European Journal of Human Genetics (2012) 20, 1096–1100

www.nature.com/ejhg

SHORT REPORT

Comprehensive oligonucleotide array-comparative genomic hybridization analysis: new insights into the molecular pathology of the *DMD* gene

Aliya Ishmukhametova*[,1,2], Philippe Khau Van Kien[3,7], Déborah Méchin[3], Delphine Thorel[3], Marie-Claire Vincent[3], François Rivier[4,5], Christine Coubes[6], Véronique Humbertclaude[1,2], Mireille Claustres[1,2,3] and Sylvie Tuffery-Giraud*[,1,2]

We report on the effectiveness of a custom-designed oligonucleotide-based comparative genomic hybridization microarray (array-CGH) to interrogate copy number across the entire 2.2-Mb genomic region of the *DMD* gene and its applicability in diagnosis. The high-resolution array-CGH, we developed, successfully detected a series of 42 previously characterized large rearrangements of various size, localization and type (simple or complex deletions, duplications, triplications) and known intronic CNVs/Indels. Moreover, the technique succeeded in identifying a small duplication of only 191 bp in one patient previously negative for *DMD* mutation. Accurate intronic breakpoints localization by the technique enabled subsequent junction fragments identification by sequencing in 86% of cases (all deletion cases and 62.5% of duplication cases). Sequence examination of the junctions supports a role of microhomology-mediated processes in the occurrence of *DMD* large rearrangements. In addition, the precise knowledge of the sequence context at the breakpoints and analysis of the resulting consequences on maturation of pre-mRNA contribute to elucidating the cause of discrepancies in phenotype/genotype correlations in some patients. Thereby, the array-CGH proved to be a highly efficient and reliable diagnostic tool, and the new data it provides will have many potential implications in both, clinics and research.
European Journal of Human Genetics **20,** 1096–1100; doi:10.1038/ejhg.2012.51; published online 18 April 2012

Keywords: duchenne muscular dystrophy; *DMD* gene; large rearrangements; comparative genomic hybridization microarray; diagnostic methods

INTRODUCTION

Mutations in the huge human Duchenne muscular dystrophy gene (*DMD*; MIM#300377), which encodes the 427-kDa muscular dystrophin protein isoform, result in dystrophinopathies. There is no simple relationship between the type or the size of the mutations in the *DMD* gene and the severity of phenotype, but the reading-frame rule holds true for 96% of Duchenne Muscular Dystrophy (DMD; OMIM#310200) and 93% of Becker Muscular Dystrophy (BMD; OMIM#300376) cases.[1] As the majority of mutations in *DMD* are large deletions and duplications, several dosage-sensitive quantitative methods mainly focused on discovering mutations in the coding regions of the gene are commonly used.[2] Here we describe the introduction into the current diagnostic practice and validation of a high-resolution custom-designed Comparative Genomic Hybridization array (array-CGH) enabling to interrogate the entire 2.2 Mb genomic region of the *DMD* gene for copy number variations. A panel of *DMD* rearrangements of various type, size and localization was selected, some of which did not conform to the reading-frame rule. Eight mutation-negative patients were also analyzed. We specifically assessed the ability of the custom-designed array-CGH to detect rearrangements within the *DMD* gene and the potential contribution of this method to the identification of breakpoint/junction sequences.

MATERIALS AND METHODS

Patients

Based on the data available in the clinical and molecular databases maintained in our laboratory,[1] we selected 50 patients' DNA from previously collected 550 non-related DMD/BMD families (French Ministry of Health, collection ID: DC-2008-417) dividing them into three groups (Table 1; Supplementary Table 1). All patients provided an agreement for further analysis on informed consent form. Eight relatives from four unrelated families were included to test the reproducibility of the technique. DNA samples of Marfan patients (two males and one female) with previously identified large rearrangements in the fibrillin type 1 (*FBN1*) gene[3] and no familial history of neuromuscular disorders, served as gender-matched internal positive controls for array-CGH.

Array-CGH

The Roche NimbleGen (Roche NimbleGen, Inc., Madison, WI, USA) custom-designed 12 × 135 K format contained 3440 exonic *DMD* probes, overlapped

[1]Université Montpellier 1, UFR médecine, Montpellier, France; [2]INSERM, U827, Montpellier, France; [3]CHU Montpellier, Hôpital Arnaud de Villeneuve, Laboratoire de Génétique Moléculaire, Montpellier, France; [4]CHRU de Montpellier, Neuropédiatrie CR Maladies Neuromusculaires, Montpellier, France; [5]U1046, INSERM, Université Montpellier 1 et 2, Montpellier, France; [6]CHU Montpellier, Hôpital Arnaud de Villeneuve, Service de Génétique Médicale, Montpellier, France
[7]Present address: CHU de Nîmes, Laboratoire de Cytologie Clinique et Cytogénétique, Nîmes, France.
*Correspondence: Dr A Ishmukhametova or Dr S Tuffery-Giraud, INSERM, U827, IURC, Institut Universitaire de Recherche Clinique, 641 Avenue du Doyen Giraud, 34093 Montpellier Cedex 5, France. Tel: +33 4 11 75 98 79; Fax: +33 4 11 75 98 82; E-mail: aliya.ishmukhametova@inserm.fr or sylvie.tuffery@inserm.fr
Received 7 November 2011; revised 27 January 2012; accepted 21 February 2012; published online 18 April 2012

Table 1 Data on the patients included in the study

	Large rearrangements			*Complex rearrangements*	*Unknown mutations*	
	Group 1[a]			*Group 2*[a]	*Group 3*[a]	
	n = *35*			n = *7*	n = *8*	
	Subgroup 1-1	*Subgroup 1-2*	*Subgroup 1-3 exception*			
	major hot spot	*minor hot spot*	*to the reading-frame rule*			*Total*
Criteria	n = *16*	n = *10*	n = *9*			n = *50*
Gender						
Male	16	8	8	7	5	44
Female	0	2	1	0	3	6
Phenotype[b]						
DMD	14	2	5	5	2	28
BMD	0	6	3	2	2	13
IMD	2	0	0	0	1	3
Symptomatic/potential carrier	0	2	1	0	3	6
Patients with rearrangements detected by array-CGH						
Deletion/Double deletion*	14	8	6	1*	0	29
Duplication/Double duplication**/Triplication***	2	2	3	2**/4***	1	10/4***
Total	16	10	9	7	1	43

[a]Groups in the study: Group 1, previously identified large rearrangements in *DMD*: subgroup 1-1, mutations in the major hot spot; subgroup 1-2, mutations in the minor hot spot; subgroup 1-3, other exceptions to the reading-frame rule. Group 2, previously identified complex rearrangements, more than one alteration found in *DMD* (*double deletion, **double duplication, ***triplication). Group 3, patients with undetected defect in *DMD* (by multiplex-PCR, semi-quantitative-PCR, Multiplex ligation-dependent probe amplification, whole-exon sequencing).
[b]Phenotype: DMD, Duchenne phenotype; BMD, Becker phenotype; IMD, intermediate phenotype; Symptomatic carrier, female with clinical symptoms of any severity; potential carrier, female with elevated creatine kinase levels.

and shifted on an average of 10 bases, and 19 294 intronic *DMD* probes interspersed by 100 bp on average. Slides were scanned by InnoScan 900 A (Inopsys, Toulouse, France) and analyzed using the CGH-segMNT algorithm of NimbleScan version 2.5 software (Roche NimbleGen, Inc.). The predicted breakpoint location was defined by the positions of the last and first probes with normal unaveraged value of $\log_2$-ratio upstream and downstream from the corresponding aberration.

PCR/Sequencing across the breakpoints

PCR primers were designed in an average distance of 0.7 kb upstream and downstream of each predicted junction and amplifications were performed using standard protocols of Promega Master Mix (Promega Corporation, Madison, WI, USA), Phusion Hot Start High-Fidelity DNA polymerase (Finnzymes Oy, Espoo, Finland) or LongRange PCR kit (Qiagen, Courtaboeuf, France). When obtained, amplified junction fragments were sequenced using the Big Dye terminator version 1.1 Cycle Sequencing Kit (Applied Biosystems, Courtaboeuf, France).

Bioinformatic analysis

UCSC Genome Browser (http://genome.ucsc.edu) and BLAST program (http://blast.ncbi.nlm.nih.gov/) were used for the mapping the particular motifs surrounding the junctions. The Position Converter Interface in Mutalyzer 2.0 β-8 was applied to convert chromosomal positions of Mar.2006 NCBI Build 36.1/hg18 (RefSeq NC_000023.9) to transcript orientated positions.[4]

RESULTS

Array-CGH: sensitivity and reproducibility

The array-CGH analysis confirmed all 35 large deletions and duplications and 7 complex contiguous and non-contiguous rearrangements (Figure 1) previously identified in patients from group 1 and 2, giving a 100% detection rate (Table 1, Supplementary Table 2). The method was also able to detect a hemizygous 191-bp duplication spanning intron 19–exon 20 junction in one DMD patient (D87, group 3), which had escaped detection both by MLPA (due to a probe-target mismatch at the 3′ end of the MLPA probe) and by genomic sequencing (due to the parameter settings for the sequencing analysis software used).

Independently derived data from eight tested relatives from four different families and from duplicate experiments performed for 12 patients showed that the reproducibility of our array-CGH platform was high, with an average accuracy in the breakpoint localization of about 700 bp (range 0–4 kb) (Supplementary Table 2). Apart from the large rearrangements already known and correctly predicted by array-CGH, we noticed some experimental artefacts (ie data not confirmed on independent and/or averaged results of array-CGH) in the vicinity of exons 13, 17, 45, Dp140, Dp71 and intron 67.

Sequence characteristics at the breakpoints

The accuracy of array-CGH breakpoint mapping enabled us to successfully design primers and obtain the breakpoint sequences in 86% of the patients (37 out of the 42 patients from group 1 and 2, and 1 patient from group 3) (Table 2). Taking into account that complex rearrangements would have more than one aberrant junction in a single patient and excluding familial cases with similar rearrangements, we expected to find 45 different junction sequences. In all, 33 of them (73.3%) were correctly identified: all simple deletions and triplications cases (25/25 and 2/2, respectively), 62.5% (5/8) of simple duplication but only 20% (2/10) of complex rearrangement junctions. All breakpoints in unrelated patients were unique with no clustering, even in the frequently rearranged introns 2, 7 or 44. Microhomology up to 9 bp was evidenced in 60.6% of the preserved ends of the rearrangement breakpoints (20 cases out of 33). In nine other cases, insertions up to 25 bp represented mostly the small duplicated parts of sequences surrounding the junctions and

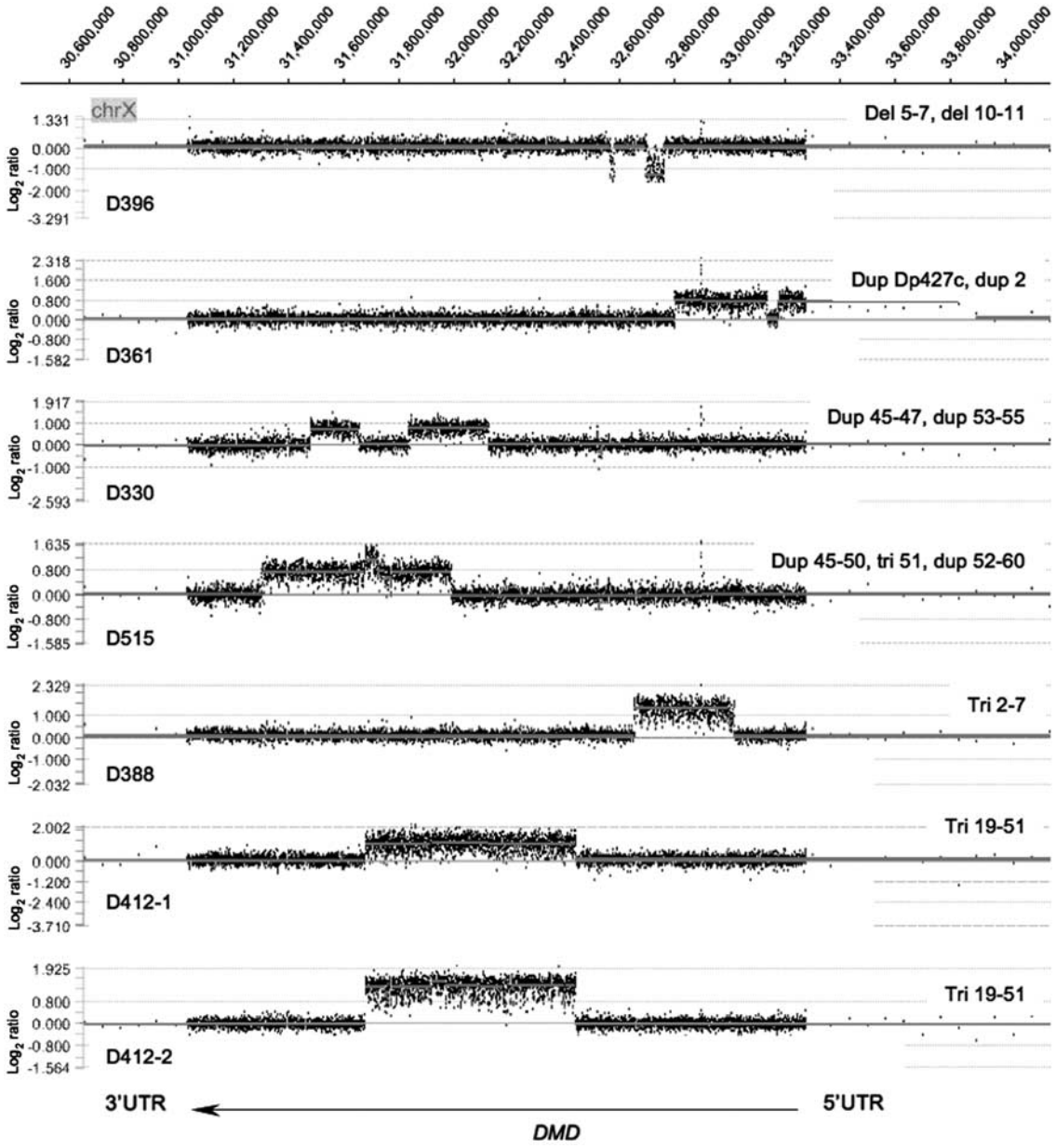

Figure 1 Array-CGH results in patients with complex rearrangements in the *DMD* gene. Array-CGH $\log_2$-ratio profiles of patients with *DMD* complex rearrangements analyzed with segMNT algorithm by NimbleScan ver.2.5 and displayed on SignalMap ver.1.9 software (Roche NimbleGen, Inc.): data for the signal of each probe were plotted indicating gain or loss of material on the y-axis *versus* X-chromosomal position of the probes on the x-axis accordingly to the GenBank NC_000023.9 and the Human Genome reference sequence Mar.2006 NCBI Build 36.1/hg18 (http://genome.ucsc.edu/). The *DMD* gene coordinates on the X-chromosome are indicated at the top (RefSeq NC_000023.9), with exon 1 to 79 from right to left. 5′UTR/3′UTR, *DMD* 5′/3′ untranslated regions; del, deletion; dup, duplication; tri, triplication; involved exons are indicated.

only four cases did not show any homology. Overall, repetitive sequences of different classes, such as LINE, LTR, SINE and DNA, were represented in 32 out of 66 junction ends (48.5%), but there was a marked difference of their involvement in aberrations of exons 3 to 7 (64.3%; 9/14) compared with mutations in the major hot spot (39.3%; 11/28). No extensive homology was visible even when repetitive elements met on the both sides of the junction with one exception: 90% homology of about 400 bp of two LINE:L1 elements situated on complementary strands was noted in a distance of 360 bp and 535 bp from the proximal and distal ends of the exons 48–50 deletion junction (D55), respectively (Supplementary Figure 1).

Five out-of-frame deletions of exons 3 to 7 were associated with BMD phenotype but no specific molecular features were found to explain the phenotype/genotype discrepancy in these patients. On the other hand, the two out-of-frame duplications of exons 3 to 7 were confirmed to be in tandem and corresponded to the severe DMD phenotype. Among the nine other cases with exception to the reading-frame rule (5 DMD, 1 symptomatic female and 3 BMD), our findings brought the explanation of severe DMD phenotype in one patient (D145) carrying an in-frame deletion of exons 35 to 42. Sequencing across the junction revealed a complex pattern on genomic level with putative splice sites in a suitable position to explain the

Table 2 Breakpoint findings

			Proximal breakpoint (toward the 5′ UTR of the gene, upstream)			Distal breakpoint (toward 3′ UTR of the gene, downstream)				
Breakpoint	*Family ID*[a]	*Mutation*	*Intron*	*Chromosomal position*[b]	*Sequence involved in rearrangement*	*Intron*	*Chromosomal position*[b]	*Sequence involved in rearrangement*	*Microhomology, insertions*	*Mutation according to the HGVS nomenclature, reflected cDNA numbering with + 1 corresponding to the A of the ATG translation initiation codon of NM_004006.2*
Group 1:										
Subgroup 1-1: mutations in the major hot spot										
1	D127	dup 44	44	32 046 211	unique	43	32 166 116	LINE:HAL1	insA, insCTCCCCTGAACATGG	c.[6291 −21015_6438 + 98743dupinsA;6291 −21008_6291 −21007insCTCCCCTGAACATGG][c]
2	D9	del 49–50	48	31 784 029	unique	50	31 709 695	unique	C	c.7096 −19169_7310 −7465del
3	D13	del 49–50	48	31 767 209	unique	50	31 720 284	unique	ACT	c.7096 −2349_7310 −18054del
4	D14-1, D14-2	del 48–50	47	31 820 498	DNA:MER33	50	31 703 366	unique	no	c.6913 −17087_7310 −1136del
5	D19	del 49–50	48	31 798 920	unique	50	31 724 110	LINE:L2	AG	c.7095 + 4309_7310 −21880del
6	D22	del 45	44	31 957 454	LINE:L2	45	31 892 020	LINE:L1MC4	insCAA	c.6439 −60902_6614 + 4357delinsCAA
7	D30	del 45–50	44	31 957 280	unique	50	31 706 265	unique	insGTTAAAGACTAAACGTTAAA GACTAAAAAGACT	c.6439 −60728_7310 −4035delins33[c]
8	D52	del 45–50	44	31 907 441	unique	50	31 728 246	LTR:MLT-int	C	c.6439 −10889_7309 + 19767del
9	D55	del 48–50	47	31 835 531	LINE:L1PA6	50	31 714 900	LTR:LTR16A	insTACATATGGCTAGCAACTG	c.6912 + 22103_7310 −12670delins TACATATGGCTAGCAACTG[c]
10	D74	del 52	51	31 694 654	unique	52	31 650 541	DNA:looper	TA	c.7542 + 7344_7660 + 7128del
11	D76	del 49–50	48	31 778 087	LINE:L1PA13	50	31 745 541	unique	GTT	c.7096 −13227_7309 + 2472del
12	D90	del 45–50	44	31 910 575	unique	50	31 722 478	unique	AT	c.6439 −14023_7310 −20248del
13	D94	del 49–52	48	31 795 083	LTR:MER67D	52	31 633 021	unique	GA	c.7095 + 8146_7660 + 24648del
14	D537	del 52	51	31 676 244	unique	52	31 639 196	unique	TA	c.7543 −18458_7660 + 18473del
Subgroup 1-2: mutations in the minor hot spot										
15	D194	dup 3–7	7	32 695 937	unique	2	32 791 392	SINE:AluSx	TGTATT…TTT	c.94 −13534_649 + 41594dup
16	D217	dup 3–7	7	32 732 409	LINE:L1P2	2	32 793 852	unique	insAAATATTTTTG	c.94 −15994_649 + 5122dupinsAAATATTTTTG
17	D32	del 3–7	2	32 816 844	LINE:L1MA8	7	32 666 495	SINE:MIR	insCTTC	c.94 −38986_650 −39164delinsCTTC
18	D49-1, D49-2	del 3–7	2	32 899 501	unique	7	32 718 634	LINE:L1MA4A	insAATGTG	c.93 + 48676_649 + 18897delinsAATGTG
19	D117-1, D117-2, D441(117-3)	del 3–7	2	32 939 450	LINE:L2	7	32 721 319	LINE:L1MA4A	CA	c.93 + 8727_649 + 16212del
20	D242	del 3–7	2	32 929 682	unique	7	32 703 400	LTR:THE1D-int	no	c.93 + 18495_649 + 34131del
21	D416	del 3–7	2	32 930 790	LTR:MER51A	7	32 692 149	unique	GG	c.93 + 17387_649 + 45382del
Subgroup 1-3: other exceptions to the frame rule										
22	D15	dup 44–59	59	31 397 817	LINE:L1PA7	43	32 171 590	DNA:MER63D	TG	c.6291 −26489_8937 + 8327dup
23	D81	del 39–43	38	32 274 758	unique	43	32 206 686	LINE:L1MA2	G	c.5449 −640_6290 + 8881del
24	D98	del 2-6	1	32 950 664	unique	6	32 738 340	LINE:L2	AAT	c.32 −2426_531 −691del
25	D26	del 18–44	17	32 450 684	LINE:L2	44	32 014 284	LINE:L1ME3A	AAG	c.2169 −4515_6439 −117732del
26	D118	del 64	63	31 151 853	LINE:L1ME4a	64	31 150 277	SINE:AluY	TAA	c.9287 −694_9361 + 808del
27	D145	del 35–42	34	32 300 123	unique	42	32 222 108	unique	insATACAATA, insATGTTCCCGTTTTATAG	c.[4846-6885_6118 −6369delinsATACAATA; 4846 −6900_4846 −6899ins17][c]
28	D281	del 61–67	60	31 334 496	unique	67	31 116 525	unique	no	c.9084 + 38023_9808 −5583del
Group 2: complex rearrangements										
29	D388	tri 2–7	7	32 654 634	LINE:L3	1	33 016 269	SINE:AluSz	CTG	c.32 −68031_650 −27303tri
30	D396	del 5–7	4	32 762 719	unique	7	32 696 959	unique	AA	c.[264 + 10102_649 + 40572del; 961 −6503_1331 + 6231delins25]
31	D396[d]	del 10–11	9	32 579 963	unique	11	32 565 939	LTR:THE1C	insATGGGTAGTACACA GCATCTGTCTG	
32	D412-1, D412-2	tri 19–51	51	31 677 878	LINE:L1PA3	18	32 442 087	unique	TTAGAA	c.2292 + 3959_7543 −20092tri
Group 3: unknown mutations										
33	D87	dup 20	Ex 20	32 419 486	unique exonic	19	32 419 676	unique	no	c.2381 −120_2451dup

[a]Patients from the same family have the same family ID (identification) but different extension number.
[b]Chromosomal positions of the first and last nucleotide(s) deleted or duplicated.
[c]Sequences across the junctions of D127, D30, D55, D145 are detailed in Supplementary Material 3.
[d]Indicate the second breakpoint found in the same patient (D396) in case of complex rearrangements.

167-bp inclusion in the mature transcripts between exons 34 and 43 detected several years ago in the patient, and of unknown origin at that time (Supplementary Figure 2). Another example of tandem out-of-frame duplication of exon 44 (D127) also held a compound breakpoint junction but in this case the 58-bp pseudoexon sequence inserted between the two duplicated copies of exon 44 on the transcripts, was originated from a DNA:MER1A element in intron 43 in a distance of 3 kb from the aberrant duplication junction (data not shown).

DISCUSSION

In this study, we present the advent of a high-resolution custom-designed oligonucleotide array-CGH into clinical practice of a reference diagnostic laboratory for DMD. This method showed to be accurate and highly sensitive, cost-effective (price 75–100€ per one patient per one experiment for reagents and consumables) and able to detect rearrangements, which are different in type, size and localization in a time less than 5 days for one experiment of 12 patients simultaneously. Based on our practice guidelines for molecular diagnosis of DMD, we recommend that the results have to be supported with alternative diagnostic methods.

With our design, we achieved a very high resolution of array-CGH (<0.2 kb) both in males and females and in one case of prenatal diagnosis. We noticed that rare experimental imperfections around particular *DMD* regions (exons 13, 17, 45, Dp140, Dp71 and intron 67) might be conditioned by poor hybridization or, in contrast, partial cross-hybridization of the probes. This fact was conclusive that the design of the probes is determinant for the reliability of the results and gives us the clue for the future probe redesigning.

Despite the high incidence of detected alterations in the *DMD* gene,[5–7] little is known about their causative molecular mechanisms. In our study, microhomology was present in 60.6% of breakpoints being comparable to the findings of Mitsui *et al* (2010)[5] for the *DMD* gene. There were no breakpoint clustering noticed and different families of known repetitive sequences, whose role has already been demonstrated[8,9] in other diseases, were found in 48.5% of the junctions. However, this frequency does not differ significantly from that of transposable repetitive elements in the human genome (46%)[9] and could explain our findings. Finally, no low-copy repeats with extensive homology that could participate in DNA secondary structure formation was evidenced except in the 120-kb deletion junction involving exons 48–50. These observations supports the microhomology mediated mechanism model, which could be either non-homologous end-joining (NHEJ) or any alternative replication-based processes,[8] in the occurrence of *DMD* rearrangements.

Although all the deletion breakpoints were obtained, the sequence of only five duplication and two triplication breakpoints were acquired, confirming the hypothesis of tandem ('head-to-tail') junction. Six cases including three duplications, two double duplications and a triplication built in the duplication remained undetected on the sequence level. Because array-CGH gives only information about size of copy number gains and losses, but not their exact position and orientation, we anticipated the difficulties to obtain the duplication/triplication breakpoints due to their unknown genomic configuration and possibility of aberrant sequence insertions inside the junctions. The absence of amplification with different combinations of primers indirectly tends to confirm this hypothesis.

In conclusion, this large survey of 50 patients confirmed the previous observations[10–13] that array-CGH is a reliable and effective tool in detecting simple and complex *DMD* rearrangements. This approach offers some advantages over exon-based detection methods as it can identify pure intronic pathogenic events and it allows precise delineation of rearrangements, some of which may affect the splicing process. This is of high importance for the deep family investigation and a more accurate genotype/phenotype correlations, but also might be decisive factor for the optimal inclusion of patients in clinical trials. In general, it could lead to better understanding of the common fundamental mutational mechanisms, clarifying pathogenesis of diseases associated with instability in the genome.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by an Association Française contre les Myopathies (AFM), grant (AFM-14178 and AFM-15258) to AI. We thank all the physicians, in particular Pr P Sarda, Dr U Walther-Louvier, Dr E Campana Salort, Dr R Bernard, Dr L Lazaro, Dr S Mercier, Dr M Cossée, Dr A David, Pr B Leheup, Dr X Ferrer, Pr PS Jouk, Pr Ph Labrune for providing clinical information and patient material. We are grateful to Céline Saquet and Sylvie Chambert for DNA and RNA analyses in some of the patients, and Nicolas Gilbert for helpful discussion about repetitive elements.

1 Tuffery-Giraud S, Beroud C, Leturcq F *et al*: Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. *Hum Mutat* 2009; **30**: 934–945.
2 Abbs S, Tuffery-Giraud S, Bakker E, Ferlini A, Sejersen T, Mueller CR: Best practice guidelines on molecular diagnostics in Duchenne/Becker muscular dystrophies. *Neuromuscul Disord* 2010; **20**: 422–427.
3 Faivre L, Khau Van Kien P, Callier P *et al*: De novo 15q21.1q21.2 deletion identified through FBN1 MLPA and refined by 244K array-CGH in a female teenager with incomplete Marfan syndrome. *Eur J Med Genet* 2010; **53**: 208–212.
4 Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE: Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat* 2008; **29**: 6–13.
5 Mitsui J, Takahashi Y, Goto J *et al*: Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. *Am J Hum Genet* 2010; **87**: 75–89.
6 Oshima J, Magner DB, Lee JA *et al*: Regional genomic instability predisposes to complex dystrophin gene rearrangements. *Hum Genet* 2009; **126**: 411–423.
7 Toffolatti L, Cardazzo B, Nobile C *et al*: Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. *Genomics* 2002; **80**: 523–530.
8 Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, Patrinos GP: Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol* 2010; **20**: 222–233.
9 Belancio VP, Deininger PL, Roy-Engel AM: LINE dancing in the human genome: transposable elements and disease. *Genome Med* 2009; **1**: 97.
10 Bovolenta M, Neri M, Fini S *et al*: A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. *BMC Genomics* 2008; **9**: 572.
11 del Gaudio D, Yang Y, Boggs BA *et al*: Molecular diagnosis of Duchenne/Becker muscular dystrophy: enhanced detection of dystrophin gene rearrangements by oligonucleotide array-comparative genomic hybridization. *Hum Mutat* 2008; **29**: 1100–1107.
12 Hegde MR, Chin EL, Mulle JG, Okou DT, Warren ST, Zwick ME: Microarray-based mutation detection in the dystrophin gene. *Hum Mutat* 2008; **29**: 1091–1099.
13 Saillour Y, Cossee M, Leturcq F *et al*: Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method. *Hum Mutat* 2008; **29**: 1083–1090.
14 Tuffery-Giraud S, Saquet C, Chambert S *et al*: The role of muscle biopsy in analysis of the dystrophin gene in Duchenne muscular dystrophy: experience of a national referral centre. *Neuromuscul Disord* 2004; **14**: 650–658.
15 Rivier F, Tuffery S, Jellali AJ, Echenne B, Mornet D, Pons F: Mosaic expression of two dystrophins in a boy with progressive muscular dystrophy. *Muscle Nerve* 1998; **21**: 1317–1320.
16 Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)