



ORIGINAL ARTICLE

Genetic diagnosis of Duchenne and Becker muscular dystrophy using next-generation sequencing technology: comprehensive mutational search in a single platform

Byung Chan Lim,¹ Seungbok Lee,^{2,3} Jong-Yeon Shin,^{2,4} Jong-Il Kim,^{2,3,4,5} Hee Hwang,¹ Ki Joong Kim,¹ Yong Seung Hwang,¹ Jeong-Sun Seo,^{2,3,4,5,6} Jong Hee Chae¹

► Additional materials are published online only. To view these files please visit the journal online (<http://jmg.bmj.com>).

¹Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, Korea

²Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Korea

³Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea

⁴Psoma Therapeutics Inc., Seoul, Korea

⁵Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea

⁶Macrogen Inc., Seoul, Korea

Correspondence to

Professor Jong Hee Chae, Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Hospital, 101, 110-744 Daehak-ro, Jongno-gu, Seoul, Korea; chaeped1@snu.ac.kr

Dr Jeong-Sun Seo, Genomic Medicine Institute, Medical Research Center, Seoul National University College of Medicine 101 Daehak-ro, Jongno-gu, 110-799, Seoul, Korea; jeongsun@snu.ac.kr

BCL and SL contributed equally to this work.

Received 14 April 2011

Accepted 7 September 2011

Published Online First

3 October 2011

ABSTRACT

Background Duchenne muscular dystrophy or Becker muscular dystrophy might be a suitable candidate disease for application of next-generation sequencing in the genetic diagnosis because the complex mutational spectrum and the large size of the dystrophin gene require two or more analytical methods and have a high cost. The authors tested whether large deletions/duplications or small mutations, such as point mutations or short insertions/deletions of the dystrophin gene, could be predicted accurately in a single platform using next-generation sequencing technology.

Methods A custom solution-based target enrichment kit was designed to capture whole genomic regions of the dystrophin gene and other muscular-dystrophy-related genes. A multiplexing strategy, wherein four differently bar-coded samples were captured and sequenced together in a single lane of the Illumina Genome Analyser, was applied. The study subjects were 25 patients: 16 with deficient dystrophin expression without a large deletion/duplication and 9 with a known large deletion/duplication.

Results Nearly 100% of the exonic region of the dystrophin gene was covered by at least eight reads with a mean read depth of 107. Pathogenic small mutations were identified in 15 of the 16 patients without a large deletion/duplication. Using these 16 patients as the standard, the authors' method accurately predicted the deleted or duplicated exons in the 9 patients with known mutations. Inclusion of non-coding regions and paired-end sequence analysis enabled accurate identification by increasing the read depth and providing information about the breakpoint junction.

Conclusions The current method has an advantage for the genetic diagnosis of Duchenne muscular dystrophy and Becker muscular dystrophy wherein a comprehensive mutational search may be feasible using a single platform.

INTRODUCTION

Duchenne muscular dystrophy (DMD; MIM #310200) and Becker muscular dystrophy (BMD; MIM #3003376) are the most common forms of childhood muscular dystrophy affecting from 1 in 3500 to 1 in 6000 male births.^{1–2} Genetic testing of

the dystrophin gene (*DMD*; Xp21.2) is now the initial method for confirming the diagnosis, although a muscle biopsy might be considered if rapid and reliable genetic testing is unavailable.³ However, full characterisation of the mutational spectrum is necessary for genetic counselling, prenatal diagnosis and selecting the patients eligible for future mutation-specific treatments. Although the proportion of mutations differs slightly between studies, possibly reflecting bias in cohort selection and application of different molecular diagnostic methods,^{4–7} the mutational spectrum can be approximated as follows: a large deletion in about 60% of patients, a large duplication in about 10% of patients and small mutations confined mostly to coding exons in about 30% of patients. To date, no genetic testing has been developed to cover this whole mutational spectrum in a single platform. In most laboratories, methods for detecting large deletions/duplications^{5–10} and methods for detecting small mutations^{11–12} are conducted separately. In addition, the large size of the dystrophin gene requires considerable effort, cost and time for direct sequencing using the Sanger method.

Massively parallel or next-generation sequencing (NGS) technologies have become essential tools for searching for new human disease genes, most of which were identified by either whole-exome sequencing or targeted sequencing of the regions identified by linkage analyses.^{13–17} NGS technology is also useful for molecular diagnosis of certain diseases where laborious sequencing efforts are required because of the large gene size or the presence of multiple causative genes in single disease entities. A few proof-of-concept studies have been published recently in this field,^{18–20} although further optimisation and validation are required. Besides small mutations including point mutations and short insertions/deletions (indels) spanning several base pairs, large deletions/duplications can also be identified by NGS technologies through read depth estimation.²¹ Because NGS provides a comprehensive mutation search from large deletions/duplications to small mutations in a single platform, DMD/BMD is a suitable candidate disease for testing whether NGS can be applied for molecular diagnosis.

Methods

To demonstrate whether large deletions/duplications and small mutations of DMD/BMD can be predicted accurately in a single NGS platform, we analysed the entire dystrophin gene regions of 25 patients with DMD or BMD using solution capture with bar-code multiplexing and massively parallel sequencing. First, the pathogenic mutations were searched in 16 patients with pathologically proven DMD/BMD who were negative for large deletions/duplications. Next, we showed how large deletion/duplication mutations could be detected in the same platform by analysing the sequencing data of nine patients with DMD/BMD with known deletion/duplication mutations.

METHODS

Patients

The Institutional Review Board of the Seoul National University Hospital approved the study protocol (H-1104-053-358). Multiplex ligation-dependent probe amplification (MLPA) was the first molecular diagnostic method used in Seoul National University Children's Hospital since 2006 when DMD/BMD was suspected from the clinical and laboratory findings. If a large deletion/duplication is not identified with MLPA, the next step is to confirm deficient dystrophin staining in the muscle biopsy specimen. We randomly selected 16 patients who were candidates for small mutations because they were negative for large deletions/duplications and positive for deficient dystrophin staining. The additional nine patients harbouring large deletion/duplication mutations, which were confirmed with MLPA, were included to validate whether large deletion/duplication mutations could be correctly predicted through NGS sequencing data. The MLPA, dystrophin immunohistochemistry findings and clinical phenotypes of the 25 male patients are summarised in table 1.

Target enrichment of genomic DNA and sequencing

After the patients provided informed consent, genomic DNA was extracted from peripheral blood leucocytes using a QIAamp DNA Blood Midi Kit (Qiagen, California, USA) according to the manufacturer's instructions. Three micrograms of DNA was sheared using a Covaris S2 (Covaris, Inc., Massachusetts, USA) to ~250 nt at a 20% duty cycle, level 5 intensity and 200 cycles per burst for 180 s. Bar-coded fragment sequencing libraries were made using an NEB sample preparation kit (New England Biolabs Inc., Massachusetts, USA) and Illumina multiplexing adaptor (Illumina, California, USA) according to the manufacturers' instructions. After ligation with the Illumina adaptor, the libraries were prepared using AMPure bead (Beckman Coulter, Inc., California, USA) rather than gel purification. Library quality was assessed using an Agilent 2100 Analyser and DNA 1000 chips (Agilent Technology, California, USA). An equimolar four-plex pool was produced for enrichment using a SureSelect Target Enrichment System Kit (Agilent Technology) and a modified protocol. Five hundred nanogrammes of pooled DNA with 5 µl (100 ng) of custom baits were used for enrichment, with blocking oligonucleotides specific for paired-end sequencing libraries and 24-h hybridisation. Biotinylated RNA library hybrids were recovered with streptavidin beads. In addition to the dystrophin gene, the target regions included 25 genes involved in congenital muscular dystrophies and limb girdle muscular dystrophies (supplementary table 1). The dystrophin gene region to be captured is at genomic position ChrX:31 047 266 to 33 139 594 (NM_004006, hg18, National Center for Biotechnology Information (NCBI) build 36), and this region was uploaded to Agilent's web-based design tool, eArray. The parameters selected were bait length (120 bp), bait tiling

Table 1 Laboratory data and clinical phenotypes of the patients

| Patient number | MLPA | Dystrophin IHC | Age at last follow-up (WC users) (years) | Phenotype* |
|----------------|-----------|----------------|------------------------------------------|--------------|
| 1 | (-) | Complete loss | 14.6 (9.5) | DMD |
| 2 | (-) | Complete loss | 10.3 | Undetermined |
| 3 | (-) | Focal loss | 21.8 | BMD |
| 4 | (-) | Complete loss | 6.6 | Undetermined |
| 5 | (-) | Complete loss | 3.2 | Undetermined |
| 6 | (-) | Focal loss | 7.7 | Undetermined |
| 7 | (-) | Complete loss | 3.4 | Undetermined |
| 8 | (-) | Complete loss | 11.0 | Undetermined |
| 9 | (-) | Complete loss | 12.7 (10.9) | DMD |
| 10 | (-) | Complete loss | 11.4 (9.4) | DMD |
| 11 | (-) | Complete loss | 10.2 (10) | DMD |
| 12 | (-) | Complete loss | 7.4 | Undetermined |
| 13 | (-) | Complete loss | 9.4 | Undetermined |
| 14 | (-) | Complete loss | 7.9 | Undetermined |
| 15 | (-) | Complete loss | 13.7 (9) | DMD |
| 16 | (-) | Complete loss | 10.0 | Undetermined |
| 17 | del 12-43 | Focal loss | 10.6 | Undetermined |
| 18 | del 3-13 | Not done | 9.6 | Undetermined |
| 19 | del 8-13 | Not done | 7.4 | Undetermined |
| 20 | del 52-53 | Not done | 14.7 | Undetermined |
| 21 | del 53-54 | Complete loss | 14.1 (11.6) | DMD |
| 22 | del 53 | Not done | 15.4 (10.1) | DMD |
| 23 | dup 30-47 | Focal loss | 23.6 | BMD |
| 24 | dup 45-52 | Focal loss | 8.4 | Undetermined |
| 25 | dup 50-55 | Complete loss | 13.8 (9.3) | DMD |

*Phenotype was determined according to the age at which wheelchair dependency began (DMD: wheelchair dependency before 13 years, BMD: wheelchair dependency after 16 years).

(-), negative for deletion/duplication; BMD, Becker muscular dystrophy; del, deleted exons; DMD, Duchenne muscular dystrophy; dup, duplicated exons; IHC, immunohistochemistry; MLPA, multiplex ligation-dependent probe amplification; WC, wheelchair.

frequency (2×), allowed overlap into avoided regions (20 bp) and avoided standard repeat masked regions, which eliminates repetitive sequences using the RepeatMasker program alignment-based method. The bait region, which was defined as the region covered by one or more capture probes in the dystrophin gene, was 1 069 974 bp (supplementary file 1). The captured libraries were amplified and sequenced on the Illumina Genome Analyser IIX by 2×69 cycles. The other experimental procedures were performed according to the protocol of each manufacturer.

Single nucleotide variant (SNV) and indel detection

The sequencing data were processed using the Genomic Short-read Nucleotide Alignment Program (GSNAP) alignment tool,^{22 23} and human genome NCBI build 36 was used as the reference. Reads aligned to multiple regions of the reference sequence were removed for further analysis. The variant detection criteria for SNVs, insertions (<10 bases) and deletions (<31 bases) were as follows: (1) ≥50% of the reads indicated variation, (2) ≥8 reads were uniquely aligned to each position of variance and (3) the mean quality score (Q score) was ≥20. The identified SNVs and indels were compared with the data from the 1000 Genomes Project, dbSNP132, data from 68 Korean healthy controls (13 from whole genome sequencing and 55 from whole-exome sequencing data) and the Leiden DMD Mutation Database.²⁴ The novel nonsense or frameshift variants were regarded as putative pathogenic mutations. The functional consequences of the novel missense variants were predicted using Sorting Intolerant from Tolerant (SIFT) and Polymorphism Phenotyping (PolyPhen).

Large deletion/duplication prediction

The exonic coverage depth of each dystrophin gene exon was calculated by dividing the sum of all read depths per base within

the specific exon by the size of that exon (base pairs) and corrected by all aligned bases within the bait region except the dystrophin gene to ensure comparability between experiments. The reference exonic coverage depth of each dystrophin gene exon was obtained by averaging the exonic coverage depths of the 16 patients without deletion/duplication. The exonic coverage depth of patients 17–25 was compared with the corresponding reference exonic coverage depth in each of the 79 exons. The threshold ratio for deletion was set at <0.2 , and that for duplication was set at >1.8 . After identifying a series of deleted or duplicated exons, the range of deleted or duplicated regions was approximated by plotting the read depth spanning the whole dystrophin gene region. Finally, the breakpoint of the deleted or duplicated region was estimated with the aid of aberrant paired-end sequencing calls. In the paired-end sequencing setting, the GSNAP alignment tool gives a ‘too long’ signal when the distance of a pair of reads is >1000 bp, indicating the possibility of a large deletion between that pair of reads. The GSNAP reports ‘inversion’ when the orientations are opposite to those expected and ‘scramble’ when the distance appears to be negative. These two signals suggest the disruption of genomic integrity and the possibility of duplication near both reads of pairs.

MLPA and validation through Sanger sequencing

MLPA analysis was performed using the SALSA MLPA KIT P034-A2/P035-A2 DMD/Becker (MRC Holland, Amsterdam, The Netherlands) according to the manufacturer’s instructions. Direct Sanger sequencing of the selected regions was performed using primer pairs designed by the authors (available upon request) to validate the coding or known pathogenic intronic variations identified through NGS. For breakpoint junction analysis, PCR primers were designed to cover the deleted or duplicated regions suggested by the paired-end sequencing (available upon request).

RESULTS

Sequencing summary and small mutation detection (patients 1–16)

Analysis of the sequence data revealed uniform coverage and high read depths in 16 patients (table 2). For estimation of

coverage, only positions having ≥ 8 uniquely aligned reads with a Q score ≥ 20 were included. The bait region including all the dystrophin gene exons was almost completely covered by ≥ 8 read depth. The mean read depth of the exons was 107 (range, 86–125). Accordingly, about 68% of the whole dystrophin gene region was covered adequately to identify SNVs and short indels. The number of total SNVs and indels and the number of variants within the coding exons are summarised in supplementary table 2. There were no SNVs or short indels involving splice sites. All coding SNVs and short indels were presented specifically and compared with the data from the 1000 Genomes Project, dbSNP132 and data from 68 Korean healthy controls; these data are provided in supplementary file 2. All coding variants were confirmed using Sanger sequencing. We also tested whether there were false negatives among other bases in the exons selected for Sanger sequencing; no false-negative variant was present. Putative small mutations were identified in 15 of 16 patients (table 3); most were either nonsense or frameshift mutations causing truncation of dystrophin. Intronic variant in patient 6 (c.4518+5G>A) was demonstrated previously to cause aberrant splicing in vivo and in vitro.²⁵ Novel nonsense (c.1006G>T, p.Glu336X) and missense (c.1005G>A, p.Met335Ile) variants were found in patient 15. This missense variant was not considered as a putative pathogenic mutation because SIFT and PolyPhen predicted that the functional consequences of this variant are ‘tolerated’ and ‘benign’. The functional consequences of the novel missense variant in patient 16 (c.450T>G, p.Asn150Lys) were also predicted by SIFT and PolyPhen as tolerated and benign.

Large deletion/duplication prediction (patients 17–25)

The reference exonic coverage depth of each dystrophin gene exon was obtained using the methods described above (supplementary file 3). With the predetermined threshold criteria (deletion <0.2 , duplication >1.8), deleted or duplicated exons in all nine patients were completely consistent with the results identified using the MLPA method (table 4, supplementary file 4). The coverage plot of these nine patients reconfirmed these results (figure 1). The junctions of deleted or duplicated regions, which were usually located in introns, were also roughly estimated from these coverage plots. To determine the breakpoint

Table 2 Sequencing summary of candidates for small mutations (patients 1–16)

| Patient number | DMD bases* aligned (≥ 8) | Bait bases† aligned (≥ 8) | Exonic bases aligned (≥ 8) | Percentage of covered DMD bases (≥ 8) | Percentage of covered bait bases (≥ 8) | Percentage of covered exonic bases (≥ 8) | Mean coverage depth (bait) | Mean coverage depth (exons) |
|----------------|---------------------------------|----------------------------------|-----------------------------------|----------------------------------------------|-----------------------------------------------|-------------------------------------------------|----------------------------|-----------------------------|
| 1 | 1 431 187 | 1 068 822 | 11 058 | 68.34 | 99.89 | 100 | 111 | 119 |
| 2 | 1 416 230 | 1 068 913 | 11 054 | 67.62 | 99.90 | 99.96 | 105 | 120 |
| 3 | 1 413 052 | 1 068 445 | 11 056 | 67.47 | 99.86 | 99.98 | 106 | 121 |
| 4 | 1 409 103 | 1 068 916 | 11 057 | 67.28 | 99.90 | 99.99 | 104 | 118 |
| 5 | 1 421 369 | 1 069 040 | 11 058 | 67.87 | 99.91 | 100 | 106 | 118 |
| 6 | 1 423 045 | 1 068 672 | 11 058 | 67.95 | 99.88 | 100 | 114 | 125 |
| 7 | 1 408 625 | 1 068 542 | 11 058 | 67.26 | 99.87 | 100 | 90 | 103 |
| 8 | 1 417 737 | 1 068 931 | 11 057 | 67.69 | 99.90 | 99.99 | 98 | 114 |
| 9 | 1 409 783 | 1 068 517 | 11 058 | 67.31 | 99.86 | 100 | 97 | 108 |
| 10 | 1 424 654 | 1 068 703 | 11 055 | 68.02 | 99.88 | 99.97 | 88 | 97 |
| 11 | 1 430 832 | 1 068 936 | 11 053 | 68.32 | 99.90 | 99.95 | 93 | 103 |
| 12 | 1 421 946 | 1 068 838 | 11 055 | 67.90 | 99.89 | 99.97 | 80 | 89 |
| 13 | 1 414 348 | 1 068 907 | 11 057 | 67.53 | 99.90 | 99.99 | 77 | 86 |
| 14 | 1 422 564 | 1 069 040 | 11 058 | 67.92 | 99.91 | 100 | 95 | 102 |
| 15 | 1 461 341 | 1 069 029 | 11 045 | 69.78 | 99.91 | 99.88 | 91 | 100 |
| 16 | 1 443 643 | 1 068 980 | 11 058 | 68.93 | 99.91 | 100 | 86 | 96 |

*DMD bases, bases within the DMD genomic region located at genomic position 31 047 266 to 33 139 594 (NM_004006, hg18, NCBI build 36).

†Bait bases, bases within the bait region defined as the region covered by more than one capture probe in the DMD genomic region (1 069 947 bp).

Methods

Table 3 Summary of putative pathogenic mutations validated via Sanger sequencing

| Patient number | Mutations identified | | | | Number of reads | | |
|----------------|----------------------|-------------------|---------------|-----------------|-----------------|-----------|-----------------------|
| | Base change | Effect | Mutation type | Leiden database | Variant | Wild type | Percentage of variant |
| 1 | c.10033C>T | p.Arg3345X | Nonsense | Reported | 132 | 0 | 100 |
| 2 | c.1615C>T | p.Arg539X | Nonsense | Reported | 142 | 2 | 98.61 |
| 3 | c.10406delA | p.His3469LeufsX11 | Frameshift | Reported | 47 | 15 | 75.80 |
| 4 | c.3964C>T | p.Gln1322X | Nonsense | Unreported | 135 | 0 | 100 |
| 5 | c.2527G>T | p.Glu843X | Nonsense | Reported | 97 | 3 | 97 |
| 6 | c.4518+5G>A | p.Ser1474LysfsX4 | Splicing | Reported | 110 | 0 | 100 |
| 7 | c.6283C>T | p.Arg2095X | Nonsense | Reported | 129 | 3 | 97.72 |
| 8 | c.8209C>T | p.Gln2737X | Nonsense | Unreported | 174 | 1 | 99.42 |
| 9 | c.10126delC | p.Leu3376X | Nonsense | Reported | 79 | 40 | 66.38 |
| 10 | c.1663C>T | p.Gln555X | Nonsense | Reported | 109 | 1 | 99.09 |
| 11 | c.3244G>T | p.Glu1082X | Nonsense | Reported | 91 | 0 | 100 |
| 12 | c.1615C>T | p.Arg539X | Nonsense | Reported | 128 | 1 | 99.22 |
| 13 | c.620T>G | p.Leu207X | Nonsense | Reported | 113 | 2 | 98.26 |
| 14 | c.583C>T | p.Arg195X | Nonsense | Reported | 159 | 0 | 100 |
| 15 | c.1006G>T | p.Glu336X | Nonsense | Unreported | 154 | 0 | 100 |

precisely, we used the raw coverage depth and paired-end sequencing data around the junctional sequence identified from the coverage plots. Subsequent Sanger sequencing confirmed the breakpoints in seven patients (patient 17 and patients 19–24; table 4). Paired-end data for patient 25 revealed an inversion between intron 55 of the dystrophin gene and intron 13 of *PHEX*, suggesting the presence of complex rearrangement in addition to duplication of exons 50 to 55 of the dystrophin gene.

DISCUSSION

Although the targeted resequencing by NGS technology has gained much attention in the field of molecular diagnostics, several limitations have delayed its application to medical genetic diagnosis.^{19 20 26} It is difficult to standardise the procedures because of the use of various sequencing platforms and target enrichment methods, which are updated rapidly. The cut-off thresholds for accurate variant identification, including the minimum read depth, range of variant percentage compared with the wild type and quality score, have not been defined fully. Variable coverage depth across target regions and misalignment between homologous sequences or pseudogenes may also affect the accuracy of sequencing data.

Accepting that some problems are inherent in NGS technologies regardless of the target regions, we believe that the dystrophin gene has an advantage over other target regions located in autosomal chromosomes. Because of hemizyosity and the absence of reported pseudogenes, a less stringent cut-off threshold can be applied. This advantage may also make target enrichment with bar-code multiplexing feasible, thus further reducing the cost and time, which are other obstacles to the

clinical application. A recent report by Cummings *et al*,²⁷ which used multiplexing and target enrichment protocols similar to those used in the present study, provides a proof-of-concept example for our study. Even after multiplexing four samples in one lane, nearly 100% of regions of all the dystrophin gene exons were covered optimally to identify the SNVs and indels. Although most research groups suggest that a minimum of 10–30 read depths is needed for accurate SNV identification,^{19 20 28–30} the cut-off criteria used in the present study might be acceptable considering the hemizyosity of the dystrophin gene.

The high proportion of variants compared with the wild type illustrates the advantage of hemizyosity of the dystrophin gene when identifying SNVs (table 3). The relatively low proportion of variants in indels (table 3) may stem from the terminal location of deleted or inserted bases in part of the aligned reads. When deleted or inserted bases are located in either end of an aligned read, they could be erroneously identified as a series of SNVs rather than indels. Although a separate algorithm for indel detection may be necessary to improve sensitivity, we applied the same criteria to both SNVs and indels because the high coverage depth in this study and the hemizyosity of the target region should decrease the likelihood of such false negatives. Because 16 patients in the present study were selected randomly and their genotypes were unconfirmed, the mutation detection rate might provide further support for the accuracy of this method. The pathogenic mutations were identified in 15 of 16 patients, and the mutation detection rate and distribution of mutation by types (12 nonsense, 2 small deletions causing frameshift and 1 splicing mutation) were similar to the results of

Table 4 Paired-end sequencing and breakpoint junction analysis of patients with deletion/duplication mutations

| Patient number | Type | Exons | Aberrant paired-end call | Breakpoint* | Size (bp) |
|----------------|-------------|-------|--------------------------|------------------------------|-----------|
| 17 | Deletion | 12–43 | Undetected | ChrX:g.32 155 793–32 548 092 | 392 299 |
| 18 | Deletion | 3–13 | Undetected | Not determined | |
| 19 | Deletion | 8–13 | Too long | ChrX:g.32 507 320–32 669 677 | 162 357 |
| 20 | Deletion | 52–53 | Too long | ChrX:g.31 606 256–31 658 241 | 51 985 |
| 21 | Deletion | 53–54 | Too long | ChrX:g.31 580 658–31 640 215 | 59 557 |
| 22 | Deletion | 53 | Too long | ChrX:g.31 599 954–31 657 066 | 57 112 |
| 23 | Duplication | 30–47 | Scramble | ChrX:g.31 807 975–32 356 021 | 548 046 |
| 24 | Duplication | 45–52 | Scramble | ChrX:g.31 653 473–32 040 375 | 386 902 |
| 25 | Duplication | 50–55 | Inversion | Not determined | |

*The genomic positions were referenced from hg18, NCBI build 36.

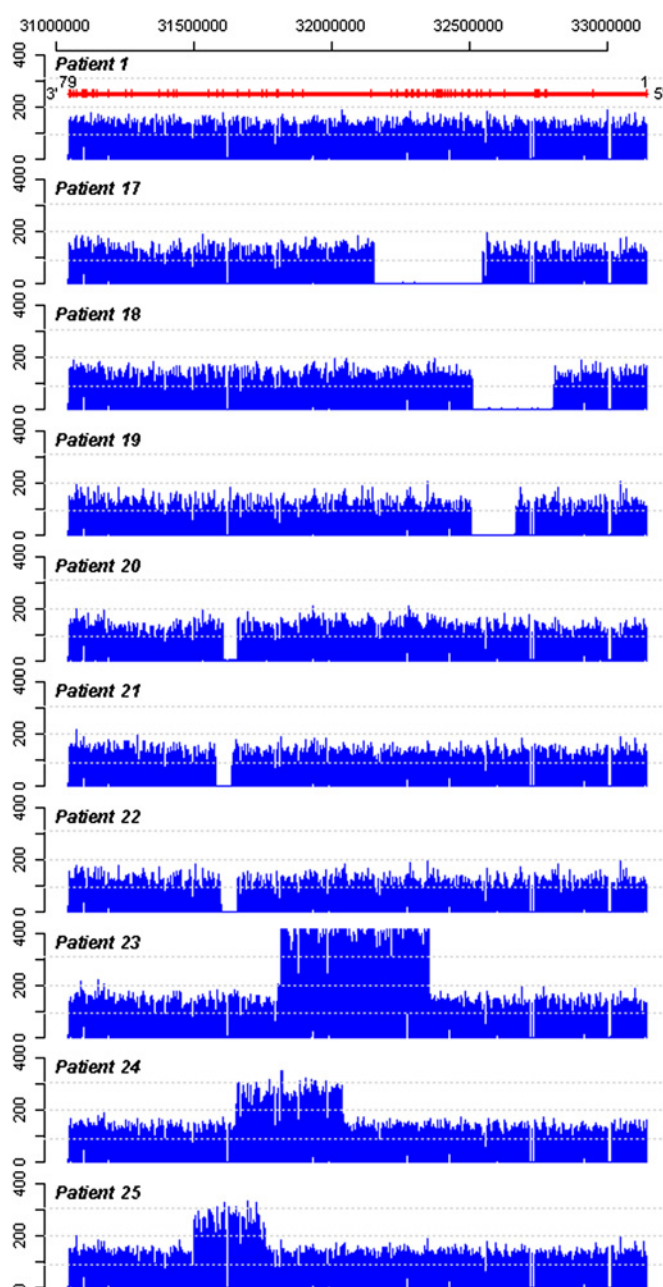


Figure 1 The coverage plot of nine patients with a large deletion (patients 17–22) or duplication (patients 23–25) mutation. The data of patient 1, who lacked a deletion or duplication mutation, are included for comparison. Red vertical bars indicate 79 DMD exons.

other studies using different methods.^{6,7} As the SIFT and PolyPhen prediction suggested tolerable or benign functional changes, the novel missense variant found in patient 16 (c.450T>G, p.Asn150Lys) was not classified as a pathogenic mutation.

As described above, variable coverage across the target regions is inherent in the targeted resequencing method. Although the coverage depth may differ among each of the 79 dystrophin gene exons in a single patient, this pattern of variation seems to be reproducible in multiple patients sharing the same target enrichment and NGS platform (supplementary file 1). This was demonstrated recently by Walsh *et al*²¹ Because a large deletion/duplication of the dystrophin gene typically includes several exons, comparison using the coverage depth per exon (exonic

coverage depth in the present study) might be justified. Compared with the target genes located in autosomal chromosomes, copy number changes would be more obvious in the dystrophin gene because of hemizygosity. This difference is particularly significant when analysing large deletions because the coverage depth of the deleted region approaches zero.

The non-coding region of the dystrophin gene, except for the repetitive area, was also captured and sequenced in the present study. Although intronic variations contribute little to the whole DMD/BMD mutational spectrum and are difficult to validate by genomic sequencing alone, inclusion of intronic regions may have several implications for DMD/BMD molecular diagnosis. First, targeting only exonic regions may produce an uneven distribution of coverage within the exon, resulting in the possible detection of a suboptimally covered region. Cummings *et al* noted that the coverage per base is reduced at the flanks and is increased in the middle of the target regions when discontinuous exons are resequenced.²⁷ Sequencing of nearly contiguous regions encompassing non-coding regions would increase the mean read depth of all the dystrophin gene exons, which would appear more evenly distributed when plotted across the whole dystrophin gene region, as indicated in figure 1. This positively affects the accuracy of the prediction of both small mutations and large deletion/duplication mutations.

Second, aberrant paired-end sequencing calls (too long, scramble and inversion) usually identified in intronic sequences could provide clues for breakpoint determination in large duplication/deletion cases. Although breakpoint determination from paired-end sequencing data may not be an indispensable step in routine genetic diagnosis, this method may provide additional information regarding complex rearrangement, as in the case of patient 25. Contrary to the prediction of in-frame duplication using the reading frame rule in patient 25 (exons 50–55), the clinical course and expression of dystrophin in muscle were consistent with DMD. In addition, insertion of *PHEX* exons into the dystrophin gene was identified by reverse transcription PCR experiments (data not shown). As it is predicted to cause the protein truncation, it could explain the severity in patient 25.

Compared with previous proof-of-concept studies that discussed the detailed methodological perspectives in limited cases, the present study emphasises a more practical approach by testing a relatively large number of patients and demonstrating complex mutational spectra from large deletions/duplications to small mutations in this single platform. This method would reduce the time necessary to reach the genetic diagnosis of DMD/BMD compared with the current method—a combination of MLPA and Sanger sequencing. Moreover, multiplexing and inclusion of other related genes might contribute to reducing the cost, which is one of the main obstacles for clinical use. We hope that the present study provides practical evidence to encourage the application of NGS technology to routine genetic diagnosis of DMD/BMD.

Funding This study was supported by a grant from the Korea Healthcare Technology R&D Project, Ministry for Health, Welfare and Family Affairs, Republic of Korea (Grant No. A080588) and Seoul Broadcasting System Grant-in-Aid for Seoul National University Children's Hospital Research (Grant No.30-2010-0190).

Competing interests None.

Ethics approval Ethics approval was provided by the Institutional Review Board of the Seoul National University Hospital.

Contributors Study concept and design: Byung Chan Lim, Seungbok Lee, Jong-Yeon Shin, Jeong-Sun Seo, Jong-Il Kim and Jong Hee Chae. Acquisition of data: Byung Chan Lim and Seungbok Lee. Genetic analysis: Seungbok Lee, Jong-Yeon Shin and Byung Chan Lim. Drafting of the manuscript: Byung Chan Lim and Seungbok Lee.

Critical revision of the manuscript for important intellectual content: Ki Joong Kim, Yong Seung Hwang and Hee Hwang.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Emery AE**. Population frequencies of inherited neuromuscular diseases—a world survey. *Neuromuscul Disord* 1991;**1**:19–29.
2. **Miller LA**, Romitti PA, Cuniff C, Druschel C, Mathews KD, Meaney FJ, Matthews D, Kantamneni J, Feng ZF, Zemblidge N, Miller TM, Andrews J, Fox D, Ciafaloni E, Pandya S, Montgomery A, Kenneson A. The muscular Dystrophy Surveillance Tracking and Research Network (MD STARnet): surveillance methodology. *Birth Defects Res A Clin Mol Teratol* 2006;**76**:793–7.
3. **Bushby K**, Finkel R, Birnkrant DJ, Case LE, Clemens PR, Cripe L, Kaul A, Kinnett K, McDonald C, Pandya S, Poysky J, Shapiro F, Tomezsko J, Constantin C; DMD Care Considerations Working Group. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet Neurol* 2010;**9**:77–93.
4. **Cuniff C**, Andrews J, Meaney FJ, Mathews KD, Matthews D, Ciafaloni E, Miller TM, Bodensteiner JB, Miller LA, James KA, Druschel CM, Romitti PA, Pandya S. Mutation analysis in a population-based cohort of boys with Duchenne or Becker muscular dystrophy. *J Child Neurol* 2009;**24**:425–30.
5. **Dent KM**, Dunn DM, von Niederhausern AC, Aoyagi AT, Kerr L, Bromberg MB, Hart KJ, Tuohy T, White S, den Dunnen JT, Weiss RB, Flanigan KM. Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort. *Am J Med Genet A* 2005;**134**:295–8.
6. **Flanigan KM**, Dunn DM, von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, Sampson JB, Mendell JR, Wall C, King WM, Pestronk A, Florence JM, Connolly AM, Mathews KD, Stephan CM, Laubenthal KS, Wong BL, Morehart PJ, Meyer A, Finkel RS, Bonnemann CG, Medne L, Day JW, Dalton JC, Margolis MK, Hinton VJ, Weiss RB; United Dystrophinopathy Project Consortium. Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. *Hum Mutat* 2009;**30**:1657–66.
7. **Takeshima Y**, Yagi M, Okizuka Y, Awano H, Zhang Z, Yamauchi Y, Nishio H, Matsuo M. Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *J Hum Genet* 2010;**55**:379–88.
8. **Flanigan KM**, von Niederhausern A, Dunn DM, Alder J, Mendell JR, Weiss RB. Rapid direct sequence analysis of the dystrophin gene. *Am J Hum Genet* 2003;**72**:931–9.
9. **Lalic T**, Vossen RH, Coffa J, Schouten JP, Guc-Scekic M, Radivojevic D, Djuricic M, Breuning MH, White SJ, den Dunnen JT. Deletion and duplication screening in the DMD gene using MLPA. *Eur J Hum Genet* 2005;**13**:1231–4.
10. **Prior TW**, Bridgeman SJ. Experience and strategy for the molecular testing of Duchenne muscular dystrophy. *J Mol Diagn* 2005;**7**:317–26.
11. **Bennett RR**, den Dunnen J, O'Brien KF, Darras BT, Kunkel LM. Detection of mutations in the dystrophin gene via automated DHPLC screening and direct sequencing. *BMC Genet* 2001;**2**:17.
12. **Mendell JR**, Buzin CH, Feng J, Yan J, Serrano C, Sangani DS, Wall C, Prior TW, Sommer SS. Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. *Neurology* 2001;**57**:645–50.
13. **Ng SB**, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;**42**:30–5.
14. **Haack TB**, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J, Rolinski B, Iuso A, Biskup S, Schmidt T, Mewes HW, Wittig I, Meitinger T, Zeviani M, Prokisch H. Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nat Genet* 2010;**42**:1131–4.
15. **Bilgüvar K**, Öztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tüysüz B, Çağlayan AO, Gökben S, Kaymakçalan H, Barak T, Bakircioğlu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yılmaz S, Dinçer A, Johnson MH, Bronen RA, Koçer N, Per H, Mane S, Pamir MN, Yalçinkaya C, Kumandaş S, Topçu M, Ozmen M, Sestan N, Lifton RP, State MW, Günel M. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010;**467**:207–10.
16. **Vermeer S**, Hoischen A, Meijer RP, Gilissen C, Neveling K, Wieskamp N, de Brouwer A, Koenig M, Anheim M, Assoum M, Drouot N, Todorovic S, Milic-Rasic V, Lochmüller H, Stevanin G, Goizet C, David A, Brice A, Kremer B, van de Warrenburg BP, Schijvenaars MM, Heister A, Kwint M, Arts P, van der Wijst J, Veltman J, Kamsteeg EJ, Scheffer H, Knoers N. Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia. *Am J Hum Genet* 2010;**87**:813–19.
17. **Nikopoulos K**, Gilissen C, Hoischen A, van Nouhuys CE, Boonstra FN, Blokland EA, Arts P, Wieskamp N, Strom TM, Ayuso C, Tilanus MA, Bouwhuis S, Mukhopadhyay A, Scheffer H, Hoefsloot LH, Veltman JA, Cremers FP, Collin RW. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *Am J Hum Genet* 2010;**86**:240–7.
18. **Amstutz U**, Andrey-Zurcher G, Suci D, Jaggi R, Haberle J, Largiadier CR. Sequence capture and next-generation resequencing of multiple tagged nucleic acid samples for parallel screening of urea cycle disorders. *Clin Chem* 2010;**57**:102–11.
19. **Benaglio P**, Rivolta C. Ultra high throughput sequencing in human DNA variation detection: a comparative study on the NDUFA3-PRPF31 region. *PLoS One* 2010;**5**:e13071.
20. **Chou LS**, Liu CS, Boese B, Zhang X, Mao R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: neurofibromatosis type 1 gene as a model. *Clin Chem* 2010;**56**:62–72.
21. **Walsh T**, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* 2010;**107**:12629–33.
22. **Kim JI**, Ju YS, Park H, Kim S, Lee S, Yi JH, Mudge J, Miller NA, Hong D, Bell CJ, Kim HS, Chung IS, Lee WC, Lee JS, Seo SH, Yun JY, Woo HN, Lee H, Suh D, Lee S, Kim HJ, Yavartanoo M, Kwak M, Zheng Y, Lee MK, Park H, Kim JY, Gokcumen O, Mills RE, Zaranek AW, Thakuria J, Wu X, Kim RW, Huntley JJ, Luo S, Schroth GP, Wu TD, Kim H, Yang KS, Park WY, Kim H, Church GM, Lee C, Kingsmore SF, Seo JS. A highly annotated whole-genome sequence of a Korean individual. *Nature* 2009;**460**:1011–15.
23. **Wu TD**, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 2010;**26**:873–81.
24. **Aartsma-Rus A**, Van Deutekom JC, Fokkema IF, Van Ommen GJ, Den Dunnen JT. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve* 2006;**34**:135–44.
25. **Thi Tran HT**, Takeshima Y, Surono A, Yagi M, Wada H, Matsuo M. A G-to-A transition at the fifth position of intron-32 of the dystrophin gene inactivates a splice-donor site both in vivo and in vitro. *Mol Genet Metab* 2005;**85**:213–19.
26. **Voelkerding KV**, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics—principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn* 2010;**12**:539–51.
27. **Cummings N**, King R, Rickers A, Kaspi A, Lunke S, Haviv I, Jowett JB. Combining target enrichment with barcode multiplexing for high throughput SNP discovery. *BMC Genomics* 2010;**11**:641.
28. **Smith DR**, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res* 2008;**18**:1638–42.
29. **Harismendy O**, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 2009;**10**:R32.
30. **Li R**, Li Y, Fang X, Yang H, Wang J, Kristiansen K. SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009;**19**:1124–32.