Human Molecular Genetics, 2009, Vol. 18, No. 12 **2188–2203**
doi:10.1093/hmg/ddp151
Advance Access published on March 26, 2009

Complex rearrangements in patients with duplications of *MECP2* can occur by fork stalling and template switching

Claudia M.B. Carvalho[1,3], Feng Zhang[1], Pengfei Liu[1], Ankita Patel[1], Trilochan Sahoo[1], Carlos A. Bacino[1], Chad Shaw[1], Sandra Peacock[1], Amber Pursley[1], Y. Jane Tavyev[2], Melissa B. Ramocki[2,5], Magdalena Nawara[4], Ewa Obersztyn[4], Angela M. Vianna-Morgante[3], Pawel Stankiewicz[1], Huda Y. Zoghbi[1,2,5], Sau Wai Cheung[1] and James R. Lupski[1,2,5,*]

[1]Department of Molecular and Human Genetics and [2]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA, [3]Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, 05508-900 São Paulo, Brazil, [4]Department of Medical Genetics, Institute of Mother and Child, Warsaw 01-211, Poland and [5]Texas Children's Hospital, Houston, TX 77030, USA

Received November 10, 2008; Revised and Accepted March 24, 2009

Duplication at the Xq28 band including the *MECP2* gene is one of the most common genomic rearrangements identified in neurodevelopmentally delayed males. Such duplications are non-recurrent and can be generated by a non-homologous end joining (NHEJ) mechanism. We investigated the potential mechanisms for *MECP2* duplication and examined whether genomic architectural features may play a role in their origin using a custom designed 4-Mb tiling-path oligonucleotide array CGH assay. Each of the 30 patients analyzed showed a unique duplication varying in size from ∼250 kb to ∼2.6 Mb. Interestingly, in 77% of these non-recurrent duplications, the distal breakpoints grouped within a 215 kb genomic interval, located 47 kb telomeric to the *MECP2* gene. The genomic architecture of this region contains both direct and inverted low-copy repeat (LCR) sequences; this same region undergoes polymorphic structural variation in the general population. Array CGH revealed complex rearrangements in eight patients; in six patients the duplication contained an embedded triplicated segment, and in the other two, stretches of non-duplicated sequences occurred within the duplicated region. Breakpoint junction sequencing was achieved in four duplications and identified an inversion in one patient, demonstrating further complexity. We propose that the presence of LCRs in the vicinity of the *MECP2* gene may generate an unstable DNA structure that can induce DNA strand lesions, such as a collapsed fork, and facilitate a Fork Stalling and Template Switching event producing the complex rearrangements involving *MECP2*.

INTRODUCTION

Loss-of-function mutations of the gene enconding X-linked methyl-CpG-binding protein 2 (*MECP2*) at Xq28 are associated with Rett syndrome [RTT (MIM 312750)], a neurodevelopmental disorder that affects 1:10 000 girls. In boys, similar mutations are associated with syndromic and non-syndromic forms of developmental delay (DD), mental retardation (MR) and fatal infantile encephalopathy. Recent reports suggest that increased *MECP2* gene copy number also conveys a clinical phenotype, resulting in a DD/MR plus seizures syndrome in males (1–4). It is increasingly apparent that many neurological disease traits do not result from coding region mutations, but instead manifest due to subtle alterations in gene copy number caused by submicroscopic genomic duplications or deletions (e.g. Charcot–Marie–Tooth disease type 1A, Pelizaeus-Merzbacher disease, Parkinson disease,

[*]To whom correspondence should be addressed at: Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604B, Houston, TX 77030, USA. Tel: +1 7137983723; Fax: +1 7137985073; Email: jlupski@bcm.tmc.edu

© The Author 2009. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

Alzheimer disease, Williams-Beuren syndrome and the reciprocal dup(7)(q11.23q11.23), 17q21.31 deletion syndrome etc.) (5,6).

Lubs *et al*. (7) described the first linkage study of a family in which five males manifested X-linked mental retardation (XLMR); the causal locus was narrowed to a 5 cM interval in Xq28, including *MECP2* [MRXSL (MIM 300260)]. Ariani *et al*. (8) reported the first duplication of the *MECP2* gene in a girl diagnosed as having the preserved-speech variant of Rett syndrome. However, they could not assess whether the duplication involved the entire gene or only the interrogated *MECP2* exon due to methodology constraints. Subsequently, duplications of the *MECP2* gene were reported in males with DD/MR and several other associated clinical signs. In contrast to the patient reported by Ariani *et al*. (8), *MECP2* duplication female carriers are generally asymptomatic mothers of affected males and display a skewed X-chromosome inactivation pattern (1–4).

The predominant clinical features of males carrying *MECP2* gene duplications are DD/MR (100%), infantile hypotonia (100%), absent speech (84%) and a history of recurrent infections (80%); other clinical signs such as genital or digital abnormalities, seizures, or lack of ambulation, sometimes followed by other sporadic signs, may be present in 40–60% of the patients (3). The reason for such phenotypic variability is not fully understood. However, duplication disorders such as Charcot–Marie–Tooth disease type 1A due to the CMT1A duplication have long been known to have variable phenotypes that can even be discordant in identical twins (9). Recently, our group observed that autism is a defining feature of boys with the *MECP2* duplication syndrome, and female carriers also manifest neuropsychiatric phenotypes despite the nearly 100% favorable skewing of X-inactivation in peripheral blood (10).

The molecular mechanism(s) associated with *MECP2* copy-number alteration is still not well defined. Characteristically, such rearrangements are non-recurrent and show a broad size variation among patients. Non-recurrent rearrangements such as these do not usually originate by non-allelic homologous recombination (NAHR), responsible for the generation of a number of recurrent rearrangements associated with genomic disorders (11). Instead, non-homologous end joining (NHEJ) and break-induced replication (BIR) have been suggested as possible mechanisms producing the *MECP2* rearrangements (3,12). An alternate possibility is the recently described replication-based mechanism, Fork Stalling and Template Switching (FoSTeS) (13). FoSTeS was proposed as the mechanism responsible for non-recurrent *PLP1* duplications at Xq22, associated with the genomic disorder Pelizaeus-Merzbacher disease [PMD (MIM 312080)]. FoSTeS is a replication-based mechanism that produces non-recurrent rearrangements potentially facilitated by the presence of low-copy repeats (LCRs) with unusual symmetry (e.g. inverted repeats). These rearrangements may be complex, for example, duplicated regions containing stretches of non-duplicated sequences, inverted regions, or triplications. Some of these complexities can be resolved only at the DNA sequence level as revealed by breakpoint junction sequencing (13). The FoSTeS model has been generalized and specific molecular details provided by the microhomology-mediated BIR (MMBIR) model that appears to be operative in all domains of life (14).

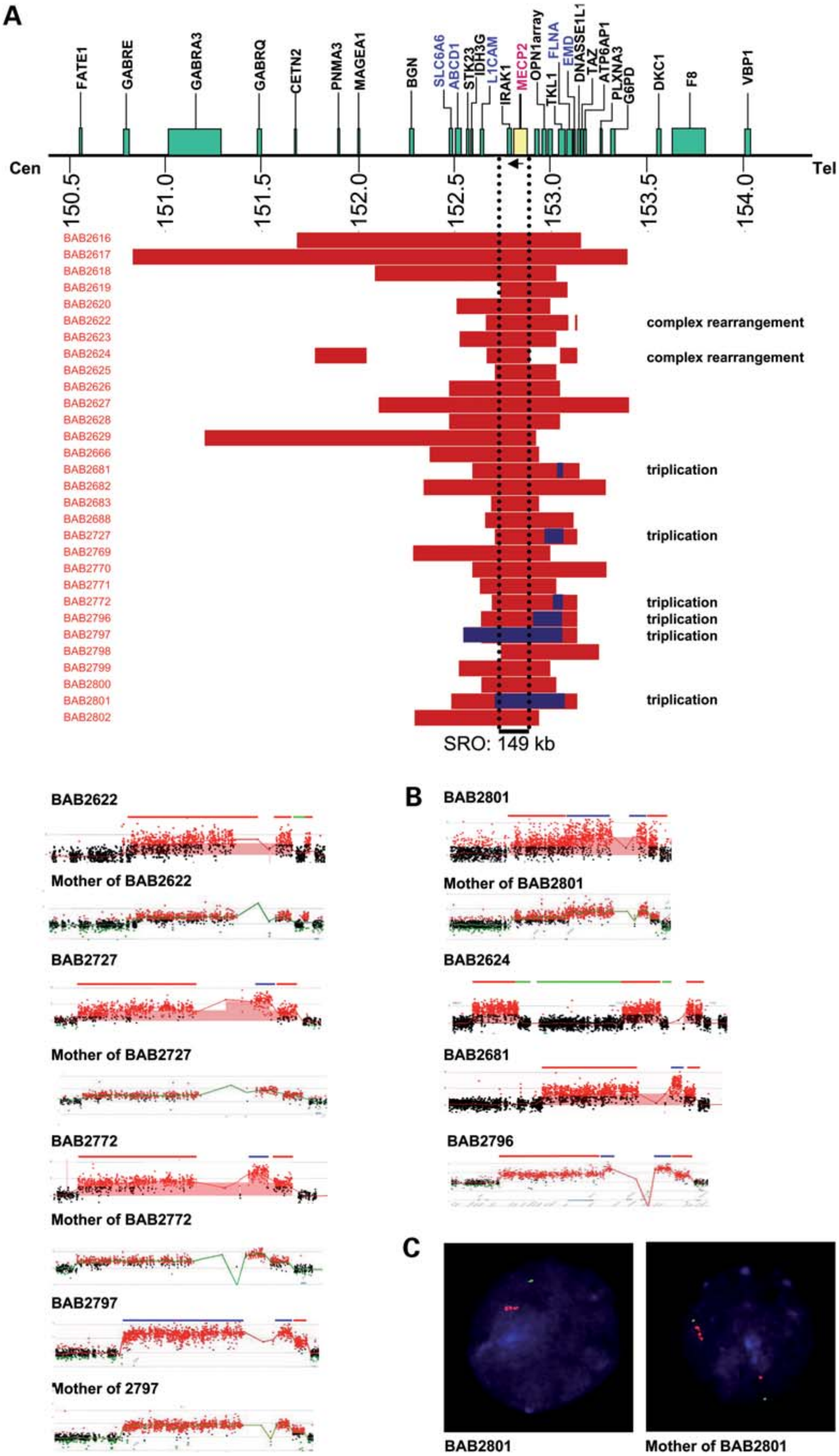
Based on the finding of a *MECP2* duplication containing a triplicated region (3), we hypothesized that FoSTeS could be a mechanism underlying the genomic duplication rearrangements including *MECP2*. To test this hypothesis, we ascertained 30 patients with *MECP2* duplications by chromosomal microarray analysis (CMA) and designed and applied a custom array for high-resolution genome analysis. This high-resolution analysis of the *MECP2* rearrangements revealed complex genomic alterations observed at the array level of resolution in eight patients (27%). Statistical and bioinformatic analyses of the genomic architecture surrounding the breakpoint regions revealed a non-random grouping of most of the distal breakpoints into two LCR clusters, both telomeric to the *MECP2* gene. Sequencing of five breakpoint junctions from four duplications revealed microhomologies at most of the junctions. One of the duplications had a complex breakpoint junction, including a gap and an inversion of the same region, consistent with the FoSTeS mechanism. FoSTeS thus appears to be a prevalent mechanism responsible for *MECP2* gene duplications, possibly triggered by genomic instability due to the presence of complex LCRs (LCRs consisting of a cluster of different repeat subunits lying in either direction) (11) that facilitate the formation of single-stranded DNA sequences.

RESULTS

Duplicated and complex rearrangements detected using custom oligonucleotide array CGH

Diagnostic testing in our Medical Genetics Laboratories and other collaborating laboratories identified 30 males with increased *MECP2* gene copy number using quantitative DNA methods and BAC or oligonucleotide-based CMA (15,16). FISH analyses confirmed duplication including *MECP2* (data not shown). To determine the size, extent, genomic content and fine map position of the breakpoints for each duplication, we designed a tiling path oligonucleotide microarray spanning 4 Mb surrounding the *MECP2* region at Xq28. Each duplication had a different size, ranging from 250 kb to 2.6 Mb (Fig. 1A, Table 1), and, therefore, varied in gene content (Fig. 1A). The smallest region of overlap (SRO) was 149 kb and included the *MECP2* and *IRAK1* genes.

The oligoarray resolution enabled us to detect the presence of complex rearrangements within the *MECP2* region in eight out of 30 patients (27%) (Fig. 1A and B). In two patients (BAB2622, BAB2624), the duplication had interspersed stretches of non-duplicated sequences. In six patients (BAB2681, BAB2727, BAB2772, BAB2796, BAB2797, BAB2801), the duplication contained a triplicated region (Fig. 1A and B). The triplications varied in size: 33 kb (BAB2681), 41 kb (BAB2772), 90 kb (BAB2727), 174 kb (BAB2796), 359 kb (BAB2801) and 537 kb (BAB2797), and, in most cases did not involve the *MECP2* gene, except in patients BAB2797 and BAB2801. Oligonucleotide array CGH revealed that all complexities were inherited from the mother in all cases tested (5 out of 8: BAB2622, BAB2727, BAB2772, BAB2797, BAB2801, Fig. 1B, Table 1). The


A
FATE1
GABRE
GABRA3
GABRQ
CETN2
PNMA3
MAGEA1
BGN
SLC6A6
ABCD1
STK23
IDH3G
L1CAM
IRAK1
MECP2
OPN1array
TKL1
FLNA
EMD
DNASSE1L1
TAZ
ATP6AP1
PLXNA3
G6PD
DKC1
F8
VBP1
Cen
Tel
150.5
151.0
151.5
152.0
152.5
153.0
153.5
154.0
BAB2616
BAB2617
BAB2618
BAB2619
BAB2620
BAB2622
BAB2623
BAB2624
BAB2625
BAB2626
BAB2627
BAB2628
BAB2629
BAB2666
BAB2681
BAB2682
BAB2683
BAB2688
BAB2727
BAB2769
BAB2770
BAB2771
BAB2772
BAB2796
BAB2797
BAB2798
BAB2799
BAB2800
BAB2801
BAB2802
complex rearrangement
complex rearrangement
triplication
triplication
triplication
triplication
triplication
triplication
SRO: 149 kb
BAB2622
Mother of BAB2622
BAB2727
Mother of BAB2727
BAB2772
Mother of BAB2772
BAB2797
Mother of 2797
B
BAB2801
Mother of BAB2801
BAB2624
BAB2681
BAB2796
C
BAB2801
Mother of BAB2801

MECP2 triplication present in patients BAB2797 and BAB2801 were inherited from the mother as shown by oligonucleotide CGH and/or FISH (Fig. 1C).

Features of a rearrangement prone region distal to *MECP2*

Bioinformatic analysis of the *MECP2* genomic region was performed to investigate for the presence and characterize the structure of LCRs mapping within the interval. We found two LCR groups, in direct and inverted orientations, 47 kb and 201 kb telomeric to the *MECP2* gene, respectively. del Gaudio *et al*. (3) designated these LCRs J and K. LCR J spans 114 kb and is formed by two genes that constitute the Opsin array, *OPN1LW* and *OPN1MW*, arranged in a head-to-tail tandem array and interrupted by copies of a gene with unknown function, *TEX28* (Fig. 2). Each *OPN1* plus *TEX28* array set has a size range of 37–38 kb and they are termed JA, JB and JC. The *OPN1LW* and *OPN1MW* genes encode the visual red and green pigments, respectively. The nearby K1 and K2 LCRs are positioned in inverted orientation, have 99% sequence identity and are 11.3 kb long. In total, the LCR-containing region spans 215 kb.

Remarkably, we mapped 77% (23/30) of the distal duplication breakpoints in one of the LCRs within the ~215 kb region (Fig. 2). This region represents <10% of the 2.6 Mb genomic interval in which breakpoints were mapped. The K1 and K2 inverted LCRs seem especially associated with the complex rearrangements since at least one of their breakpoints mapped at these repeats. The distal breakpoints of all six triplications were mapped to the proximal K LCR (Fig. 2).

As one way to model the expected distribution of the distal endpoints for these segmental copy number changes, we used the locations of the proximal endpoints as an internal control. Assuming a completely random model for breakpoint occurrence, we observed 30 proximal endpoints over an ~1.8 Mb region. Using the Poisson distribution and Poisson process assumptions, these observations suggest a rate of breakpoint occurrence of approximately one breakpoint every 60 kb in our patient cohort. We can use this distribution of proximal endpoints to model a completely random-based case for the distribution of distal breakpoints. Under this Poisson model, a 215 kb region would have approximately 3.58 (about 3 or 4) breakpoint events in a cohort of 30 patients. However, 23/30 *MECP2* duplication distal breakpoints group within this 215 kb LCR containing region giving us an excess of 19.42 or 6-fold more than expected ($P < 2.2 \times 10^{-16}$). It is important to note that the breakpoints of the complex rearrangements (e.g. the triplications) were not considered in the calculations, thus, even though we used a conservative sum, this simple Poisson model suggests an extremely unusual concentration of distal breakpoints.

To further refine our analysis of breakpoint positioning in the genome, we performed 10 000 replicate Monte Carlo simulations to better evaluate the non-random distribution of breakpoints considering a random allocation of segments and requiring them to overlap the *MECP2* gene. To summarize these Monte Carlo outcomes, we determined two summary statistics for each of the simulation runs: the variance of the segment locations and the mean segment distal breakpoint. We find that our segments have far less variable locations (as measured by average breakpoint location) than would be expected for randomly simulated segments as only 415 of 10 000 or 0.0415 of the simulations segments had locations as variable or less variable than what we observed. We also found our observed segments to be very highly shifted relative to what would be expected. Only 165 of 10 000 or 0.0165 of the simulation runs had an average distal breakpoint that mapped as centromeric as what we observed (Fig. 2A). These results are represented in the graphic seen in Figure 2B. Moreover, the average distal breakpoint of our data maps at 153.0888 Mb, included within the 215 kb laden-LCR region, while the simulation results have an average breakpoint of 153.1899 Mb, which is outside of the LCR region (Fig. 2A and B). Taken together, these Monte Carlo results demonstrate not only that our segments are highly non-randomly distributed across this genomic region but also they are grouping around the shifted average breakpoint (153.0888 Mb), even accounting for the bias introduced by the fact that the segments overlap *MECP2*.

Additionally, five out of seven distal breakpoints that do not map to the 215 kb LCR region map in proximity to another pair of LCRs located >400 kb telomeric to the *MECP2* gene. This LCR pair is constituted by 99% identical subunits (L1 and L2), lying in inverted orientation to each other (3) (Fig. 2A).

Breakpoint junction analysis reveals both simple and complex rearrangements

The breakpoint junctions were amplified using outward facing primers positioned at both ends of the duplication, as determined by aCGH, under the assumption that the repeated

Figure 1. (**A**) Genomic region harboring duplications in our cohort of 30 patients analyzed by oligonucleotide array CGH. Solid red bars represent oligonucleotide probes for which the mean normalized $\log_2$ (Cy5/Cy3) ratio of the CGH signal, amid a 5 kb window, reached a threshold of 0.6, indicating a duplication. The graphical normalized data for each patient was obtained by applying the most distal and proximal oligonucleotide genomic probe coordinates to the custom track at UCSC site http://genome.ucsc.edu/cgi-bin/hgGateway. Positions are given relative to NCBI Build 35 for the X chromosome. The blue region represents triplicated regions. 'Complex rearrangements' refers to array results from those patients who presented duplicated regions interrupted by non-duplicated regions as well as duplicated regions interrupted by triplicated regions. Eight out of 30 (27%) in our cohort were found to have complex rearrangements by array CGH. The Smallest Region of Overlap (SRO) is 149 kb and contains two genes, *IRAK1* and *MECP2*. Note BAB2626 and BAB2628 have the same size duplication, as expected since they are affected sibs who share the same mother. (**B**) Agilent 4X44K oligoarray CGH results for the eight patients (plus mothers in five cases) carrying complex rearrangements. Red lines above the diagram represent duplicated segments, blue lines represent triplicated segments and green lines represent normal copy number stretches of DNA contained within duplicated sequences. The gaps between the lines represent intervals where we could not infer the copy number status due to poor probe coverage secondary to LCRs. The complexities were inherited from the mother as shown by oligonucleotide array CGH in five cases. (**C**) FISH detection of *MECP2* triplication in male patient BAB2801 and in his mother using RP11-119A22 clone containing the *MECP2* gene (red signal) and RP11-137H15 (green signal) as control; that result confirms the triplication revealed by oligonucleotide array CGH.

Table 1: Summary of results for the *MECP2* duplication observed in 30 male patients.

Patient	Duplication rearrangement size (kb)	Coordinates: first nucleotide dup to last nucleotide dup (hg17)	Origin of the duplication	Presence of complexities at the OligoCGH level	Presence of complexities at the breakpoint junctions	Mechanism for formation
BAB2616	1 470	151 680 732:153 151 518	NA	No	-	ND
BAB2617	2 560	150 831 388:153 392 578	Maternal	No	–	ND
BAB2618	940	152 085 597:153 025 406	*de novo*	No	–	ND
BAB2619	352	152 732 560:153 084 321	NA	No	–	ND
BAB2620	486	152 508 915:152 994 739	*de novo* – insert to Y chr	No	–	ND
BAB2622	474	152 660 054:153 134 047	Maternal	dup-nl-dup	–	FoSTeS
BAB2623	500	152 525 081:153 025 406	Maternal	No	–	ND
BAB2624	1 360	151 775 349:153 133 440	NA	dup-nl-dup-nl-dup	–	FoSTeS
BAB2625	321	152 704 295:153 025 406	NA	No	–	ND
BAB2626	573	152 470 635:153 044 062	NA	No	–	ND
BAB2627	1 295	152 104 966:153 399 819	Maternal	No	–	ND
BAB2628	573	152 470 635:153 044 062	NA	No	–	ND
BAB2629	1 715	151 204 101:152 918 778	Maternal	No	No; No microhomology	NHEJ
BAB2666	565	152 370 215:152 935 149	Maternal	No	–	ND
BAB2681	558	152 588 414:153 146 624	Maternal	dup-trip-dup	–	ND
BAB2682	944	152 338 152:153 282 262	Maternal	No	–	ND
BAB2683	248	152 687 147:152 935 149	Maternal	No	–	ND
BAB2688	461	152 654 266:153 114 864	Maternal	No	No; TCC microhomology	FoSTeS × 1 or NHEJ
BAB2727	430	152 704 295:153 134 047	Maternal	dup-trip-dup	Yes: inversion GG microhomology ACA insertion + CA microhomology	FoSTeS × 3
BAB2769	711	152 284 103:152 994 739	Maternal	No	–	ND
BAB2770	693	152 591 273:153 283 750	Maternal	No	No; GCCT microhomology	FoSTeS × 1 or NHEJ
BAB2771	396	152 629 017:153 025 406	Maternal	No	–	ND
BAB2772	444	152 690 284:153 134 047	Maternal	dup-trip-dup	–	FoSTeS
BAB2796	497	152 637 272:153 134 047	Maternal	dup-trip-dup	–	FoSTeS
BAB2797	586	152 548 079:153 134 047	Maternal	nl-trip-dup	–	FoSTeS
BAB2798	506	152 737 680:153 244 088	Maternal	No	–	ND
BAB2799	476	152 518 706:152 994 739	NA	No	–	ND
BAB2800	389	152 637 719:153 026 519	Maternal	No	–	ND
BAB2801	653	152 481 328:153 134 047	Maternal	dup-trip-dup	–	FoSTeS
BAB2802	645	152 290 619:152 935 149	Maternal	No	–	ND

NA, parents not available; ND, not defined; dup, duplicated region; nl; normal copy number region; trip, triplication; —, junction sequences not obtained

copies were arranged in tandem. Difficulties amplifying unique junctions and resolving them at the nucleotide level were anticipated, because the genomic interval surrounding *MECP2* is laden with LCRs (3) thus often hampering the ability to uniquely position the exact breakpoint in the genome. Several other breakpoints were not obtained under the above assumptions despite multiple attempts, consistent with potential further unanticipated complexity at the breakpoints. Five breakpoint junctions from four duplications were obtained by long-range PCR and directly sequenced (Fig. 3A–D). Three out of four rearrangements were simple, resulting in tandem duplications (BAB2629, BAB2688 and BAB2770). The fourth one (BAB2727) was complex, and included an inversion.

The breakpoint junction from the duplication in BAB2629 was amplified using forward and reverse primers facing outwards, designed at the very end of the duplication breakpoints as determined by aCGH. The PCR product was sequenced and the breakpoint junction showed no microhomology between the distal and proximal reference sequences (Fig. 3A). The distal breakpoint occurred within an *Alu*Jo element, whereas the proximal breakpoint occurred within one of the introns of the *GABRA3* gene. No SINEs/LINEs were detected within the surrounding 175 nt genomic interval (Fig. 3A).

The breakpoint junctions from the duplication in patients BAB2688 and BAB2770 were obtained as described for BAB2629. The distal breakpoint of BAB2688 duplication occurred in one of the *FLNA* introns and the proximal occurred in one of the *L1CAM* introns, very close to the next exon boundary. No SINEs/LINEs were detected within a 175 nt window around either proximal or distal breakpoint reference sequences. A TCC microhomology was observed at the junction point (Fig. 3B). In patient BAB2770, a microhomology of 4 nt (GCCT) was detected. The distal breakpoint was located in an *Alu*Sx element, whereas the proximal one was located in one of the exons of the *PDZK4* gene (Fig. 3C).

Patient BAB2727 has a complex rearrangement observed at oligonucleotide array resolution: a triplication spanning ~90 kb encompassing the Opsin array and the proximal LCR K (Fig. 2A and 3D). The non-functional copy of the Opsin array plus the functional gene, *TKL1*, are involved in the triplication. We designed forward and reverse primers for the duplicated and triplicated segment ends based on the oligoarray results. Long-range PCR reactions were performed using all

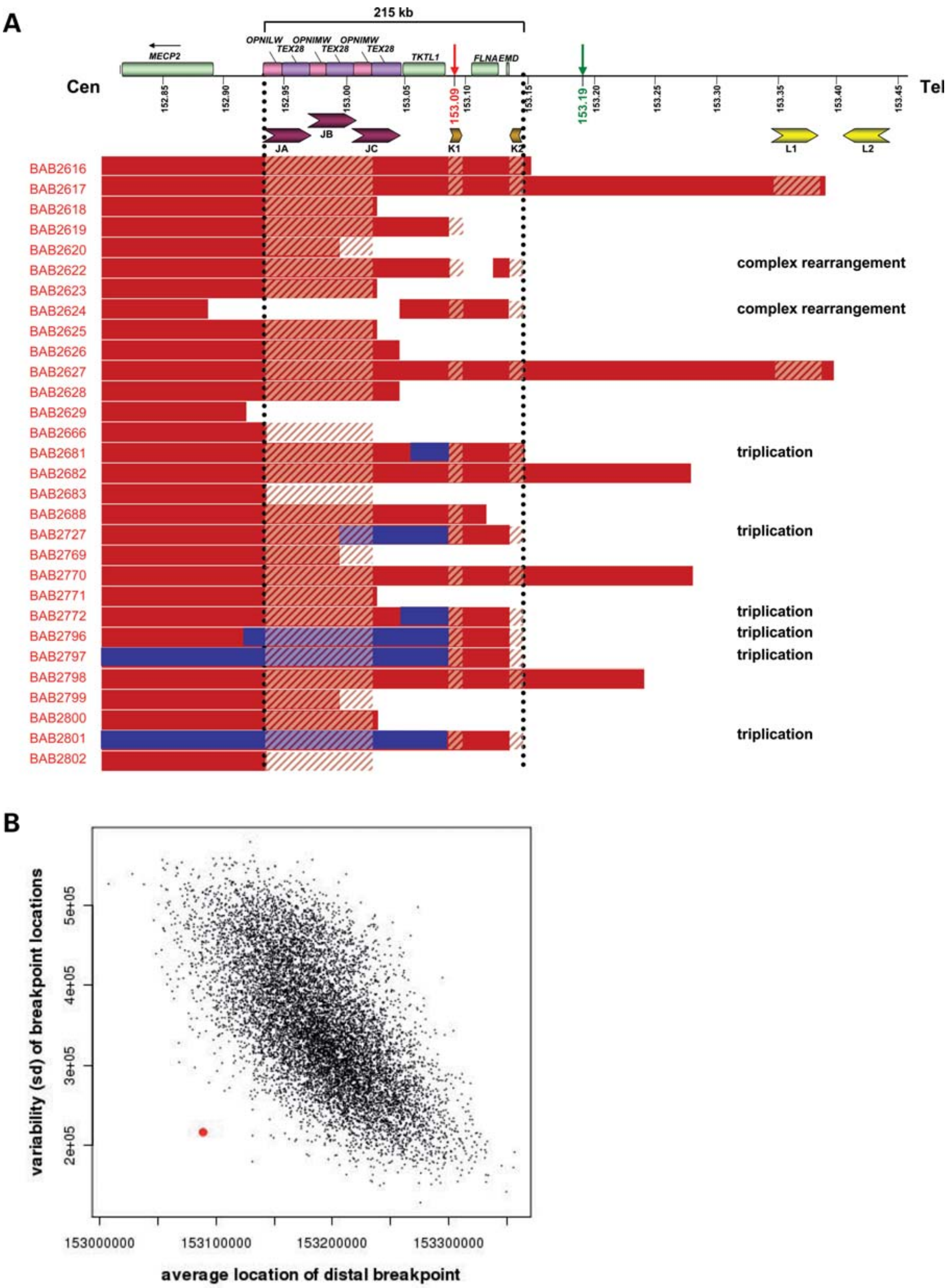


Figure 2. Breakpoints/join points group near or at the SDs around the *MECP2* gene. **(A)** Solid red bars represent duplicated regions and blue bars represent triplicated regions. LCR J spans 114 kb and is formed by three genes that constitute the Opsin array, *OPN1LW* and *OPN1MW*, arranged in head-to-tail tandem array interrupted by copies of *TEX28*. The nearby LCRs, K1 and K2, are positioned in inverted orientation, have 99% sequence identity and are 11.3 kb in length. Hatched bars represent intervals with poor probe coverage. The red and dark green arrows point out, respectively, the genomic coordinates of the observed and the expected average distal breakpoint as calculated by the Monte Carlo algorithm. **(B)** Scatter plot of the results of the Monte Carlo simulation with the two summary outcomes shown. The red dot represents the average distal breakpoint for our cohort; black dots show the distribution of 10 000 replicate Monte Carlo simulations. Y-axis: variance distribution of the segment locations; X-axis: location of the calculated average distal breakpoints. Positions are given relative to NCBI Build 35 for the X chromosome.

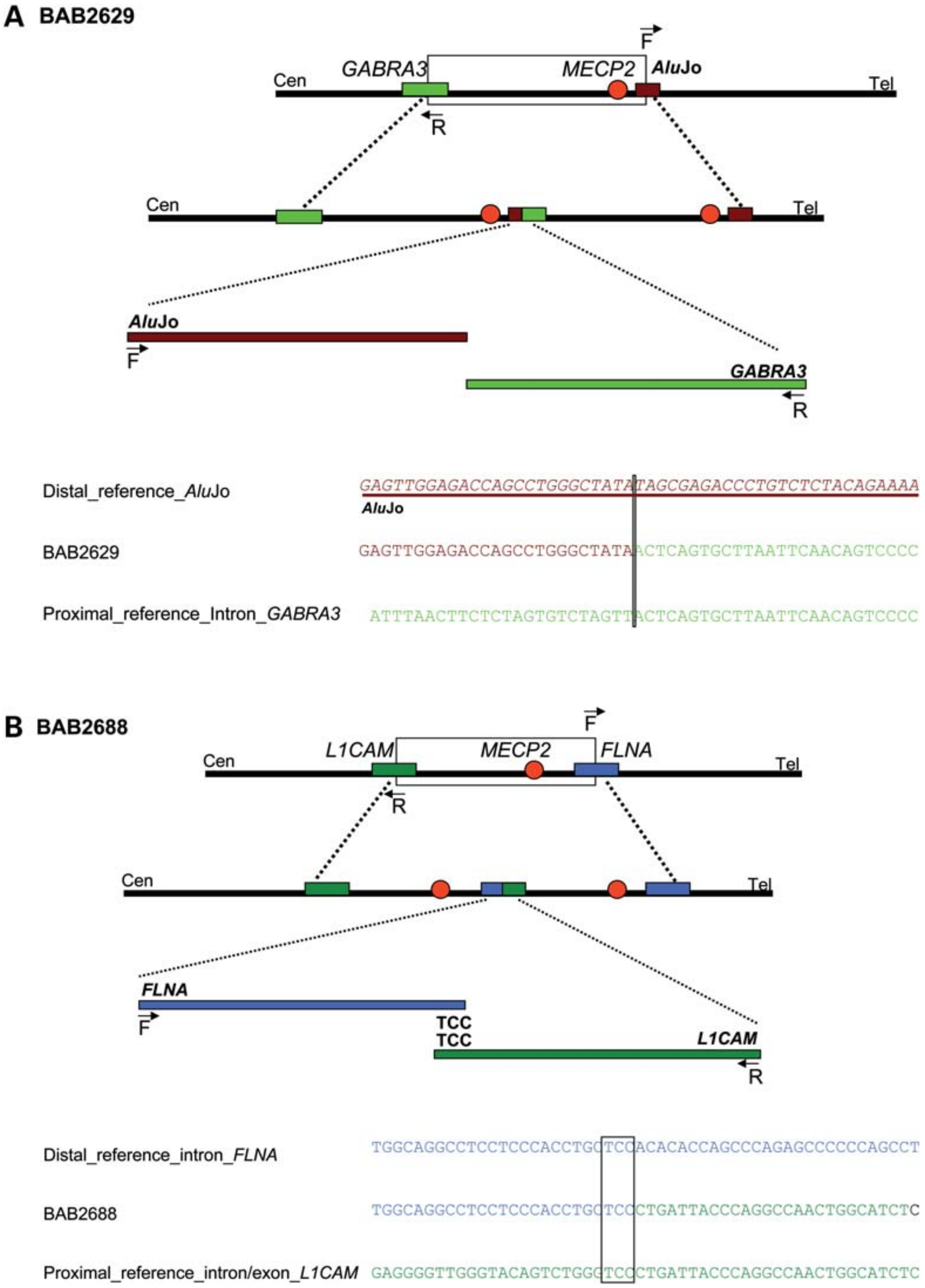

A BAB2629
F
Cen
GABRA3
MECP2
AluJo
Tel
R
Cen
Tel
AluJo
F
GABRA3
R
Distal_reference_AluJo
GAGTTGGAGACCAGCCTGGGCTATATAGCGAGACCCTGTCTCTACAGAAAA
AluJo
BAB2629
GAGTTGGAGACCAGCCTGGGCTATAACTCAGTGCTTAATTCAACAGTCCCC
Proximal_reference_Intron_GABRA3
ATTTAACTTCTCTAGTGTCTAGTTACTCAGTGCTTAATTCAACAGTCCCC
B BAB2688
F
Cen
L1CAM
MECP2
FLNA
Tel
R
Cen
Tel
FLNA
F
TCC
TCC
L1CAM
R
Distal_reference_intron_FLNA
TGGCAGGCCTCCTCCCACCTGCTCCACACACCAGCCCAGAGCCCCCCAGCCT
BAB2688
TGGCAGGCCTCCTCCCACCTGCTCCCTGATTACCCAGGCCAACTGGCATCTC
Proximal_reference_intron/exon_L1CAM
GAGGGGTTGGGTACAGTCTGGGTCCCTGATTACCCAGGCCAACTGGCATCTC

Figure 3. (**A**–**C**) Breakpoints/Join points of the duplications in patients BAB2629, BAB2688, BAB2770, respectively. Reference and rearranged genomic structures are shown. The duplicated region is boxed with a black rectangle. Primers used to obtain breakpoint junctions by long-range PCR are shown (F and R). Colored rectangles expanded by dotted lines represent the proximal and distal breakpoint areas. The breakpoint/join point is expanded and the sequence is aligned to the distal and proximal genomic references (to facilitate visualization sequence colors match the colored rectangles). The microhomologies, are boxed. Inserted nucleotides are written in black. The repetitive elements present near or at the join points are underlined and their sequences italicized. Patient BABA2629 has no microhomology; the thin rectangle represents the breakpoint junction location. **(D)** Complex rearrangements observed for BAB2727. The duplicated region is boxed in the black rectangle and the triplicated/inverted region is boxed in the blue rectangle. Primers used to obtain the breakpoint junctions by long-range PCR are shown (R and R). We sequenced two out of three join points (represented as FoSTeS × 1, ×2 and ×3), enlarged below. All sequences are aligned to the distal and proximal genomic references in anti-sense orientation except the light green one for the Proximal_reference_intron_*ARHGAP4*, which is in sense orientation (representing the join point sequence as we observed). A complete alignment of this join point is available as Supplementary Material, Figure S1. **(E)** Predicted order, origins, and relative orientations of duplicated/triplicated sequences for patient BAB2727. Arrowheads show direction of DNA relative to the positive strand; circled numbers represent a FoSTeS event. Asterisk (*) indicates a non-sequenced join junction.

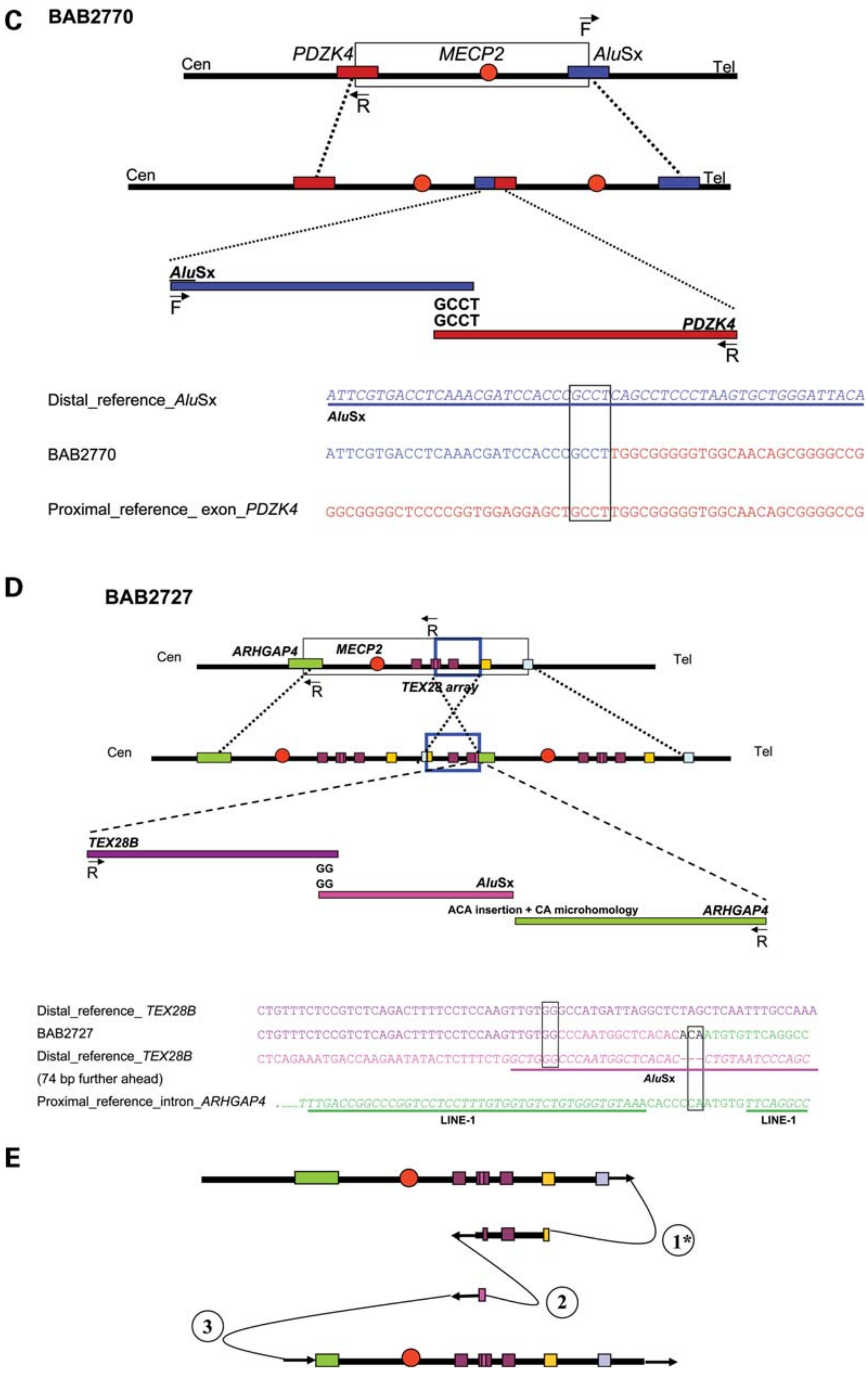

C
BAB2770
F
PDZK4
MECP2
AluSx
Cen
Tel
R
Cen
Tel
AluSx
F
GCCT
GCCT
PDZK4
R
Distal_reference_AluSx
ATTCGTGACCTCAAACGATCCACCCGCCTCAGCCTCCCTAAGTGCTGGGATTACA
AluSx
BAB2770
ATTCGTGACCTCAAACGATCCACCCGCCTTGGCGGGGGTGGCAACAGCGGGGCCG
Proximal_reference_ exon_PDZK4
GGCGGGGCTCCCCGGTGGAGGAGCTGCCTTGGCGGGGGTGGCAACAGCGGGGCCG
D
BAB2727
R
Cen
ARHGAP4
MECP2
Tel
R
TEX28 array
Cen
Tel
TEX28B
R
GG
GG
AluSx
ACA insertion + CA microhomology
ARHGAP4
R
Distal_reference_ TEX28B
CTGTTTCTCCGTCTCAGACTTTTCCTCCAAGTTGTGGGCCATGATTAGGCTCTAGCTCAATTTGCCAAA
BAB2727
CTGTTTCTCCGTCTCAGACTTTTCCTCCAAGTTGTGGCCCAATGGCTCACACACAATGTGTTCAGGCC
Distal_reference_TEX28B
(74 bp further ahead)
CTCAGAAATGACCAAGAATATACTCTTTCTGGCTGGGCCCAATGGCTCACAC---CTGTAATCCCAGC
AluSx
Proximal_reference_intron_ARHGAP4
.......TTTGACCGGCCCGGTCCTCCTTTGTGGTGTCTGTGGGTGTAAACACCCCAATGTGTTCAGGCC
LINE-1
LINE-1
E
1*
2
3

Figure 3. Continued.

possible primer combinations in both direct and inverted orientation. One of the breakpoint junctions was obtained using the proximal duplicated reverse primer plus the proximal triplicated reverse primer (Fig. 3D; the complete alignment of 175 nt flanking the breakpoints can be seen in Supplementary Material, Fig. S1). The sequence data revealed that the triplicated segment was inserted in an inverted orientation with respect to the reference genome. The breakpoint junction alignment plus the oligoarray probes involved in the triplication revealed that the first template switch probably occurred at the second *TEX28* gene, located within the Opsin array (here termed as *TEX28B*). A second template switch, with the same strand orientation, occurred again and produced a gap of 74 bp at the junction. The new alignment region (an *Alu*Sx element) showed a 2 bp microhomology (GG) between proximal and distal reference sequences. An additional ACA triplet was inserted at the third breakpoint junction. Interestingly, the end of the second breakpoint junction was a short CACA repeat, suggesting that the polymerase enzyme may have 'stuttered' at this point and inserted the ACA by mistake. It may be generated by slipped-strand mispairing during DNA replication as there is a CA repeat before the junction (Fig. 3D). The third breakpoint occurred within the *Alu*Sx element, 15 bp from the second junction. This new junction aligned to the proximal duplicated region, extending into one of the introns of the *ARHGAP4* gene. There was a 2 bp nucleotide microhomology (CA) between the breakpoint junction and the reverse strand within the intron of the *ARHGAP4* gene; thus the ACA insertion due to the polymerase 'stuttering' preceded a switch of strand orientation. Analysis of this flanking region revealed a LINE element (L1 family) starting 5 bp from this junction point. Figure 3D summarizes our findings. We repeated the complete assay for this patient starting from the genomic DNA and confirmed the above findings. This complex rearrangement can be explained by a FoSTeS × 3 event (Fig. 3E). A summary of these results is presented in Table 1.

Presence of repetitive elements and high incidence of CTG/CAG motifs at the breakpoint junctions

Repetitive elements such as SINEs and LINEs were present at some, but not all, breakpoints in our cohort. *Alu* families were observed in four breakpoints in different patients and a LINE element was observed just once, 5 nt away from the junction (BAB2727). An *Alu*Jo family member was detected at the distal breakpoint of the BAB2629 duplication and an *Alu*Sx was observed at the distal breakpoint in BAB2770 duplication. In neither BAB2629 nor BAB2770 duplication did the proximal breakpoint have repetitive elements within a window of 175 nt. Interestingly, BAB2727 had two break events at the same *Alu*Sx element, only 15 bp apart from each other (Fig. 3D). In both events, the similarity between the exchanging strands was restricted to microhomologies at their join points, as the *Alu* elements are present in just one of the strands. BAB2688 duplication presented no repetitive sequences either at or within a window of 175 nt around the breakpoint junctions.

We analyzed the DNA sequence around the join points of the three *MECP2* duplications potentially generated by FoSTeS (BAB2688, BAB2727 and BAB2770) for the presence of the trinucleotide sequence 5′-CTG-3′ and its complementary 5′-CAG-3′ sequence. Slack *et al.* (41) found a high incidence of the 5′-CTG-3′motif at the junctions produced by gene amplification apparently induced under stress in *Escherichia coli*. For our analysis, we defined a 175 nt window containing the breakpoint junctions at both distal and proximal reference genomic sequences. The CG median value for those genomic intervals was calculated (52%) in order to estimate the expected number of the CTG/CAG sequences. The expected number of CTG/CAG sequences was 45, whereas the observed number was 78, a statistically significant difference (Fisher exact test, $P < 0.003$).

We analyzed the breakpoint junction data reported for patients with *PLP1* duplications in order to examine whether the frequency of that trinucleotide motif is increased in other regions shown to have complexities at the breakpoint junctions [data provided from Lee *et al.* (13)]. Considering a window of 175 nt and a GC median value of 41.7%, we also found an increased frequency of CTG/CAG motifs at the *PLP1* duplication breakpoint junctions; the expected number of CTG/CAG sequences was 44, whereas the observed was 79 (Fisher exact test, $P < 0.0017$). Interestingly, Kornreich *et al.* (17) reported a high incidence of the 5′-CCTG-3′motif, which contains the trinucleotide CTG, at the junctions of non-homologous recombination sites causing deletions and duplications (including one complex rearrangement) in the human *GLA* gene.

DISCUSSION

High-resolution oligonucleotide array CGH confirms a broad size variation of duplications involving *MECP2*

MECP2 duplications are among the most frequently identified subtelomeric rearrangements in a cohort of 5380 cases referred for clinical array CGH (18). Duplications involving the *MECP2* gene are characteristically non-recurrent, that is, each rearrangement has a different size and distinct breakpoints. In our cohort of 30 patients, they vary from ∼250 kb to ∼2.6 Mb in size. Syndrome-associated non-recurrent rearrangements usually share a common genomic region-of-overlap, the smallest region of overlap or SRO, encompassing the locus associated with the conveyed genomic disorder. The analysis of 30 patients with *MECP2* rearrangements using high-density arrays enabled us to narrow the SRO to 149 kb. This region contains two genes, *IRAK1* and *MECP2*, and the increased dosage of either one or both may cause the clinical phenotypes; although the current data in the aggregate suggest over-expression of the *MECP2* gene as the main culprit. Supporting evidence comes from animal models as *MECP2* duplication syndrome was predicted by the observation that mice engineered to over-express *MECP2* develop a progressive neurological disorder, stereotypic and repetitive movements, epilepsy, spasticity, hypoactivity and early death (19). Furthermore, a patient with a complex rearrangement including *MECP2* triplication, had the most severe phenotype consistent with an *MECP2* dosage role in the development of the salient clinical endophenotypes (3).

The reason for the observed phenotypic variability among patients with *MECP2* duplication is not fully understood. Each patient has a different duplication size, as observed in this present work and by others (3,12), consequently, different genes are involved. Some are well-known genes implicated in human diseases when deleted or harboring pathological point mutations, for example, *ABCD1*, can cause adrenoleukodystrophy [ALD (MIM 300100)], *L1CAM*, can cause hydrocephalus due to congenital stenosis of the aqueduct of Sylvius [HSAS (MIM 307000)], rearrangements and point mutations of the *OPN1* array can cause colorblindness, deutan series [CBD (MIM 303800)] or protan series [CBP (MIM 303900)], *FLNA* can cause X-linked periventricular heterotopia [MIM 300049] and *EMD* can cause Emery–Dreifuss muscular dystrophy [EDMD (MIM 310300)]. Their roles, if any, when in functional disomy are not established. Reported data (2,4,20) suggest that duplication size is not correlated to severity or specificity of any neuropsychiatric phenotype and our data confirmed this contention (10). It is possible that genetic background plays a role in such variability, including SNP variation within the rearranged genomic segment on the non-duplicated allele (21), but further studies are required to clarify this hypothesis. Variable expressivity is a hallmark of duplication disorders; it occurs in patients carrying dup(7)(q11.23q11.23) (22–25), dup(17)(p11.2p11.2) (26), dup(17)(q12q12) (27), dup(22)(q11.2q11.2) (28), etc. Such a characteristic was remarkably shown by discordant phenotypes in identical twins with the CMT1A duplication (9).

A genomic interval prone to polymorphic and disease-causing rearrangements

A first glimpse of *MECP2* duplications at the array level reveals proximal and distal breakpoints scattered along the *MECP2* flanking regions. The size of each duplication is unique, and the breakpoints are different, thus revealing the prominent feature of such rearrangements, their non-recurrent nature. However, careful analysis of the distal breakpoints reveals a remarkable grouping pattern: 77% or 23/30 of the breakpoints mapped within a 215 kb region 47 kb distal from the *MECP2* gene. This 215 kb region is formed by the J LCR group (JA, JB, JC) and the K LCR group (K1 and K2) sharing 99% sequence identity within each LCR (3,29). Using a Monte Carlo algorithm and considering the bias introduced by the fact that all duplications must overlap the *MECP2* gene, we showed that not only the distal breakpoints are non-randomly distributed across the genomic region but also they present an unexpected average location within the LCR-laden region. This result confirmed a Poisson analysis, which revealed that this breakpoint grouping is highly significant ($P < 2.2 \times 10^{-16}$). Despite not considered in calculations, we can also clearly observe that the distal breakpoints of all six triplications detected in our cohort are grouped at the proximal K LCR.

LCR J is formed by two genes that constitute the Opsin array, *OPN1LW* and *OPN1MW*, plus three copies of the *TEX28* gene. The *OPN1LW* and *OPN1MW* genes provide one of the earliest examples of tandem repeats and NAHR associated with a common human trait, color blindness (30). Usually, more than one *OPN1MW* gene is present, but only the proximal is functional. Approximately 25% of Caucasians have one copy, 50% two copies and 25% three or more copies of the gene (reviewed in 31). The Opsin genes share 99% sequence identity and undergo frequent rearrangements, probably by unequal crossing-over between the highly similar units, and gene conversions producing common variation in color vision and red–green color vision defects.

The region between the LCRs K is inverted in 33% of females of European descent and represents a common polymorphism within this population (32). Non-recurrent deletions involving one of the K LCRs and the *EMD* gene have been reported to cause EDMD (32,33). Remarkably, another LCR inverted pair (L1 and L2, >400 kb telomeric to the *MECP2* gene) is associated with the two most distal breakpoint duplications. Each L subunit displays one copy of the *NEMO* gene (L1 has a functional copy and L2 has a non-functional copy), in which intragenic deletions produce most of the *de novo* cases of Incontinentia Pigmenti [IP (MIM 308300)] (34). Rearrangements and gene conversion involving L1 and L2 have been reported (35). In total, 83% (25/30) of the distal duplication breakpoints map within an LCR region, suggesting an association between the local genomic structure (presence of LCRs) and the occurrence of rearrangements therein.

In aggregate, these findings suggest that the genomic interval telomeric to *MECP2* is a highly unstable region that undergoes frequent rearrangements leading to either human population polymorphism or disease. The genomic architecture neighboring the *MECP2* gene, in this case the presence of the LCRs J and K, seems to have an important role in the origin of such events. LCRs mediating recurrent rearrangements through NAHR can cause deletions and duplications in many genomic disorders (5,11). For non-recurrent rearrangements there is evidence for an alternative role by which LCRs lead to genomic instability and consequently stimulate such events rather than mediate them (13,29,36–38). Bacolla *et al.* (39) showed an association of genomic rearrangement breakpoints with alternative DNA conformations, represented by non-B DNA (i.e. slipped, cruciform structures, etc.) that can potentially trigger genomic rearrangements in humans because they are expected to increase the rate of single-strand lesions at these sequences. Such alternative conformations can be formed by regions presenting, for instance, direct, mirror or inverted repeats. Thus, by inference LCRs may be appropriate substrates in which alternative conformations can form. We propose that J and K LCRs form non-B DNA structures, potentially cruciforms, and cause DNA single-strand lesions and rearrangements involving the *MECP2* gene. Single strand lesions can lead to collapsed replication forks and the formation of one-ended, double-stranded DNA (dsDNA) ends that stimulate FoSTeS/MMBIR (13,14). A similar ‘grouping’ of breakpoints on one side near complex genome architecture was also observed with *PLP1* duplications (13).

Complex rearrangements associated with *MECP2* duplications implicate FoSTeS

The use of high-resolution tiling-path oligonucleotide arrays to investigate rearrangements involving the *MECP2* gene enabled us to detect complexities at the array level. Such complexities were observed either as triplications of varying sizes interrupting the duplicated regions or duplicated regions interrupted by long stretches of normal copy number. At the sequencing level, we also detected a complex rearrangement in the breakpoint junction of patient BAB2727, involving at least three DNA strand discontinuities (FoSTeS × 3), including a triplication and an inversion of one of the DNA stretches (Fig. 3D). These complex alterations as described are not readily explained by either NAHR or NHEJ, the usual recombination-based mechanisms proposed to explain human genomic rearrangement events. NAHR is mostly mediated by LCRs with recombination hotspots, gene conversion and apparent minimal processing fragments (MEPS), usually producing recurrent rearrangements and clustering of breakpoints (reviewed in 36). Here, we demonstrate an absence of such characteristics in *MECP2* duplication samples.

NHEJ and homologous recombination (HR) are the repair pathways responsible for double-strand break (DSB) repair in cells. Following detection of DSBs, NHEJ re-joins the broken DNA ends without homology requirements; this process requires the preparation of damaged ends using base removal and insertions of new bases, without ensuring sequence restoration around the break. Woodward *et al*. (40) and Lee *et al*. (29) used a two-step model of HR and NHEJ DSB repair to explain the origin of duplications involving the proteolipid protein 1 (*PLP1*) gene in patients with PMD.

However, Lee *et al*. (13), using high-resolution oligoarray CGH and breakpoint junction sequencing, detected complex rearrangements not observed before with *PLP1* duplications, likely reflecting technology limitations. The presence of complex rearrangements therein, similar to those observed in *MECP2* duplication patients described herein, led Lee *et al*. (13) to propose a new replication-based mechanism, FoSTeS to explain them. This is a modified mechanism based on long-distance template switching during DNA replication, proposed by Slack *et al*. (41) as a model to explain gene amplification induced under stress in *Escherichia coli*. According to Slack *et al*. (41), events are initiated by collapsed forks that are repaired by recombination, initiating replication. If this replication stalls, 3′-ends are released, and these might resume DNA synthesis on a different template, possibly on a second nearby replication fork. Microhomologies at the junctions or 'join point' reflect priming on the switched template strand, possibly later stalling, disengaging and switching template again. Serial replication fork and lagging strand invasion could occur several times (e.g. FoSTeS × 2, FoSTeS × 3, etc.) before resumption of replication on the original template by the tethered strand (13,36). Such long-distance template switches driven by microhomologies can explain the origin of duplications and triplications (if the switches occur to a fork that is replicating the region before the stalling point), inversions (switching to a template in the reverse orientation), as well as the occurrence of gaps and duplications interspersed by normal copy number (14). Slack *et al*. (41) showed that the proposed template switching events can involve a switch of a nascent lagging strand to a lagging strand template, but they also reported switching between leading and lagging strands producing DNA inversions.

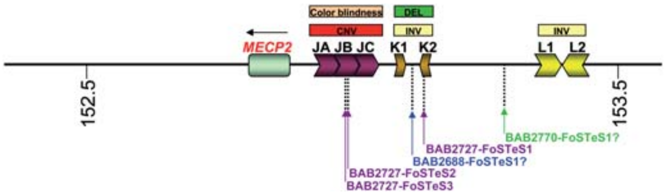


Figure 4. Alignment of the join points to the genomic locations of each FoSTeS in the context of regional LCRs. Positions are given relative to NCBI Build 35 for the X chromosome. Rearrangements reported previously in the literature are represented in different color rectangles above the figure (refer to the text for further details).

Hastings *et al*. (14) proposed a novel MMBIR pathway. This model hypothesizes that chromosomal structural changes result from aberrant repair of collapsed (broken) replication forks. In certain cell environments, the normal repair of collapsed forks by BIR is unavailable because of down-regulation of Rad51, the protein that allows 3′-ends at collapsed forks to re-invade the sister molecule. Under these circumstances, after the processing of the single dsDNA end generated from the collapsed fork, the single-strand 3′-tails will anneal with microhomology to any single-stranded DNA nearby, prime low processivity polymerization with multiple template switches and eventual re-establishment of processive replication. Hastings *et al*. (14) proposed MMBIR as a mechanistically specific general model applicable to all life forms that encompasses the FoSTeS model for human genomic rearrangements. Recent work in yeast shows that broken forks are the precursor lesions directly processed into segmental duplications (SDs) (42). Importantly, a functional Pol32, one of the subunits from the major replicative DNA polymerase in *S. cerevisae*, is a crucial requirement for the formation of SD in yeast, therefore, suggesting that genomic duplications can arise from DNA synthesis rather than unequal crossing-over (42).

Similar to what is observed with *MECP2*, duplications involving the *PLP1* gene are non-recurrent. Most *PLP1* duplication distal breakpoints group in the vicinity of complex LCRs, intervals to which complex rearrangements also map (13,29,43). Furthermore, the complex LCRs, wherein the *PLP1* duplication breakpoints group, are highly polymorphic in the population, a similar observation to that described here for *MECP2* rearrangements and summarized in Figure 4.

Distinguishing features at the *MECP2* duplication breakpoint junctions/join points

We were able to sequence five breakpoints/join points in four patients. One of them (BAB2629) had an apparent simple duplication by aCGH and showed no microhomology at the breakpoint junction suggesting NHEJ as the mechanism for formation. Two of them (BAB2688 and BAB2770 with apparent simple duplication by aCGH) had microhomologies of 3–4 nt, suggesting either the NHEJ mechanism or, alternatively, FoSTeS × 1. Remarkably, though, the occurrence of complex rearrangements in 27% of patients is strongly suggestive of the FoSTeS mechanism and that mechanism alone is sufficient to explain the observed junction complexity at the sequencing level. Under the FoSTeS hypothesis, breakpoint junctions might be better referred to as 'join points' because junctions arise by the juxtaposition of genomic sequences and microhomology reflecting the priming of DNA replication. The join point of patient BAB2727 (Fig. 3D and E), who has a complex duplication–triplication–duplication by aCGH, is consistent with the occurrence of FoSTeS a total of three times, including two switches of the template strand direction resulting in an inversion (FoSTeS × 3). The second breakpoint junction, deleted for 74 nt, can also be explained by the serial replication slippage model (SRS) (44). SRS assumes that the 3′-end of the nascent strand dissociates from the original template and invades other templates based on microhomology. SRS and FoSTeS share some general features. SRS can promote rearrangements between sequences at the same replication fork, and requires nascent strand disassociation presumably by a helicase. FoSTeS allows switching of replication forks, thus, the rearrangements can occur over longer distances, and initiates from a collapsed fork that generates a single dsDNA end with an available 3′ end for priming DNA synthesis on the switched template (36,45). Noteworthy, there is a region of perfect complementary match GCCA/CGGT just posterior or anterior to each GG direct repeats. The presence of this tract may approximate both direct repeats in the same strand and stimulate either FoSTeS or SRS to occur.

We examined for the occurrence of particular sequences flanking the breakpoint regions in both *MECP2* and *PLP1* duplication join points; such consistent presence might provide clues to the mechanism/enzymes involved on the rearrangements. In this regard, we found an increased frequency of the sequences 5′-CTG-3′/5′-CAG-3′ flanking the breakpoint junctions of the rearrangements. In prokaryotes, the 5′-CTG-3′ trinucleotide represents a minimal primase recognition site (the so-called G-site) (46,47). It signals the start of the synthesis of primer RNAs that initiate DNA replication leading and lagging-strand synthesis. In *Escherichia coli*, the high incidence of the 5′-CTG-3′ trinucleotide in or near the junctions produced by gene amplification induced under stress has been suggested as evidence for a relationship between such junction occurrences and the ends of the Okazaki fragments (41). The participation of the lagging strand and the involvement of the ends of the Okazaki fragments are features seen with the Long Distance Template Switching model proposed by Slack *et al*. (41). The observation of increased incidence of 5′-CTG-3′ trinucleotides near the junction of both *MECP2* and *PLP1* duplications raises the intriguing question as to whether such a motif might represent a *cis* acting sequence that may be a recognition site for proteins involved in priming DNA replication in eukaryotes.

The presence of repetitive sequences, especially *Alu* families, is a remarkable characteristic of the *MECP2* duplication breakpoints, despite the fact that they do not seem fundamental to the formation of the breakpoint junctions. Bauters *et al*. (12) reported a 2.5-fold enrichment of *Alu* elements at the breakpoint of *MECP2* duplications. In our cohort, we found *Alu* elements in 40% of the proximal/distal reference sequences involved at the breakpoints/join points (4 in 10) but none has an *Alu* present on both sides, a requirement for substrates participating in NAHR. Repetitive elements may have an important role in homologous recombination. Shaw and Lupski (48) showed that *Alu* elements could be used as substrates for NAHR in non-recurrent 17p11.2 chromosomal deletions causing Smith–Magenis Syndrome. The role of such sequences in other mechanisms, such as NHEJ and FoSTeS, is not well understood, but they are more frequent than expected if the breakpoints occurred randomly throughout the genome (48). Woodward *et al*. (40) found a high incidence of interspersed repetitive elements (69%) at the breakpoint junctions of patients with *PLP1* duplications. Toffolatti *et al*. (49) also found a high incidence of such elements at the breakpoints of deletions within the dystrophin gene. Importantly, *Alu* repeats have been postulated to play a role in the spread of SDs in humans. Bailey *et al*. (50) analyzed 9464 SD junctions and found a statistically significant enrichment of *Alu* elements near or within the junctions; and similar to our results, most of them showed *Alu* repeat sequences in only one side of the alignment. A similar result was obtained by Babcock *et al*. (51) when they analyzed the genomic sequence of known LCR22 genes and their duplicated derivatives. Thus, an association between *Alu* elements and rearrangements throughout the genome is consistently observed; however, its biological significance remains elusive.

The occurrence of similar characteristics at the breakpoint junctions of non-recurrent rearrangements in different genomic locations such as duplications involving *MECP2*, *PLP1* and dystrophin gene, with or without complexities, raises the intriguing hypothesis that the underlying mechanisms may be the same. Since neither NAHR nor NHEJ are sufficient to explain the more complex rearrangements, we propose that the mechanism is FoSTeS/MMBIR (13,14).

Bauters *et al*. (12) reported sequencing of three *MECP2* duplicated breakpoint junctions in three patients, all of them with 2–5 bp microhomologies at the juxtaposition region. Remarkably, two out of three patients have duplicated DNA stretches of the Xq28 distal region inserted amid the *MECP2* duplications. Bauters *et al*. (12) proposed BIR as the mechanism by which such complexities have arisen. Indeed, BIR and its extended version, Template Switching BIR (52), have been suggested as mechanisms underlying SDs and other structural changes in yeast and human. This mechanism is strongly RecA/Rad51-dependent and homology-dependent, and because of that it is usually an accurate process involving long lengths of homology between DNA sequences (reviewed by 14). This, however, does not fit with the microhomology junctions (usually between 2 and 6 nt) found at the breakpoint junctions of the rearrangements generated by MMBIR/FoSTeS (13). Therefore, we propose that the presence of short microhomologies, such as those reported by Bauters *et al*. (11), is rather consistent with a FoSTeS/MMBIR mechanism.

In conclusion, the presence of complex rearrangements, characteristically observed in rearrangements generated by FoSTeS/MMBIR, imply that this mechanism underlies many *MECP2* rearrangements. Recently, Bi *et al*. (53) identified patients carrying submicroscopic duplication in 17p13.3 involving the *PAFAH1B1* (*LIS1*) and/or the *YWHAE* (*14-3-3ϵ*) genes; three out of seven patients with complex rearrangements are also suggestive of FoSTeS as a mechanism for formation. Likewise, the FoSTeS mechanism can potentially explain some of the complex LCR-associated rearrangements observed in the literature, such as a patient with aplasia of Mullerian ducts carrying a complex rearrangement at the 22q11.21 region (54), and the patients carrying duplication/triplication/inversion involving the olfactory receptors/

defensin repeats at 8p23.1 (55). The LCRs associated with these rearrangements, even though not mediating FoSTeS directly, might be able to bring the replication forks together and facilitate the replication fork-switching event (36). Alternatively, the presence of LCRs in the *MECP2* vicinity can generate an unstable DNA structure, which can induce DNA strand lesions and collapsed forks that facilitate FoSTeS/MMBIR during the replication process. Additionally, the local presence of SINEs, LINEs or the enrichment of specific sequences, such as the 5′-CTG-3′/5′-CAG-3′, may influence the occurrence of FoSTeS at a given location, reinforcing the importance of genomic architecture for this mechanism.

MATERIALS AND METHODS

Samples

Peripheral blood samples from patients were submitted for clinical testing to the Baylor Medical Genetics Laboratories for either *MECP2* deletion/duplication analysis or CMA. The methodology used to screen the samples (quantitative DNA methods and BAC-based chromosomal microarray analysis) has been described (3). All the duplications were confirmed by either aCGH platform [such as a commercial aCGH for patient BAB2797 or aCGH X-chromosome tiling-path (56) for patient BAB2688] or by another methodology such as FISH or MLPA as described (3): BAB2616, BAB2617, BAB2618, BAB2619, BAB2620, BAB2622, BAB2623, BAB2624, BAB2625, BAB2626, BAB2627, BAB2628, BAB2629, BAB2769, BAB2770, BAB2771, BAB2796, BAB2798, BAB2799, BAB2800, BAB2801, BAB2802; FISH and/or MLPA as described (20): BAB2681, BAB2682 and BAB2683; Agilent 244K Oligonucleotide array CGH: BAB2727. Microduplications in patients BAB2666 and BAB2772 were identified by Signature Genomics Laboratories and confirmed by FISH or MLPA. We identified 21 males with increased *MECP2* gene copy number. Three laboratories (one from the Department of Pediatrics at Baylor College of Medicine, Houston, TX, USA, and two from genetics institutes in Poland and Brazil) identified nine additional patients based on clinical characteristics associated with *MECP2* duplication and further *MECP2* screening analysis. Genomic DNA was extracted from blood leukocytes using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN), and DNA concentration was measured using the NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Rockland, DE). The screening protocol was approved by Baylor College of Medicine Institutional Review Board.

Parental analysis was done by FISH or MLPA, when DNA was available. Most of the patients (21 out of 30) inherited the duplication from their mother. Two duplications were *de novo* (BAB2618 and BAB2620). In patient BAB2620 the duplication inserted into the Y-chromosome, but the father is not a carrier. In seven cases parental DNA was not available (BAB2616, BAB2619, BAB2624, BAB2625, BAB2626, BAB2628, BAB2799). Five mothers of patients carrying complexities were analyzed by oligonucleotide array CGH. The complexities were inherited from the mother for all patients, including the complex junction found in BAB2727, as confirmed by DNA sequencing.

Probes design

We designed a tiling-path oligonucleotide microarray spanning 4 Mb around the *MECP2* region at Xq28. The custom 4X44K Agilent Technologies microarray (#G4426A) was designed using the Agilent website (http://earray.chem.agilent.com/earray/). We selected 8323 probes covering the ChrX: 150 400 000–154 400 000 (NCBI build 35), including the *MECP2* gene, which represents an average resolution of one probe per 500 bp. Probe labeling and hybridization were performed following the manufacturer's protocol (Agilent Oligonucleotide Array based CGH for Genomic DNA analysis, version 4.0 plus the 4X44K complementary protocol with modifications unique to the four-pack format). Briefly, 1.5 μg of genomic male/female reference and patient DNA were digested with *Alu*I (5 U) and *Rsa*I (5 U) (Promega) for 2 h at 37°C. Digestions were verified by agarose gel electrophoresis. Labeling reactions with Cy5-dUTP for patient DNA and Cy3-dUTP for male/female reference DNA were performed according to the manufacturer's instructions (Agilent Genomic DNA Labeling Kit Plus, 5188–5309). Individual dye-labeled reference and patient samples were purified using Microcon Ym-30 filters (Millipore Corporation). DNA yield was determined using a NanoDrop ND-1000 UV-VIS spectrophotometer. Each dye-labeled patient and gender-matched reference DNA was combined with 5 μg human Cot-1 DNA (Invitrogen Corporation), Agilent Blocking Agent, and Agilent hybridization buffer (#5188–5220). These mixtures were denatured at 95°C for 3 min, pre-incubated at 37°C for 30 min, and hybridized to the array in a hybridization chamber (Agilent Technologies) for 40 h at 65°C in a rotating oven (Agilent Technologies). Array slides were washed using Agilent Wash solutions 1 and 2 (#5188–5226), Acetonitrile (Sigma-Aldrich), and Stabilization and Drying Solution (#5185–5979), according to the manufacturer's instructions.

Slides were scanned on a GenePix 4000B Microarray Scanner (Axon Instruments). Images were analyzed and data were extracted, background subtracted, and normalized using Agilent Feature Extraction Software A.7.5.1. These data were subsequently imported into array CGH Analytics software v3.1.28 (Agilent Technologies). The genomic copy number was defined by the analysis of the normalized $\log_2$ (Cy5/Cy3) ratio average of the CGH signal, amid a 5 kb window. Regions that reached a threshold of 0.6 were considered as duplication, whereas thresholds of 1.2 were considered as triplication.

Based on the high-resolution data, sets of primers were designed at the apparent boundaries of each duplicated segment(s) (as determined by a transition from normal copy number to gain in copy number) for all patients but BAB2796 and BAB2797, and used for long-range PCR amplification as described below (primers facing outward, inward and in the same direction).

Long-range PCR amplification using outward-facing primers

Assuming tandem duplications, and using outward-facing primers (with respect to the reference sequence), long-range PCR was performed using TaKaRa LA *Taq* (TAKARA Bio) or Phusion high-fidelity polymerase with HF buffer (Finnzymes Oy). A 50 μl PCR analysis was performed using 2.5 U TaKaRa with 1× PCR buffer, 0.4 mM dNTP, 10 pmol of each primer, 1 μl DMSO and 200 ng DNA template. The PCR conditions were as follows: 98°C for 30 s, 32 cycles at 94°C for 1 min, 65°C for 20 s and 68°C for 20 min, followed by 68°C for 10 min. The Phusion polymerase reaction conditions were as follows: 98°C for 30 s, followed by 12 cycles of 98°C for 10 s, 67°C for 30 s (−1°C/cycle) and 72°C for 30 s per kb (initial attempt: 15 kb or 7 min and 30 s), followed by 30 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 30 s per kb finalizing with 72°C for 10 min. Amplification products were electrophoresed on 0.8–1% agarose gels, PCR products from any unique bands were purified, using a gel extraction kit, and were analyzed by DNA sequencing.

To investigate the possibility of small inversions at breakpoint junctions, PCR amplification was attempted using forward and reverse primers at the apparent boundaries of each duplicated segment(s) using both pairs of primers facing in only one direction and pairs of primers facing inward.

Bioinformatic and sequence analysis

The human reference genome sequence for the *MECP2* gene and flanking regions covering 4 Mb of repeat-masked DNA sequence was analyzed using NCBI build 35 at the Genome Browser Gateway at the University of California Santa Cruz (http://genome.ucsc.edu/cgi-bin/hgGateway) and Blast 2 browser with default parameters (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). We aimed to define the genomic architecture in the Xq28 region that potentially facilitates the *MECP2* duplication events. Specifically, this analysis was performed by dividing the entire analyzed region into ~1 Mb segments; the sequence for each ~1 Mb segment was then compared with itself and with each of the other segments by BLAST 2 to identify LCR sequence and genomic architecture within the analyzed region. CG content was calculated using the web-based software MCLAB (http://www.mclab.com/toolbox/oligo_calculator2.jsp). DNA motifs were determined using The Sequence Manipulation suite (http://www.bioinformatics.org/sms2/).

Statistical analysis

We performed 10 000 replicate Monte Carlo simulations to evaluate the non-random distribution of breakpoints considering a random allocation of segments and requiring them to overlap the *MECP2* gene. To perform the simulation, we first determined each segment's size. We then simulated locations of segments in our cohort as follows: we randomly chose a start position for each segment by drawing uniformly from the values from the interval (*MECP2*-start-S, *MECP2*-stop), where S= segment size−1. This procedure was repeated for each segment adjusting the maximal start and stop position accordingly. Each simulation run determined a randomly distributed set of segment locations with the same distribution of segment sizes as for our observed data and with the restriction that each segment would overlap *MECP2*. To summarize these Monte Carlo outcomes, we determined two summary statistics for each of the simulation runs: the variance of the segment locations and the mean segment distal breakpoint.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors are grateful to Philip J. Hastings for his valuable critical reviews. We thank Marjorie A. Withers for her outstanding technical support.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Thrasher Research Fund (NR-0017 to C.M.B.C.); the National Institute of Child Health and Human Development (NICHD) and Mental Retardation and Developmental Disabilities Research Center (HD024064); Texas Children's Hospital GCRC grant (M01RR00188). Additional support was obtained from Conselho Conselho Nacional de Desenvolvimento Científico e Tecnológico of Brazil (CNPq) (to C.M.B.C.); National Institutes of Health/National Institute of Neurological Disorders and Stroke (NIH/NINDS) (T32 NS43124 and 1K08NS062711-01 to M.B.R); Polish Ministry of Education and Science grant PBZ KBN-122/PO5/2004/01-9. The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, NINDS, or NIH.

REFERENCES

1. Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K. and Epplen, J.T. (2005) Submicroscopic duplication in Xq28 causes increased expression of the *MECP2* gene in a boy with severe mental retardation and features of Rett syndrome. *J. Med. Genet.*, **42**, e12.
2. Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gecz, J. *et al.* (2005) Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.*, **77**, 442–453.
3. del Gaudio, D., Fang, P., Scaglia, F., Ward, P.A., Craigen, W.J., Glaze, D.G., Neul, J.L., Patel, A., Lee, J.A., Irons, M. *et al.* (2006) Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.*, **8**, 784–792.
4. Friez, M.J., Jones, J.R., Clarkson, K., Lubs, H., Abuelo, D., Bier, J.A., Pai, S., Simensen, R., Williams, C., Giampietro, P.F. *et al.* (2006) Recurrent infections, hypotonia, and mental retardation caused by duplication of *MECP2* and adjacent region in Xq28. *Pediatrics*, **118**, e1687–e1695.
5. Gu, W. and Lupski, J.R. (2008) CNV and nervous system disease - what's new? *Cytogenet. Genome Res.*, **123**, 54–64.

6. Lee, J.A. and Lupski, J.R. (2006) Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, **52**, 103–121.
7. Lubs, H., Abidi, F., Bier, J.A., Abuelo, D., Ouzts, L., Voeller, K., Fennell, E., Stevenson, R.E., Schwartz, C.E. and Arena, F. (1999) XLMR syndrome characterized by multiple respiratory infections, hypertelorism, severe CNS deterioration and early death localizes to distal Xq28. *Am. J. Med. Genet.*, **85**, 243–248.
8. Ariani, F., Mari, F., Pescucci, C., Longo, I., Bruttini, M., Meloni, I., Hayek, G., Rocchi, R., Zappella, M. and Renieri, A. (2004) Real-time quantitative PCR as a routine method for screening large rearrangements in Rett syndrome: Report of one case of *MECP2* deletion and one case of *MECP2* duplication. *Hum. Mutat.*, **24**, 172–177.
9. Garcia, C.A., Malamut, R.E., England, J.D., Parry, G.S., Liu, P. and Lupski, J.R. (1995) Clinical variability in two pairs of identical twins with the Charcot-Marie-Tooth disease type 1A duplication. *Neurology*, **45**, 2090–2093.
10. Ramocki, M.B., Peters, S.U., Tavyev, J.Y., Zhang, F., Carvalho, C.M.B., Schaaf, C.P., Richman, R., Fang, P., Glaze, D.G., Lupski, J.R. *et al.* (2009) Autism and other Neuropsychiatric symptoms are prevalent in individuals with the *MECP2* duplication syndrome. *Ann Neurol.*, in press.
11. Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.*, **18**, 74–82.
12. Bauters, M., Van Esch, H., Friez, M.J., Boespflug-Tanguy, O., Zenker, M., Vianna-Morgante, A.M., Rosenberg, C., Ignatius, J., Raynaud, M., Hollanders, K. *et al.* (2008) Nonrecurrent *MECP2* duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. *Genome Res.*, **18**, 847–858.
13. 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.
14. 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.
15. Cheung, S.W., Shaw, C.A., Yu, W., Li, J., Ou, Z., Patel, A., Yatsenko, S.A., Cooper, M.L., Furman, P., Stankiewicz, P. *et al.* (2005) Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet. Med.*, **7**, 422–432.
16. 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.
17. Kornreich, R., Bishop, D.F. and Desnick, R.J. (1990) Alpha-galactosidase A gene rearrangements causing Fabry disease. Identification of short direct repeats at breakpoints in an *Alu*-rich gene. *J. Biol. Chem.*, **265**, 9319–9326.
18. 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 5,380 cases. *Am. J. Med. Genet. A*, **146A**, 2242–2251.
19. Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., Sweatt, J.D. and Zoghbi, H.Y. (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.*, **13**, 2679–2689.
20. Smyk, M., Obersztyn, E., Nowakowska, B., Nawara, M., Cheung, S.W., Mazurczak, T., Stankiewicz, P. and Bocian, E. (2008) Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.*, **147B**, 799–806.
21. Kurotaki, N., Shen, J.J., Touyama, M., Kondoh, T., Visser, R., Ozaki, T., Nishimoto, J., Shiihara, T., Uetake, K., Makita, Y. *et al.* (2005) Phenotypic consequences of genetic variation at hemizygous alleles: Sotos syndrome is a contiguous gene syndrome incorporating coagulation factor twelve (FXII) deficiency. *Genet. Med.*, **7**, 479–483.
22. Somerville, M.J., Mervis, C.B., Young, E.J., Seo, E.J., del Campo, M., Bamforth, S., Peregrine, E., Loo, W., Lilley, M., Perez-Jurado, L.A. *et al.* (2005) Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N. Engl. J. Med.*, **353**, 1694–1701.
23. Berg, J.S., Brunetti-Pierri, N., Peters, S.U., Kang, S.H., Fong, C.T., Salamone, J., Freedenberg, D., Hannig, V.L., Prock, L.A., Miller, D.T. *et al.* (2007) Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams-Beuren syndrome region. *Genet. Med.*, **9**, 427–441.
24. Depienne, C., Heron, D., Betancur, C., Benyahia, B., Trouillard, O., Bouteiller, D., Verloes, A., LeGuern, E., Leboyer, M. and Brice, A. (2007) Autism, language delay and mental retardation in a patient with 7q11 duplication. *J. Med. Genet.*, **44**, 452–458.
25. Torniero, C., Dalla Bernardina, B., Novara, F., Cerini, R., Bonaglia, C., Pramparo, T., Ciccone, R., Guerrini, R. and Zuffardi, O. (2008) Dysmorphic features, simplified gyral pattern and 7q11.23 duplication reciprocal to the Williams-Beuren deletion. *Eur. J. Hum. Genet.*, **16**, 880–887.
26. Potocki, L., Bi, W., Treadwell-Deering, D., Carvalho, C.M., Eifert, A., Friedman, E.M., Glaze, D., Krull, K., Lee, J.A., Lewis, R.A. *et al.* (2007) Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am. J. Hum. Genet.*, **80**, 633–649.
27. Mefford, H.C., Clauin, S., Sharp, A.J., Moller, R.S., Ullmann, R., Kapur, R., Pinkel, D., Cooper, G.M., Ventura, M., Ropers, H.H. *et al.* (2007) Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am. J. Hum. Genet.*, **81**, 1057–1069.
28. Wentzel, C., Fernstrom, M., Ohrner, Y., Anneren, G. and Thuresson, A.C. (2008) Clinical variability of the 22q11.2 duplication syndrome. *Eur. J. Med. Genet.*, **51**, 501–510.
29. Lee, J.A., Inoue, K., Cheung, S.W., Shaw, C.A., Stankiewicz, P. and Lupski, J.R. (2006) Role of genomic architecture in *PLP1* duplication causing Pelizaeus-Merzbacher disease. *Hum. Mol. Genet.*, **15**, 2250–2265.
30. Nathans, J., Piantanida, T.P., Eddy, R.L., Shows, T.B. and Hogness, D.S. (1986) Molecular genetics of inherited variation in human color vision. *Science*, **232**, 203–210.
31. Deeb, S.S. (2005) The molecular basis of variation in human color vision. *Clin. Genet.*, **67**, 369–377.
32. Small, K., Iber, J. and Warren, S.T. (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat. Genet.*, **16**, 96–99.
33. Small, K. and Warren, S.T. (1998) Emerin deletions occurring on both Xq28 inversion backgrounds. *Hum. Mol. Genet.*, **7**, 135–139.
34. Smahi, A., Courtois, G., Vabres, P., Yamaoka, S., Heuertz, S., Munnich, A., Israel, A., Heiss, N.S., Klauck, S.M., Kioschis, P. *et al.* (2000) Genomic rearrangement in *NEMO* impairs NF-kappaB activation and is a cause of incontinentia pigmenti. The International Incontinentia Pigmenti (IP) Consortium. *Nature*, **405**, 466–472.
35. Aradhya, S., Bardaro, T., Galgoczy, P., Yamagata, T., Esposito, T., Patlan, H., Ciccodicola, A., Munnich, A., Kenwrick, S., Platzer, M. *et al.* (2001) Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the *NEMO* and *LAGE2* genes. *Hum. Mol. Genet.*, **10**, 2557–2567.
36. Gu, W., Zhang, F. and Lupski, J.R. (2008) Mechanisms for human genomic rearrangements. *Pathogenetics*, **1**, 4.
37. Padiath, Q.S., Saigoh, K., Schiffmann, R., Asahara, H., Yamada, T., Koeppen, A., Hogan, K., Ptacek, L.J. and Fu, Y.H. (2006) Lamin B1 duplications cause autosomal dominant leukodystrophy. *Nat. Genet.*, **38**, 1114–1123.
38. Stankiewicz, P., Shaw, C.J., Dapper, J.D., Wakui, K., Shaffer, L.G., Withers, M., Elizondo, L., Park, S.S. and Lupski, J.R. (2003) Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am. J. Hum. Genet.*, **72**, 1101–1116.
39. Bacolla, A., Jaworski, A., Larson, J.E., Jakupciak, J.P., Chuzhanova, N., Abeysinghe, S.S., O'Connell, C.D., Cooper, D.N. and Wells, R.D. (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc. Natl Acad. Sci. USA*, **101**, 14162–14167.
40. Woodward, K.J., Cundall, M., Sperle, K., Sistermans, E.A., Ross, M., Howell, G., Gribble, S.M., Burford, D.C., Carter, N.P., Hobson, D.L. *et al.* (2005) Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination. *Am. J. Hum. Genet.*, **77**, 966–987.
41. Slack, A., Thornton, P.C., Magner, D.B., Rosenberg, S.M. and Hastings, P.J. (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.*, **2**, e48.
42. Payen, C., Koszul, R., Dujon, B. and Fischer, G. (2008) Segmental duplications arise from Pol32-dependent repair of broken forks through two alternative replication-based mechanisms. *PLoS Genet.*, **4**, e1000175.

43. 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.
44. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. and Cooper, D.N. (2005) Complex gene rearrangements caused by serial replication slippage. *Hum. Mutat.*, **26**, 125–134.
45. Chen, J.M., Chuzhanova, N., Stenson, P.D., Ferec, C. and Cooper, D.N. (2005) Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. *Hum. Mutat.*, **26**, 362–373.
46. Hiasa, H., Sakai, H., Komano, T. and Godson, G.N. (1990) Structural features of the priming signal recognized by primase: mutational analysis of the phage G4 origin of complementary DNA strand synthesis. *Nucleic Acids. Res.*, **18**, 4825–4831.
47. Tanaka, K., Rogi, T., Hiasa, H., Miao, D.M., Honda, Y., Nomura, N., Sakai, H. and Komano, T. (1994) Comparative analysis of functional and structural features in the primase-dependent priming signals, G sites, from phages and plasmids. *J. Bacteriol.*, **176**, 3606–3613.
48. Shaw, C.J. and Lupski, J.R. (2005) Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum. Genet.*, **116**, 1–7.
49. Toffolatti, L., Cardazzo, B., Nobile, C., Danieli, G.A., Gualandi, F., Muntoni, F., Abbs, S., Zanetti, P., Angelini, C., Ferlini, A. *et al.* (2002) Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics*, **80**, 523–530.
50. Bailey, J.A., Liu, G. and Eichler, E.E. (2003) An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.*, **73**, 823–834.
51. Babcock, M., Pavlicek, A., Spiteri, E., Kashork, C.D., Ioshikhes, I., Shaffer, L.G., Jurka, J. and Morrow, B.E. (2003) Shuffling of genes within low-copy repeats on 22q11 (LCR22) by *Alu*-mediated recombination events during evolution. *Genome Res.*, **13**, 2519–2532.
52. Smith, C.E., Llorente, B. and Symington, L.S. (2007) Template switching during break-induced replication. *Nature*, **447**, 102–105.
53. 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.
54. Cheroki, C., Krepischi-Santos, A.C., Szuhai, K., Brenner, V., Kim, C.A., Otto, P.A. and Rosenberg, C. (2008) Genomic imbalances associated with mullerian aplasia. *J. Med. Genet.*, **45**, 228–232.
55. Barber, J.C., Maloney, V.K., Huang, S., Bunyan, D.J., Cresswell, L., Kinning, E., Benson, A., Cheetham, T., Wyllie, J., Lynch, S.A. *et al.* (2008) 8p23.1 duplication syndrome; a novel genomic condition with unexpected complexity revealed by array CGH. *Eur. J. Hum. Genet.*, **16**, 18–27.
56. Veltman, J.A., Yntema, H.G., Lugtenberg, D., Arts, H., Briault, S., Huys, E.H., Osoegawa, K., de Jong, P., Brunner, H.G., Geurts van Kessel, A. *et al.* (2004) High resolution profiling of X chromosomal aberrations by array comparative genomic hybridisation. *J. Med. Genet.*, **41**, 425–432.