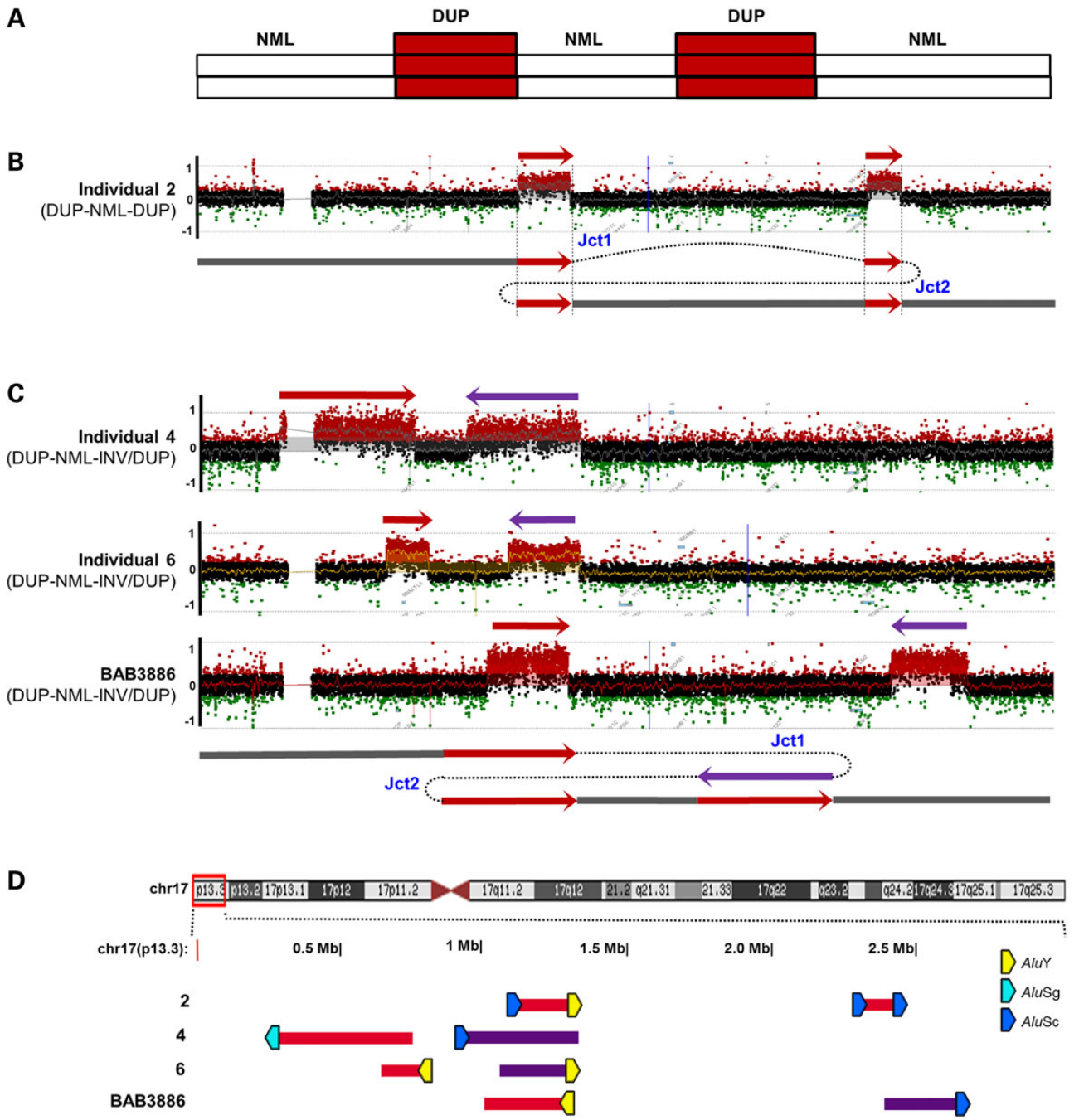


Figure 3. Complex rearrangements distinguished by a DUP–NML–DUP pattern of 17p13.3. (**A**) Illustration of the DUP–NML–DUP rearrangement pattern from aCGH results. (**B**) Array CGH $\log_2$ plot and proposed generation mechanism of DUP–NML–DUP rearrangement in individual 2. (**C**) Array CGH $\log_2$ plot and proposed generation mechanism of DUP–NML–INV/DUP rearrangement in individuals 4, 6 and BAB3886. (**D**) CNVs were plotted by exact genomic coordinates according to breakpoint junction sequencing results using the UCSC genome browser custom track function. Red blocks, duplication; Purple blocks, inverted duplication. Upper panel shows an ideogram of human chromosome 17 with the 17p13.3 region involved highlighted in the red box. *Alu* pairs that mediated the complex rearrangements are denoted by right-facing (+ strand) or left-facing (– strand) pentagons. Pentagons filled with different colors indicate different *Alu* families. Breakpoint sequences of all the DUP–NML–DUP individuals are shown in Supplementary Material, Figure S3. Jct1 and Jct2, junction 1 and junction 2.

junction 1 in individual 28 and breakpoint junction 2 in individual 33 (Supplementary Material, Fig. S5D–F).

Additional complex rearrangement patterns

The most complex rearrangement in this study was observed in individual 23. The rearrangement exhibited a 'DEL–NML–DUP–TRP–DUP–TRP–DUP–NML–DUP–TRP–DUP' pattern encompassing a significant portion of the short arm of chromosome 17 from 17p13.1 to 17p13.3 with multiple copy-number transition states and potentially six breakpoint junctions (Fig. 5 and Supplementary Material, Fig. S6A). We successfully mapped these breakpoint junctions to nucleotide resolution. Notably, although the copy-number changes were in 17p13.3, none of the breakpoint junctions had both distal and proximal reference sequences present at 17p13.3, but rather at 17p13.1 or 17p13.2. In addition, none

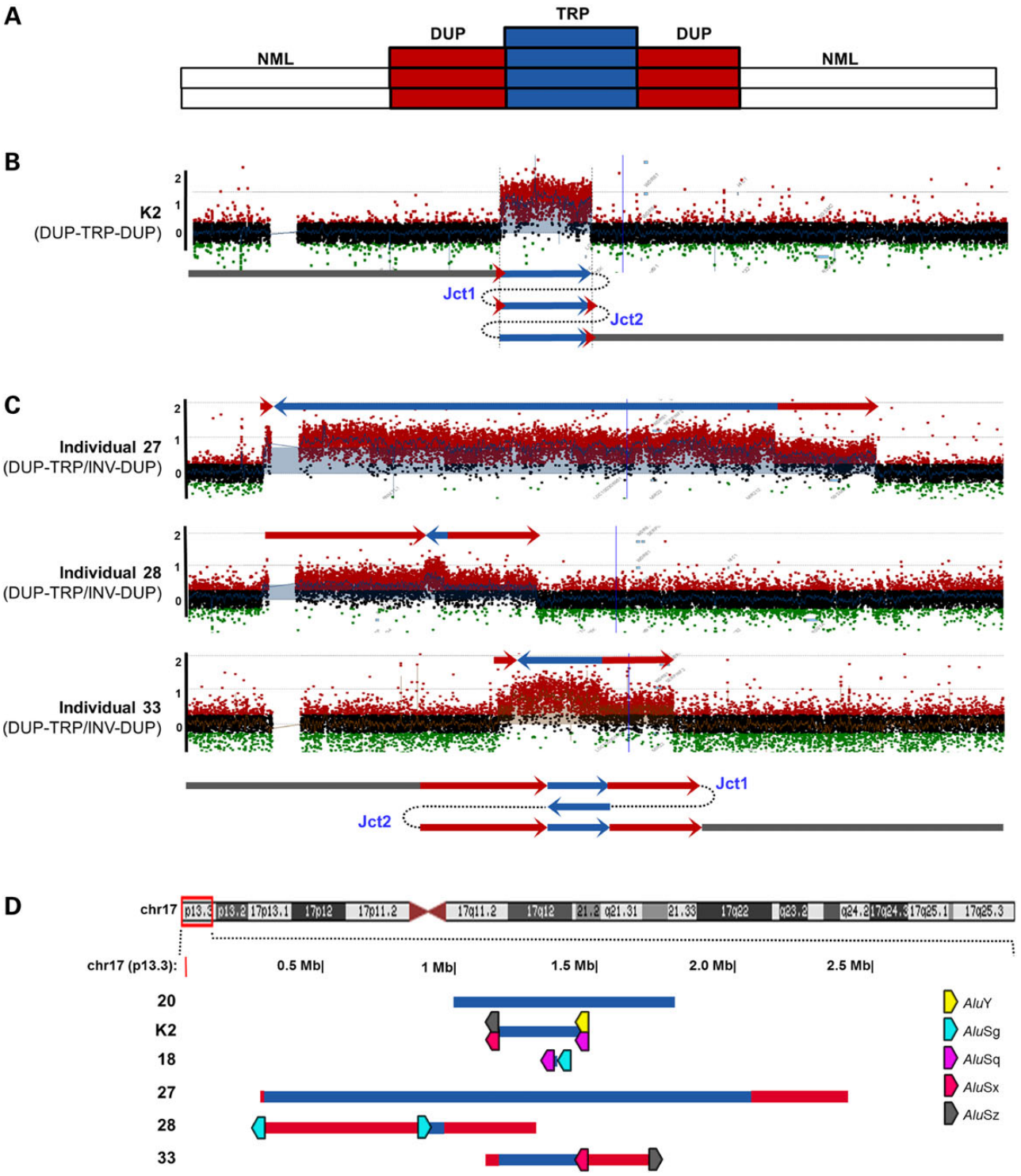


Figure 4. Triplications of 17p13.3. (**A**) Illustration of the DUP–TRP–DUP rearrangement pattern from aCGH results. (**B**) Array CGH $\log_2$ plot and proposed generation mechanism of triplication in individual K2. Note that both sides of the triplicated region contained short duplicated sequences (2413 bp in the distal side and 792 bp in the proximal side) as indicated by small red arrows. (**C**). Array CGH $\log_2$ plot and proposed generation mechanism of DUP-TRP/INV-DUP pattern rearrangement in individuals 27, 28 and 33. (**D**) CNVs were plotted by exact genomic coordinates according to breakpoint junction sequencing results using the UCSC genome browser custom track function. Red blocks, duplication; Blue blocks, triplication. Upper panel shows an ideogram of human chromosome 17 with the 17p13.3 region involved highlighted in the red box. *Alu* pairs that mediated the complex rearrangements are denoted by right-facing (+ strand) or left-facing (– strand) pentagons. Pentagons filled with different colors indicate different *Alu* families. Note that the triplication in individual K2 was generated by two breakpoints that both were *Alu*–*Alu*-mediated, and the two pairs of *Alu* repetitive elements are *AluSz*/*AluY* and *AluSx*/*AluSq*. Array plots and breakpoint sequences of additional cases with triplication are shown in Supplementary Material, Figure S5. Jct1 and Jct2, junction 1 and junction 2.

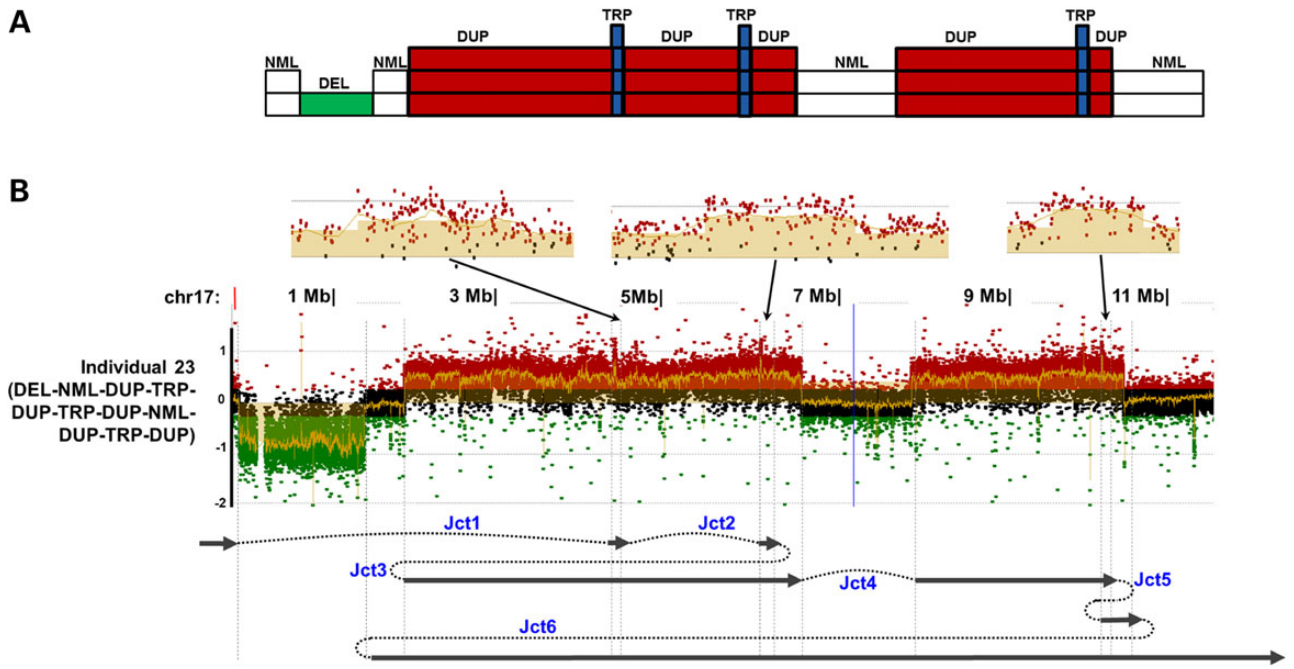


Figure 5. Individual 23 with complex rearrangement. (**A**) Illustration of the DEL–NML–DUP–TRP–DUP–TRP–DUP–NML–DUP–TRP–DUP rearrangement pattern in individual 23 from aCGH results. (**B**) Array CGH $\log_2$ plot and proposed generation mechanism of the complex rearrangement in individual 23. Enlarged array plots of the three small triplicated regions are shown on the top.

of the six breakpoint junctions appeared to be *Alu*–*Alu*-mediated, potentially indicating a reduced probability to form *Alu*–*Alu*-mediated CNVs outside the 17p13.3 region. Microhomology at breakpoint junction 2 (2 bp) and small templated insertions at breakpoint junction 1 (47 bp), junction 3 (15 bp), junction 4 (22 bp, mapped to another locus of chromosome 17), junction 5 (1 bp) and junction 6 (9 bp) were observed (Supplementary Material, Fig. S6A). The complexity of the rearrangement pattern and the small templated insertions at the breakpoints indicated a possible chromoanasynthesis/chromothripsis event (32,34) that was likely mediated by multiple template switches during the replication process (35).

Individual 17 had an interstitial 0.55 Mb deletion. Upon careful inspection, a 7 kb region with normal copy number was found within and near the distal end of the deletion (Supplementary Material, Fig. S6B). Breakpoint analysis revealed a joining of the distal and proximal normal regions of the 0.55 Mb deletion (junction 1 of individual 17) with microhomology (6 bp) observed. Analysis of the other breakpoint (junction 2) indicated that a deletion occurred on one homolog of the chromosome, whereas a tandem duplication of 7 kb was present on the other homolog, resulting in an overall DEL–NML–DEL pattern from the array.

Individual 24 exhibited a 4.1 Mb duplication at 17pter and a 0.69 Mb deletion at 17qter. Breakpoint junction analysis indicated that the duplicated 17pter region inverted and joined to the truncated chromosome 17 long arm (Supplementary Material, Fig. S6C). Similar rearrangements have previously been observed on chromosome 17, and a pericentric inversion in one of the parents likely underlies the rearrangement (36).

Telomeric healing of terminal deletions

Seven individuals in this study harbored terminal deletions of 17p13.3, ranging in size from 1.38 to 5.06 Mb (individuals 11, 12, 38, 25, 35, BAB3291 and K1, Fig. 6 and Supplementary Material, Fig. S7). All seven deletions contained *YWHAE*, and the deletions in individuals 25, 11 and 12 also involved *PAFAH1B1*. Individual 12 contained a larger 5.06 Mb terminal deletion spanning from 17p13.2 to 17p13.3 (Fig. 6 and Supplementary Material, Fig. S7F). Breakpoint junctions showed addition of telomere repeats $(TTAGGG)_n$ to the truncated chromosome 17, indicating telomerase-mediated healing events in all except individual 25 in whom we could not amplify the junction fragment (26,37).

Alu–*Alu*-mediated rearrangements

The mutational signatures present at each of the breakpoints are shown in Table 1. Strikingly, 47% of the breakpoints within the 17p13.3 region (with both distal and proximal reference sequence present within 17p13.3) were mediated by *Alu* elements as defined by the generation of a chimeric *Alu* hybrid. If one restricts the analysis to interstitial rearrangements, 53% of breakpoint junctions were *Alu*–*Alu*-mediated. Using the reference human genome and RepeatMasker annotations, we found that 30.0% of the 17p13.3 region is comprised of *Alu* elements; this is markedly higher than the 9.95% observed on a genome-wide scale and even higher than the 18.4% elsewhere on chromosome 17.

We hypothesized that particular *Alu* elements in the region are providing sequence similarity to serve as a substrate for a template switch or alternatively for a homologous or homeologous recombination. To investigate this, we performed Smith–Waterman alignments of the pairs of *Alu* actually mediating the CNVs identified in this study and compared their alignment scores to randomly selected *Alu* pairs in the region. We found that CNV-mediating pairs align significantly better than randomly selected pairs (Wilcoxon signed rank test, $P = 5 \times 10^{-6}$, Fig. 7A). Sequence alignment of the *Alu* pairs revealed microhomology/crossover locations throughout the *Alu* structures (Fig. 7B).

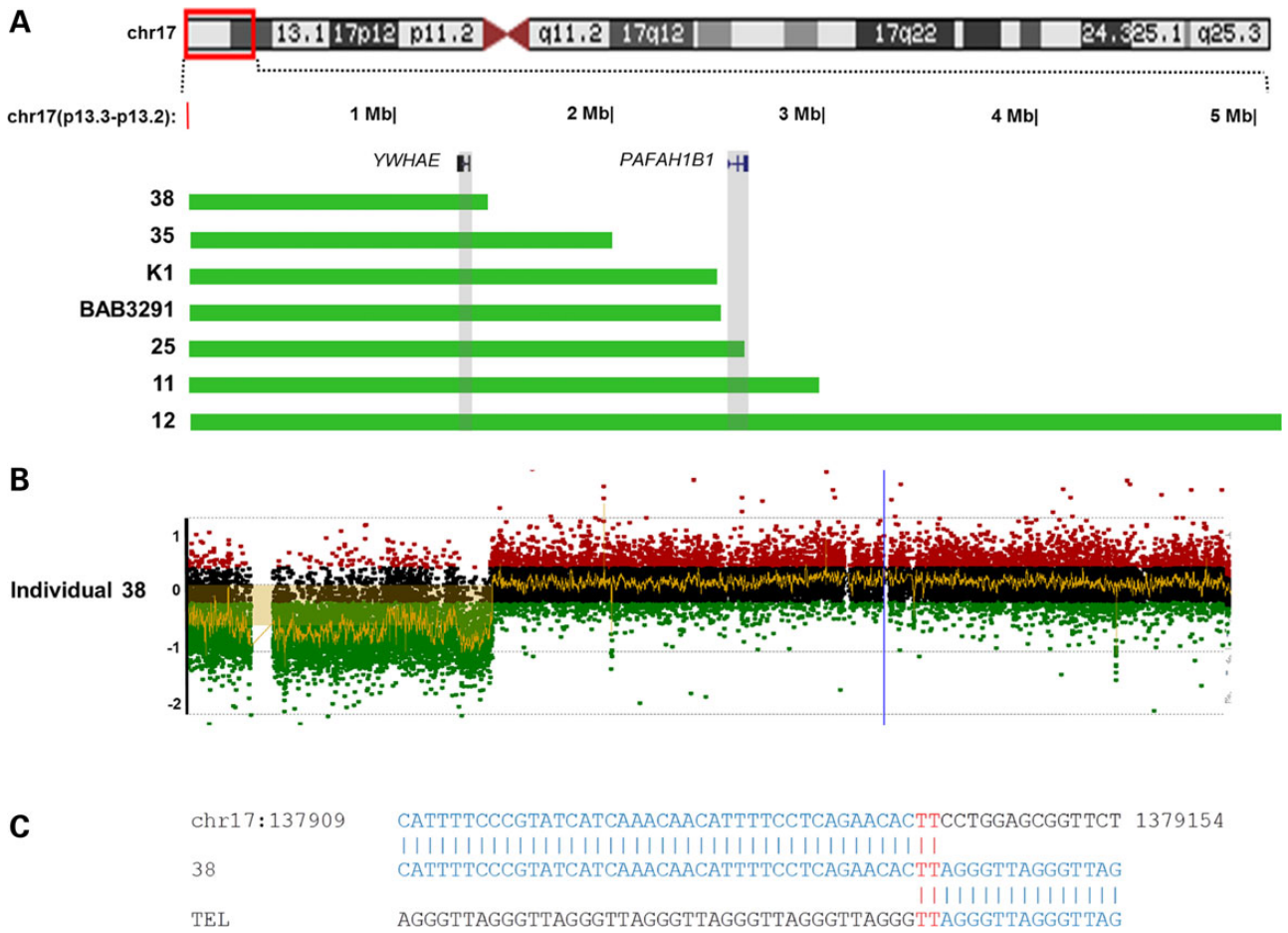


Figure 6. Terminal 17p deletion cases. (**A**) CNVs were plotted by exact genomic coordinates according to breakpoint junction sequencing results using the UCSC genome browser custom track function (except for individual 25 without breakpoint junctions mapped to nucleotide resolution, which were plotted according to aCGH result, with resolution higher than 200 bp per probe). Green blocks represent deletion. Upper panel shows an ideogram of human chromosome 17 with the 17p13.2–17p13.3 region involved highlighted in the red box. Two critical genes in 17p13.3 deletion syndromes, *YWHAE* and *PAFAH1B1*, are highlighted in gray boxes. (**B**, **C**) Representative array CGH $\log_2$ plot and breakpoint junction sequence of individual 38 with terminal deletion. Array plots and breakpoint sequences of additional cases with terminal deletion are shown in Supplementary Material, Figure S7.

Notably, the *Alu* elements that mediated the 17p13.3 CNVs more frequently utilized different families of *Alu* elements, within a given *Alu*–*Alu* substrate pair (13 out of 20 [65%]). The recombination join point or breakpoint junction often occurred in a region of higher identity, although some breakpoint junctions (e.g. K2_1) localize to homeologous regions with a number of mismatches.

Discussion

Although the first discovery of the connection between MDS and 17p13.3 deletion occurred three decades ago (38), the mechanisms responsible for the generation of the CNVs within the 17p13.3 region have remained poorly understood. Much of the literature regarding CNVs in this region is focused on genotype–phenotype correlations. Studies that attempted to explore mechanisms for CNV formation have mainly focused on deletions (10,26), whereas investigations of duplications and CGRs have rarely been reported (27).

Here, we gain mechanistic insights by exploring the patterns of copy-number changes, as evidenced by deviations from the normal diploid state and determining the breakpoint junction sequences in 39 individuals, 5 with interstitial deletion, 14 with tandem duplication, 7 with terminal deletion and 13 with CGRs. Among the 13 CGRs, we observed diverse patterns of copy-number changes: four DUP–NML–DUP structures, six CGRs with triplicated segments (of which four were flanked by duplications), one inverted duplication/translocation, one DEL–NML–DEL and one with an extremely complex 'DEL–NML–DUP–TRP–DUP–TRP–DUP–NML–DUP–TRP–DUP' pattern. In all but three individuals, we mapped the breakpoint junctions to nucleotide resolution by PCR and Sanger sequencing. The information obtained from these breakpoint junction sequences allowed us to surmise potential rearrangement mechanisms involved in 17p13.3.

Subtelomeric regions exhibit increased genome instability and are associated with various pathogenic CNVs that can cause autism, intellectual disability, neurocognitive deficits and multiple congenital anomalies. Subtelomeric rearrangements have been identified on every chromosome (39,40), including some regions with well-described genotype–phenotype correlations and mutational mechanisms, most notably, deletions on 1p (MIM #607872) (41,42), 9q (MIM #158170) (43,44) and 22q (MIM #606232) (45). Subtelomeric CNVs are typically non-recurrent and have unique breakpoints that are generated through mechanisms other than NAHR (44,46). The 17p13.3 CNVs analyzed

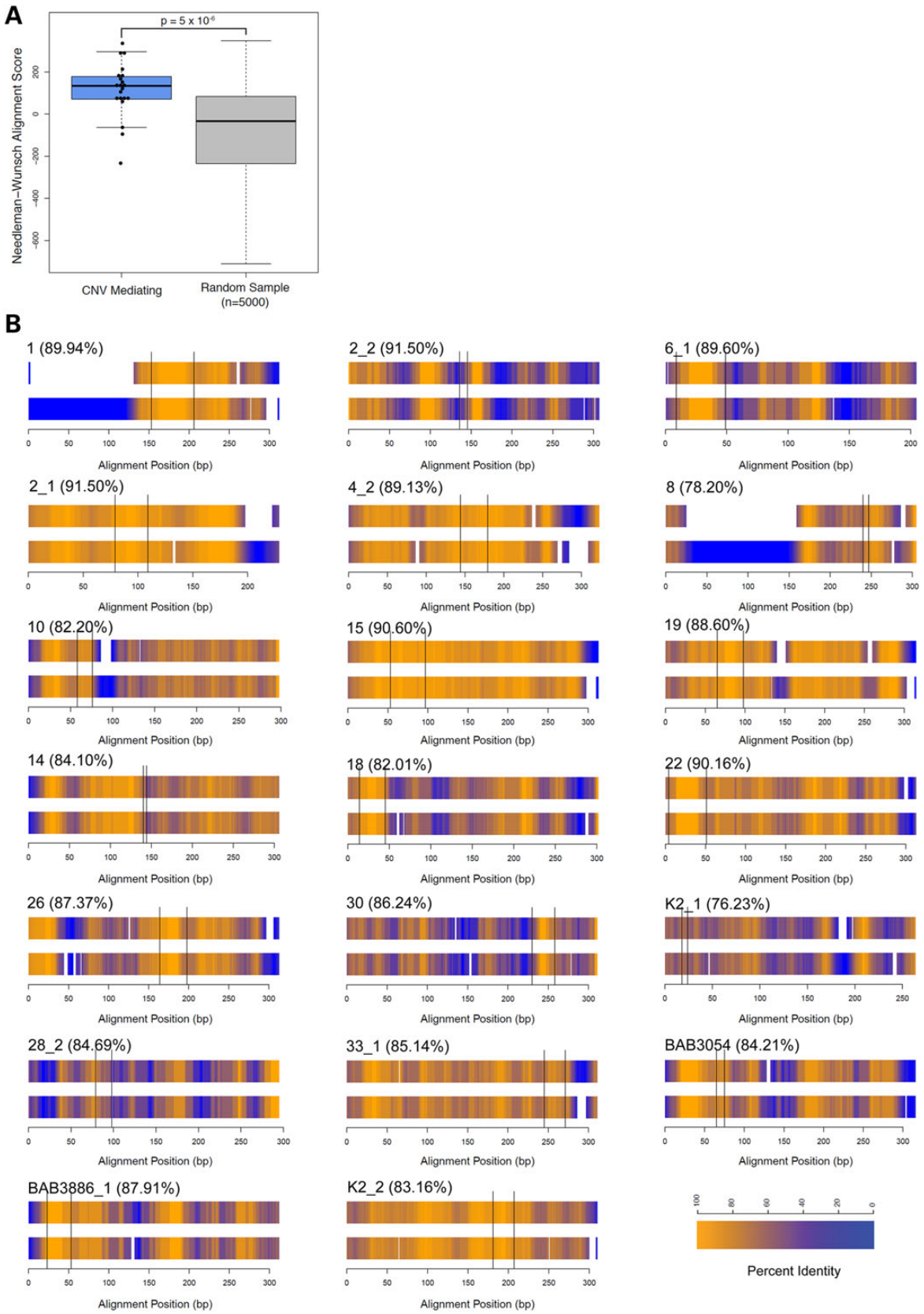


Figure 7. Analysis of *Alu*–*Alu*-mediated breakpoints. (**A**) Smith–Waterman alignment scores of *Alu* pairs that mediate CNVs are significantly higher than randomly selected pairs in the region ($P = 5 \times 10^{-6}$, Wilcoxon signed rank test). (**B**) *Alu* pairs aligned for each *Alu*-mediated breakpoint. The telomeric *Alu* is on top in each pairing. Heat map shading depicts a 20 bp sliding window of identity. Microhomology/crossover spots are flanked by black lines, which occupy various positions within the *Alu* pairs. The percent identity for each aligned pair, excluding gaps, is also presented.

in this study reproduced these characteristics—the CNVs varied in size and none of them shared identical breakpoint junctions and exhibited the hallmarks of various mutational mechanisms (Table 1). Due to the non-recurrent nature and variable size of 17p13.3 CNVs, it is understandable that diverse phenotypes have been observed, as different rearrangements involve different dosage-sensitive genes (26,27,47).

Unlike some subtelomeric genomic regions that have abundant segmental duplications (48), the 17p13.3 region contains few of these repeats, but rather consists of gene-rich sequences. Although lacking long repeat sequences (low-copy repeats or LCRs), 17p13.3 is highly enriched for *Alu* repetitive elements. This region contains ~30% *Alu* sequences in the human genome reference (hg19/GRCh37), compared with ~18% for chromosome 17 and ~10% at the genome-wide scale. This places the 17p13.3 region in the 98.8th percentile of *Alu* repeats in a single region when compared with equally-sized regions genome-wide.

The abundance of *Alu* repetitive elements provides a potential substrate for the generation of CNVs. Among the breakpoint junctions mapped to nucleotide resolution within 17p13.3, breakpoint junctions in 80% of the interstitial deletions (4 out of 5), 46% of the tandem duplications (6 out of 13) and 50% of the CGRs (10 out of 20) were mediated by *Alu*–*Alu* rearrangement. In total, 47% of the 17p13.3 breakpoint junctions mapped to nucleotide resolution appeared to be *Alu*–*Alu*-mediated (Table 2). In addition, at the 17p13.3 region, 30 of 35 (86%) subjects (36 of 43 sequenced breakpoints, 84%) had at least one *Alu* or other transposable element involved in facilitating rearrangement formation (Table 1). Previous studies of randomly selected non-recurrent CNVs identified in subjects referred for clinical diagnostic testing but not ascertained for a particular phenotype have shown *Alu*–*Alu*-mediated rearrangements in ~7–29.2% of CNVs (49,50). Thus, rearrangements of the 17p13.3 region appear to be significantly enriched in *Alu*–*Alu*-mediated CNVs (Fisher's exact test, $P < 0.005$). Therefore, we speculate that the 17p13.3 region favors *Alu*–*Alu*-mediated rearrangement mechanisms, in turn leading to non-recurrent CNVs resulting in phenotypically diverse syndromes.

Breakpoint junctions in simple deletions and duplications that occurred without the formation of a new *Alu* hybrid but showed 2–6 bp microhomology (in BAB3106 with an interstitial deletion and in individuals 29, 3, 9, 36, 7, 31 and 21 with tandem duplications) may be generated through various mechanisms (Table 2). Non-homologous end joining (NHEJ) (51,52) or microhomology-mediated end joining (MMEJ) (53,54) could potentially explain the simple deletion in BAB3106. Although NHEJ has been proposed to generate duplications (55), due to a gain of genomic material, MMEJ or replicative mechanisms, e.g. FoSTeS/MMBIR, may more readily explain the increased copy number that can accompany a replicative process (23,24). Other mechanisms with the capacity to generate deletions or duplications, e.g. single-strand annealing (SSA) and synthesis-dependent strand annealing, require long uninterrupted repeat sequences for strand invasion, which was absent from the breakpoint junctions we observed in this study (56).

The molecular mechanism of *Alu*–*Alu*-mediated rearrangement has long been proposed to be due to NAHR (9,57). This was partially because simple deletions and duplications were thought to be much more frequently observed with *Alu*–*Alu* events, and complex rearrangements were rarely described (7). However, in this study, we observed that 50% of the breakpoints of CGRs (10 out of 20) within 17p13.3 were *Alu*–*Alu*-mediated events. Moreover, the orientation of the *Alu* pairs—whether directly oriented or in opposite-orientation—determined the overall rearrangement pattern when chimeric *Alu*s were generated. For example, although all individuals 2, 4, 6 and BAB3886 exhibited a 'DUP–NML–DUP' pattern of rearrangement from the aCGH results, DNA from individual 2 had both breakpoint junctions mediated by directly oriented *Alu* pairs, therefore resulting in duplications without inversion. In contrast, in individuals 4, 6 and BAB3886, one of the two breakpoint junctions was mediated by *Alu* pairs in opposite orientation, resulting in one inverted duplication (Fig. 3 and Supplementary Material, Fig. S3). Similar observations of CNVs with 'DUP–NML–INV/DUP' pattern mediated by *Alu*–*Alu* rearrangement have been previously reported in other genomic regions (19,32).

Rearrangement patterns consistent with *Alu* pair orientations were also observed in individuals with triplication. Both breakpoint junctions of the triplication in individual K2 were mediated by directly oriented *Alu* pairs, whereas in individuals 28 and 33, one of the breakpoint junctions was mediated by *Alu* pairs in opposite orientation. As a result, the triplication in the former was not inverted, whereas the latter two had an inverted triplicated region (Fig. 4 and Supplementary Material, Fig. S5). The

Table 2. Mutational signatures of the 17p13.3 breakpoints for CNVs characterized in this study[a]

Signature	CGR	Simple DEL or DUP	Percentage
Alu–*Alu*	10[b]	10[c]	46.5% (20 out of 43)
FoSTeS/MMBIR	10[d]	8[e]	41.9% (18 out of 43)[f]
NHEJ or MMEJ	0		
Telomeric healing	N/A	5[g]	11.6% (5 out of 43)
Total	20 (46.5%)	23 (53.5%)	100% (43 out of 43)

Abbreviations: DEL, deletion; DUP, duplication; CGR, complex genomic rearrangement; *Alu*–*Alu*, *Alu*–*Alu*-mediated rearrangement; MMBIR, microhomology-mediated break-induced replication; FoSTeS, fork stalling and template switching; NHEJ, non-homologous end joining; MMEJ, microhomology-mediated end joining; N/A, not applicable.

[a]Only breakpoint junctions mapped within the 17p13.3 region and at nucleotide resolution are shown (e.g. breakpoint junctions in individuals 5, 12, 23 and 24 are not considered due to either or both proximal and distal sequences of their breakpoints mapping outside the 17p13.3 region).

[b]These 10 breakpoint junctions include: 2_1, 2_2, 4_2, 6_1, BAB3886_1, 28_2, 33_1, K2_1, K2_2 and 18 as shown in Table 1.

[c]These 10 breakpoint junctions include: 10, 15, 19, 30, 1, 8, 14, 22, 26 and BAB3054 as shown in Table 1.

[d]These 10 breakpoint junctions include: 4_1, 6_2, BAB3886_2, 27_1, 27_2, 28_1, 33_2, 20, 17_1 and 17_2 as shown in Table 1.

[e]These eight breakpoint junctions include: BAB3106, 3, 7, 9, 21, 29, 31 and 36 as shown in Table 1. These eight breakpoint junctions in simple deletions or duplications may be generated through FoSTeS/MMBIR, NHEJ or MMEJ. Refer to the 'Results' section for detailed descriptions of each breakpoint.

[f]Percentage of these types of mutational signatures, FoSTeS/MMBIR, NHEJ and MMEJ are calculated together.

[g]These five breakpoint junctions include: 11, 35, 38, BAB3291 and K1 as shown in Table 1.

'DUP-TRP/INV-DUP' CNV pattern was originally described at Xq28 (*MECP2*), and these rearrangements were found to be mostly mediated by inverted long repeats (856 bp–11.3 kb) with at least 98% sequence identity (33). Here, we showed that inverted small repetitive elements like *Alu*s (~300 bp in length) with ~85% sequence identity could also generate this type of complex rearrangement with inverted triplication embedded in a duplication (Table 1 and Fig. 7). Additionally, for CGR in individual 27, neither of the breakpoint junctions appeared to be mediated by repetitive elements (Table 1), indicating that the absence of inverted repeats can also generate this type of DUP-TRP/INV-DUP rearrangement product structure.

Recent studies of rearrangement breakpoint junctions determined in samples from subjects with spastic paraplegia 4 (SPG4, MIM #182601) caused by pathogenic deletions or duplications of the *SPAST* gene demonstrated that 70% (38 out of 54) of CNVs were mediated by directly oriented *Alu* pairs (15,16). The *Alu* pairs that mediate the 17p13.3 CNVs in this study have similar characteristics with the *Alu* pairs that mediate CNVs involving *SPAST*. The average percent identity for each *Alu* pair in our study is 86.0% (range 76.2–91.5%; Fig. 7) compared with 82.7% (range 75.8–90.7%) involving *SPAST*. Notably, the *Alu* elements that mediated the CNVs in both scenarios more frequently utilized different *Alu* families within a pair, a finding inconsistent with an NAHR-mediated event (33 out of 39, 85% in the SPG4 study versus 13 out of 20, [65%] in this study) (Table 1). The average *Alu*–*Alu*-mediated CNV breakpoint microhomology also had a wider range in our study (mean 26 bp, range 3–52 bp, Table 1) compared with that evaluating CNVs in *SPAST* (mean 17 bp, range 5–38 bp).

Several mechanisms were proposed to potentially generate *Alu*-mediated CNVs other than NAHR in the SPG4 study, including MMEJ, SSA, FoSTeS/MMBIR and homeologous recombination. These newly proposed mechanisms were based primarily on the observation that: (i) many of the simple deletions and duplications were *Alu*–*Alu* events, but the overwhelming majority formed a recombinant hybrid from substrate constituents representing different *Alu* family members and (ii) the percent identity of the involved *Alu* was <90%, inconsistent with an homologous recombination process, but rather consistent with an homeologous and/or microhomology-mediated process (16). In addition, a previous smaller study at the *SPAST* locus, which investigated the mechanism for deletion of the same exon in multiple unrelated subjects, revealed: (i) different-sized genomic deletions underlying the same exonic loss, (ii) the utilization of different *Alu* sequence pairs as substrates for the recombinant/hybrid *Alu* formed and (iii) evidence for potential multiple template switches (15). All the *Alu*-mediated CNVs in the larger SPG4 study appeared to be simple deletions or duplications generated through directly oriented *Alu* pairs. The complex nature of the *Alu*-mediated CGRs with multiple template switches, as we observed in the current study, especially mediated by opposite-oriented *Alu* pairs, are more consistent with the features we observe in replicative mechanisms. In addition, even apparently simple deletions with breakpoint junctions revealing evidence for *Alu*–*Alu*-mediated events can potentially arise from multiple iterative *Alu*–*Alu* template switches (15).

One such mechanism is classical break-induced replication (BIR). BIR is initiated by invasion of a single-ended double-strand break into homologous sequence followed by extensive DNA synthesis. BIR has been proposed to be one of the mechanisms responsible for generating non-recurrent CNVs in humans (33,58). Notably, BIR is dependent on the strand invasion and homology search protein Rad51 (59). Previous studies in humans suggested that the minimal efficient processing segment substrates of Rad51 were 300–500 bp with uninterrupted homology (60), which would preclude the use of *Alu* elements. More recent studies of Rad51^{-} or apparently Rad51-dependent processes, however, questioned these strict limits (61,62). Thus, Rad51 or another of its paralogs in humans (63) may indeed potentially mediate BIR using *Alu* elements as substrates, although less efficiently, also explaining the rarity of identical *Alu*–*Alu*-mediated CNVs. Alternatively, in the MMBIR mechanism, which is proposed to be independent of Rad51, single-stranded DNA could potentially switch to any template with 2–15 bp of shared microhomology (23,64). Whether homeology (such as provided by the *Alu* elements) helps guide strand invasion during template switching in recombination coupled DNA synthesis is unclear at this time.

Recent studies in the model organism *Saccharomyces cerevisiae* further support the above hypothesis. Ty elements in yeast contain a large central region flanked by two long terminal repeats (LTRs, or δ elements) (65). These LTRs, either within a Ty element or present as solo LTRs as the product of excision of the central region of Ty elements (66), are analogous to *Alu* elements in humans, which are present in many copies in the yeast genome and share high sequence homology (67). Ty-mediated BIR followed by DNA double-strand break (DSB) leading to deletions, duplications and translocations are well documented and studied (67–69). By inducing low levels of replicative DNA polymerase and therefore introducing more deletions and duplications in yeast, it was observed that most of the breakpoints of these CNVs were located at Ty elements (70). In budding yeast, it was recently demonstrated that Ty-mediated template switching during BIR/MMBIR was a prominent mechanism causing complex chromosome rearrangements (71). Ty and δ element-mediated multiple template switches resulted in CGRs with multiple copy-number transitions, which was similar to the *Alu*-mediated CGRs we observed in humans in this study (71).

Overall, our results show that *Alu*–*Alu*-mediated rearrangement is a common cause of the non-recurrent CNVs of chromosome 17 at p13.3 that mediate intricate phenotypes causing genomic disorders in humans. Our data suggest that the high concentration of *Alu* elements in the region serve as a substrate with increased homology/homeology for formation of CNVs. Moreover, we provide evidence that *Alu*–*Alu*-mediated mechanisms contribute considerably to the formation of complex CNVs and that the orientation of the *Alu* pair may determine whether an inversion of a genomic segment is involved, and these mechanisms are unlikely attributed to NAHR. Additional studies of genome-wide *Alu* elements may identify additional loci that are particularly susceptible to such structural variations.

Materials and Methods

Subjects

Individuals with CNVs in the 17p13.3 region were identified by chromosomal microarray analysis (CMA) clinical testing at Baylor College of Medicine (BCM) Medical Genetics Laboratories. Of the 33 800 individuals tested from 27 May 2008 to 23 December 2013, a total of 39 unrelated subjects were identified with copy-number changes involving known disease genes including *PAFAH1B1*, *YWHAE* and *bHLHA9*, or disease candidate gene *RPA1* within the 17p13.3 region. Individuals with interstitial deletions involving *PAFAH1B1*, but no other disease genes, those with unbalanced translocations or likely benign copy-number gains at the one end of a disease gene, were not included in this study.

This study was approved by the Institutional Review Board for Human Subject Research at BCM (IRB No. H-22231 or H-25466). Informed consent was obtained prior to collecting identifiable DNA samples (BAB3054, BAB3106, BAB3291, BAB3886, BAB3887, BAB3888, K1 and K2). The remaining DNA samples were de-identified for breakpoint and mechanistic studies (named individual 1, individual 2, individual 3, etc.).

Clinical CMA

Custom-designed BCM OLIGO V6.5, V7, V8 or V9 oligonucleotide arrays were performed as previously described (72,73). Arrays were designed to specifically interrogate clinically significant regions with an average resolution of 30 kb between probes.

High-density aCGH

To further characterize the CNVs identified by CMA in the 17p13.3 region, two custom, high-density Agilent arrays were designed. A 4 × 180 K array with ~200 bp per probe coverage spanning 17pter to 17p13.2 and ~2000 bp per probe coverage spanning the rest of chromosome 17 was used to study individuals 5, 23 and 24. For the remaining individuals, an 8 × 60 K array with ~150 bp per probe covering the first 6 Mb of chromosome 17p (17pter to 17p13.2) and ~5000 bp per probe covering the remaining chromosome 17 was used. Hybridization controls were gender-matched (HapMap individual NA10851 as male control and HapMap individual NA15510 as female control). For the anonymous individuals with unknown gender, the male control DNA (HapMap individual NA10851) was used. Scanned array images were processed using Agilent Feature Extraction software (version 10) and extracted files were analyzed using Agilent Genomic Workbench (version 7.0.4.0). Array designs and sequence alignment for breakpoint analysis were based on the February 2009 genome build (GRCh37/hg19 assembly).

Breakpoint junction mapping

To further confirm the CNVs identified by high-density arrays and map the breakpoint junctions, primers flanking the predicted breakpoints were designed, and long-range PCRs were conducted using TaKaRa LA *Taq* according to the manufacturer's protocol (TaKaRa Bio Company, Cat. No. RR002). Generally, 200 ng of genomic DNA and 0.5 μl of each primer (10 μM) were put into a 25 μl scale PCR reaction mixture containing 2.5 μl 10× LA *Taq* buffer, 4 μl dNTP mixture (2.5 mM each) and 0.25 μl LA *Taq* enzyme. Annealing temperature was set at 68°C and extension time was set for 1 min per kb. PCR products were prepared for sequencing using ExoSAP-IT (Affymetrix, Cat. No. 78201) according to the manufacturer's protocol or gel extracted and purified with the Zymoclean Gel DNA Recovery Kit (Zymo Research, Cat. No. D4001). Purified PCR products were then sequenced by Sanger dideoxynucleotide sequencing (BCM Sequencing Core, Houston, TX, USA).

Alu pair analysis

We obtained the annotations for each *Alu* element in the 17p13.3 region from the RepeatMasker track of the UCSC genome browser (74). We obtained the sequence encoding each *Alu* (the + or – reference sequence depending on the orientation of the element) using the element annotation and the BSgenome package implemented in the R Statistical Programing Language. We subsequently aligned the *Alu* sequences to each other using the R Biostrings package and computed the identity as the fraction of identical positions over the total number of aligned positions excluding gaps.

Statistical analysis

To assess how the 17p13.3 region's *Alu* content compares with the rest of the genome, we adopted a Monte Carlo approach. We randomly selected windows across the genome of the same size as the 17p13.3 region. We rejected windows that overlapped any gaps in the hg19 assembly using the UCSC gaps track, including centromeres and telomeres. We subsequently determined fractional *Alu* content using the RepeatMasker track. This process was repeated 10 000 times. We next determined the empirical cumulative distribution function of the simulated data and evaluated it at the fractional content of 17p13.3. To assess enrichment in *Alu*–*Alu*-mediated CNVs at 17p13.3, we assumed a null hypothesis that the CNVs from previous studies (49,50,75) and those we detected in the current study were sampled from the same distribution. We performed a Fisher's exact test with the pooled data from the previous studies versus our observations.

Authors' Contributions

S.G., W.B. and J.R.L. designed the study. S.G. conducted the experiments. S.G., B.Y., C.M.B.C. and J.R.L. analyzed the data. I.M.C. performed the bioinformatics analysis. S.G. and J.R.L. wrote the manuscript. B.Y., I.M.C., C.R.B., C.M.B.C., S.C.S.N. and C.A.B. revised the manuscript. A.P., C.A.B., S.C.S.N., A.E., C.A.S., P.S., S.W.C. and W.B. identified the cases through clinical diagnosis. All authors read and approved the final manuscript.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Conflict of Interest statement. J.R.L. has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, has stock options in Lasergen, Inc. and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases and bacterial genomic fingerprinting. The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from the chromosomal microarray analysis (CMA) and clinical exome sequencing offered in the Baylor Miraca Genetics Laboratory (BMGL: http://www.bmgl.com/BMGL/Default.aspx).

Funding

We thank the patients and their families for participating in this study. I.M.C. is a fellow of the Baylor College of Medicine Medical Scientist Training Program (T32GM007330). He was supported by a fellowship from the National Institute of Neurological Disorders and Stroke (F31 NS083159). C.R.B. is an HHMI Fellow of the Damon Runyon Cancer Research Foundation (DRG 2155-13). S.C.S.N. is a recipient of the DDCF Clinical Scientist Development Award. This work was supported in part by grants from the US National Institute of Neurological Disorders and Stroke (R01NS058529),

National Institute of General Medical Sciences (R01GM106373) and National Human Genome Research Institute (NHGRI)/ National Heart Lung and Blood Institute (NHLBI) (U54 HG006542) to the Baylor-Hopkins Center for Mendelian Genomics. This work was also supported by the Doris Duke Charitable Foundation (DDCF) (Grant #2013095). The project described was supported by Baylor College of Medicine IDDRC from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (Grant Number 1 U54 HD083092).

References

1. Batzer, M.A. and Deininger, P.L. (2002) *Alu* repeats and human genomic diversity. *Nat. Rev. Genet.*, **3**, 370–379.
2. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
3. Deininger, P. (2011) *Alu* elements: know the SINEs. *Genome Biol.*, **12**, 236.
4. Witherspoon, D.J., Zhang, Y., Xing, J., Watkins, W.S., Ha, H., Batzer, M.A. and Jorde, L.B. (2013) Mobile element scanning (ME-Scan) identifies thousands of novel *Alu* insertions in diverse human populations. *Genome Res.*, **23**, 1170–1181.
5. Shen, M.R., Batzer, M.A. and Deininger, P.L. (1991) Evolution of the master *Alu* gene(s). *J. Mol. Evol.*, **33**, 311–320.
6. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–876.
7. Deininger, P.L. and Batzer, M.A. (1999) *Alu* repeats and human disease. *Mol. Genet. Metab.*, **67**, 183–193.
8. Brooks, E.M., Branda, R.F., Nicklas, J.A. and O'Neill, J.P. (2001) Molecular description of three macro-deletions and an *Alu–Alu* recombination-mediated duplication in the HPRT gene in four patients with Lesch–Nyhan disease. *Mutat. Res.*, **476**, 43–54.
9. Sen, S.K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P.A., Dyer, M., Cordaux, R., Liang, P. and Batzer, M.A. (2006) Human genomic deletions mediated by recombination between *Alu* elements. *Am. J. Hum. Genet.*, **79**, 41–53.
10. Mei, D., Lewis, R., Parrini, E., Lazarou, L.P., Marini, C., Pilz, D.T. and Guerrini, R. (2008) High frequency of genomic deletions—and a duplication—in the LIS1 gene in lissencephaly: implications for molecular diagnosis. *J. Med. Genet.*, **45**, 355–361.
11. Franke, G., Bausch, B., Hoffmann, M.M., Cybulla, M., Wilhelm, C., Kohlhase, J., Scherer, G. and Neumann, H.P. (2009) *Alu–Alu* recombination underlies the vast majority of large VHL germline deletions: molecular characterization and genotype–phenotype correlations in VHL patients. *Hum. Mutat.*, **30**, 776–786.
12. Shlien, A., Baskin, B., Achatz, M.I., Stavropoulos, D.J., Nichols, K.E., Hudgins, L., Morel, C.F., Adam, M.P., Zhukova, N., Rotin, L. *et al.* (2010) A common molecular mechanism underlies two phenotypically distinct 17p13.1 microdeletion syndromes. *Am. J. Hum. Genet.*, **87**, 631–642.
13. Strout, M.P., Marcucci, G., Bloomfield, C.D. and Caligiuri, M.A. (1998) The partial tandem duplication of ALL1 (MLL) is consistently generated by *Alu*-mediated homologous recombination in acute myeloid leukemia. *Proc. Natl Acad. Sci. USA*, **95**, 2390–2395.
14. O'Neil, J., Tchinda, J., Gutierrez, A., Moreau, L., Maser, R.S., Wong, K.K., Li, W., McKenna, K., Liu, X.S., Feng, B. *et al.* (2007) *Alu* elements mediate MYB gene tandem duplication in human T-ALL. *J. Exp. Med.*, **204**, 3059–3066.
15. Boone, P.M., Liu, P., Zhang, F., Carvalho, C.M., Towne, C.F., Batish, S.D. and Lupski, J.R. (2011) *Alu*-specific microhomology-mediated deletion of the final exon of SPAST in three unrelated subjects with hereditary spastic paraplegia. *Genet. Med.*, **13**, 582–592.
16. Boone, P.M., Yuan, B., Campbell, I.M., Scull, J.C., Withers, M.A., Baggett, B.C., Beck, C.R., Shaw, C.J., Stankiewicz, P., Moretti, P. *et al.* (2014) The *Alu*-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am. J. Hum. Genet.*, **95**, 143–161.
17. Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D. and Lupski, J.R. (2009) The DNA replication FoSTeS/ MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.*, **41**, 849–853.
18. Zhang, F., Seeman, P., Liu, P., Weterman, M.A., Gonzaga-Jauregui, C., Towne, C.F., Batish, S.D., De Vriendt, E., De Jonghe, P., Rautenstrauss, B. *et al.* (2010) Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am. J. Hum. Genet.*, **86**, 892–903.
19. Szafranski, P., Golla, S., Jin, W., Fang, P., Hixson, P., Matalon, R., Kinney, D., Bock, H.G., Craigen, W., Smith, J.L. *et al.* (2014) Neurodevelopmental and neurobehavioral characteristics in males and females with CDKL5 duplications. *Eur. J. Hum. Genet.*, doi: 10.1038/ejhg.2014.217.
20. Carvalho, C.M., Pehlivan, D., Ramocki, M.B., Fang, P., Alleva, B., Franco, L.M., Belmont, J.W., Hastings, P.J. and Lupski, J.R. (2013) Replicative mechanisms for CNV formation are error prone. *Nat. Genet.*, **45**, 1319–1326.
21. Hedges, D.J. and Deininger, P.L. (2007) Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.*, **616**, 46–59.
22. Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K. and Eichler, E.E. (2010) A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell*, **143**, 837–847.
23. Hastings, P.J., Ira, G. and Lupski, J.R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.*, **5**, e1000327.
24. Lee, J.A., Carvalho, C.M. and Lupski, J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, **131**, 1235–1247.
25. Bruno, D.L., Anderlid, B.M., Lindstrand, A., van Ravenswaaij-Arts, C., Ganesamoorthy, D., Lundin, J., Martin, C.L., Douglas, J., Nowak, C., Adam, M.P. *et al.* (2010) Further molecular and clinical delineation of co-locating 17p13.3 microdeletions and microduplications that show distinctive phenotypes. *J. Med. Genet.*, **47**, 299–311.
26. Nagamani, S.C., Zhang, F., Shchelochkov, O.A., Bi, W., Ou, Z., Scaglia, F., Probst, F.J., Shinawi, M., Eng, C., Hunter, J.V. *et al.* (2009) Microdeletions including YWHAE in the Miller–Dieker syndrome region on chromosome 17p13.3 result in facial dysmorphisms, growth restriction, and cognitive impairment. *J. Med. Genet.*, **46**, 825–833.
27. Bi, W., Sapir, T., Shchelochkov, O.A., Zhang, F., Withers, M.A., Hunter, J.V., Levy, T., Shinder, V., Peiffer, D.A., Gunderson, K.L. *et al.* (2009) Increased LIS1 expression affects human and mouse brain development. *Nat. Genet.*, **41**, 168–177.
28. Luk, H.M., Wong, V.C., Lo, I.F., Chan, K.Y., Lau, E.T., Kan, A.S., Tang, M.H., Tang, W.F., She, W.M., Chu, Y.W. *et al.* (2014) A

prenatal case of split-hand malformation associated with 17p13.3 triplication—a dilemma in genetic counseling. *Eur. J. Med. Genet.*, **57**, 81–84.
29. Outwin, E., Carpenter, G., Bi, W., Withers, M.A., Lupski, J.R. and O'Driscoll, M. (2011) Increased RPA1 gene dosage affects genomic stability potentially contributing to 17p13.3 duplication syndrome. *PLoS Genet.*, **7**, e1002247.
30. Han, K., Lee, J., Meyer, T.J., Remedios, P., Goodwin, L. and Batzer, M.A. (2008) L1 recombination-associated deletions generate human genomic variation. *Proc. Natl Acad. Sci. USA*, **105**, 19366–19371.
31. Carvalho, C.M., Bartnik, M., Pehlivan, D., Fang, P., Shen, J. and Lupski, J.R. (2012) Evidence for disease penetrance relating to CNV size: Pelizaeus–Merzbacher disease and manifesting carriers with a familial 11 Mb duplication at Xq22. *Clin. Genet.*, **81**, 532–541.
32. Liu, P., Erez, A., Nagamani, S.C., Dhar, S.U., Kolodziejska, K.E., Dharmadhikari, A.V., Cooper, M.L., Wiszniewska, J., Zhang, F., Withers, M.A. *et al.* (2011) Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. *Cell*, **146**, 889–903.
33. Carvalho, C.M., Ramocki, M.B., Pehlivan, D., Franco, L.M., Gonzaga-Jauregui, C., Fang, P., McCall, A., Pivnick, E.K., Hines-Dowell, S., Seaver, L.H. *et al.* (2011) Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.*, **43**, 1074–1081.
34. Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A. *et al.* (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, **144**, 27–40.
35. Maher, C.A. and Wilson, R.K. (2012) Chromothripsis and human disease: piecing together the shattering process. *Cell*, **148**, 29–32.
36. Yokoyama, Y., Narahara, K., Teraoka, M., Koyama, K., Seino, Y., Yagi, S., Konishi, T. and Miyawaki, T. (1997) Cryptic pericentric inversion of chromosome 17 detected by fluorescence *in situ* hybridization study in familial Miller–Dieker syndrome. *Am. J. Med. Genet.*, **71**, 236–237.
37. Flint, J., Craddock, C.F., Villegas, A., Bentley, D.P., Williams, H.J., Galanello, R., Cao, A., Wood, W.G., Ayyub, H. and Higgs, D.R. (1994) Healing of broken human chromosomes by the addition of telomeric repeats. *Am. J. Hum. Genet.*, **55**, 505–512.
38. Stratton, R.F., Dobyns, W.B., Airhart, S.D. and Ledbetter, D.H. (1984) New chromosomal syndrome: Miller–Dieker syndrome and monosomy 17p13. *Hum. Genet.*, **67**, 193–200.
39. Shao, L., Shaw, C.A., Lu, X.Y., Sahoo, T., Bacino, C.A., Lalani, S.R., Stankiewicz, P., Yatsenko, S.A., Li, Y., Neill, S. *et al.* (2008) Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: a study of 5380 cases. *Am. J. Med. Genet. A*, **146A**, 2242–2251.
40. Ballif, B.C., Sulpizio, S.G., Lloyd, R.M., Minier, S.L., Theisen, A., Bejjani, B.A. and Shaffer, L.G. (2007) The clinical utility of enhanced subtelomeric coverage in array CGH. *Am. J. Med. Genet. A*, **143A**, 1850–1857.
41. Ballif, B.C., Gajecka, M. and Shaffer, L.G. (2004) Monosomy 1p36 breakpoints indicate repetitive DNA sequence elements may be involved in generating and/or stabilizing some terminal deletions. *Chromosome Res.*, **12**, 133–141.
42. Ballif, B.C., Yu, W., Shaw, C.A., Kashork, C.D. and Shaffer, L.G. (2003) Monosomy 1p36 breakpoint junctions suggest premeiotic breakage-fusion-bridge cycles are involved in generating terminal deletions. *Hum. Mol. Genet.*, **12**, 2153–2165.
43. Yatsenko, S.A., Brundage, E.K., Roney, E.K., Cheung, S.W., Chinault, A.C. and Lupski, J.R. (2009) Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome. *Hum. Mol. Genet.*, **18**, 1924–1936.
44. Yatsenko, S.A., Hixson, P., Roney, E.K., Scott, D.A., Schaaf, C.P., Ng, Y.T., Palmer, R., Fisher, R.B., Patel, A., Cheung, S.W. *et al.* (2012) Human subtelomeric copy number gains suggest a DNA replication mechanism for formation: beyond breakage-fusion-bridge for telomere stabilization. *Hum. Genet.*, **131**, 1895–1910.
45. Bonaglia, M.C., Giorda, R., Beri, S., De Agostini, C., Novara, F., Fichera, M., Grillo, L., Galesi, O., Vetro, A., Ciccone, R. *et al.* (2011) Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genet.*, **7**, e1002173.
46. Luo, Y., Hermetz, K.E., Jackson, J.M., Mulle, J.G., Dodd, A., Tsuchiya, K.D., Ballif, B.C., Shaffer, L.G., Cody, J.D., Ledbetter, D.H. *et al.* (2011) Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. *Hum. Mol. Genet.*, **20**, 3769–3778.
47. Armour, C.M., Bulman, D.E., Jarinova, O., Rogers, R.C., Clarkson, K.B., DuPont, B.R., Dwivedi, A., Bartel, F.O., McDonell, L., Schwartz, C.E. *et al.* (2011) 17p13.3 microduplications are associated with split-hand/foot malformation and long-bone deficiency (SHFLD). *Eur. J. Hum. Genet.*, **19**, 1144–1151.
48. Linardopoulou, E.V., Williams, E.M., Fan, Y., Friedman, C., Young, J.M. and Trask, B.J. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature*, **437**, 94–100.
49. Vissers, L.E., Bhatt, S.S., Janssen, I.M., Xia, Z., Lalani, S.R., Pfundt, R., Derwinska, K., de Vries, B.B., Gilissen, C., Hoischen, A. *et al.* (2009) Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum. Mol. Genet.*, **18**, 3579–3593.
50. Verdin, H., D'Haene, B., Beysen, D., Novikova, Y., Menten, B., Sante, T., Lapunzina, P., Nevado, J., Carvalho, C.M., Lupski, J.R. *et al.* (2013) Microhomology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain. *PLoS Genet.*, **9**, e1003358.
51. Inoue, K., Osaka, H., Thurston, V.C., Clarke, J.T., 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.
52. Lieber, M.R., Ma, Y., Pannicke, U. and Schwarz, K. (2003) Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.*, **4**, 712–720.
53. Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y. and Pomerantz, R.T. (2015) Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. *Nat. Struct. Mol. Biol.*, **22**, 230–237.
54. McVey, M. and Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.*, **24**, 529–538.
55. Carvalho, C.M., Zhang, F. and Lupski, J.R. (2011) Structural variation of the human genome: mechanisms, assays, and role in male infertility. *Syst. Biol. Reprod. Med.*, **57**, 3–16.
56. Paques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.*, **63**, 349–404.
57. Ade, C., Roy-Engel, A.M. and Deininger, P.L. (2013) Alu elements: an intrinsic source of human genome instability. *Curr. Opin. Virol.*, **3**, 639–645.

58. Malkova, A. and Ira, G. (2013) Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.*, **23**, 271–279.
59. Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S. and Malkova, A. (2013) Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature*, **502**, 389–392.
60. Reiter, L.T., Hastings, P.J., Nelis, E., De Jonghe, P., Van Broeckhoven, C. and Lupski, J.R. (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am. J. Hum. Genet.*, **62**, 1023–1033.
61. Mundia, M.M., Desai, V., Magwood, A.C. and Baker, M.D. (2014) Nascent DNA synthesis during homologous recombination is synergistically promoted by the rad51 recombinase and DNA homology. *Genetics*, **197**, 107–119.
62. Shuvarikov, A., Campbell, I.M., Dittwald, P., Neill, N.J., Bialer, M.G., Moore, C., Wheeler, P.G., Wallace, S.E., Hannibal, M.C., Murray, M.F. *et al.* (2013) Recurrent HERV-H-mediated 3q13.2-q13.31 deletions cause a syndrome of hypotonia and motor, language, and cognitive delays. *Hum. Mutat.*, **34**, 1415–1423.
63. Park, J.Y., Yoo, H.W., Kim, B.R., Park, R., Choi, S.Y. and Kim, Y. (2008) Identification of a novel human Rad51 variant that promotes DNA strand exchange. *Nucleic Acids Res.*, **36**, 3226–3234.
64. Hastings, P.J., Lupski, J.R., Rosenberg, S.M. and Ira, G. (2009) Mechanisms of change in gene copy number. *Nat. Rev. Genet.*, **10**, 551–564.
65. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell*, **16**, 739–751.
66. Gafner, J. and Philippsen, P. (1980) The yeast transposon Ty1 generates duplications of target DNA on insertion. *Nature*, **286**, 414–418.
67. Hoang, M.L., Tan, F.J., Lai, D.C., Celniker, S.E., Hoskins, R.A., Dunham, M.J., Zheng, Y. and Koshland, D. (2010) Competitive repair by naturally dispersed repetitive DNA during non-allelic homologous recombination. *PLoS Genet.*, **6**, e1001228.
68. Lemoine, F.J., Degtyareva, N.P., Lobachev, K. and Petes, T.D. (2005) Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell*, **120**, 587–598.
69. Argueso, J.L., Westmoreland, J., Mieczkowski, P.A., Gawel, M., Petes, T.D. and Resnick, M.A. (2008) Double-strand breaks associated with repetitive DNA can reshape the genome. *Proc. Natl Acad. Sci. USA*, **105**, 11845–11850.
70. Song, W., Dominska, M., Greenwell, P.W. and Petes, T.D. (2014) Genome-wide high-resolution mapping of chromosome fragile sites in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **111**, E2210–E2218.
71. Anand, R.P., Tsaponina, O., Greenwell, P.W., Lee, C.S., Du, W., Petes, T.D. and Haber, J.E. (2014) Chromosome rearrangements via template switching between diverged repeated sequences. *Genes Dev.*, **28**, 2394–2406.
72. Boone, P.M., Bacino, C.A., Shaw, C.A., Eng, P.A., Hixson, P.M., Pursley, A.N., Kang, S.H., Yang, Y., Wiszniewska, J., Nowakowska, B.A. *et al.* (2010) Detection of clinically relevant exonic copy-number changes by array CGH. *Hum. Mutat.*, **31**, 1326–1342.
73. Ou, Z., Kang, S.H., Shaw, C.A., Carmack, C.E., White, L.D., Patel, A., Beaudet, A.L., Cheung, S.W. and Chinault, A.C. (2008) Bacterial artificial chromosome-emulation oligonucleotide arrays for targeted clinical array-comparative genomic hybridization analyses. *Genet. Med.*, **10**, 278–289.
74. 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.
75. Campbell, I.M., Yuan, B., Robberecht, C., Pfundt, R., Szafranski, P., McEntagart, M.E., Nagamani, S.C., Erez, A., Bartnik, M., Wisniowiecka-Kowalnik, B. *et al.* (2014) Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. *Am. J. Hum. Genet.*, **95**, 173–182.