ARTICLE

Challenges in detecting genomic copy number aberrations using next-generation sequencing data and the eXome Hidden Markov Model: a clinical exome-first diagnostic approach

Toshiyuki Yamamoto[1], Keiko Shimojima[1], Yumiko Ondo[1], Katsumi Imai[2], Pin Fee Chong[3], Ryutaro Kira[3], Mitsuhiro Amemiya[4], Akira Saito[4] and Nobuhiko Okamoto[5]

Next-generation sequencing (NGS) is widely used for the detection of disease-causing nucleotide variants. The challenges associated with detecting copy number variants (CNVs) using NGS analysis have been reported previously. Disease-related exome panels such as Illumina TruSight One are more cost-effective than whole-exome sequencing (WES) because of their selective target regions (~21% of the WES). In this study, CNVs were analyzed using data extracted through a disease-related exome panel analysis and the eXome Hidden Markov Model (XHMM). Samples from 61 patients with undiagnosed developmental delays and 52 healthy parents were included in this study. In the preliminary study to validate the constructed XHMM system (microarray-first approach), 34 patients who had previously been analyzed by chromosomal microarray testing were used. Among the five CNVs larger than 200 kb that were considered as non-pathogenic CNVs and were used as positive controls, four CNVs was successfully detected. The system was subsequently used to analyze different samples from 27 patients (NGS-first approach); 2 of these patients were successfully diagnosed as having pathogenic CNVs (an unbalanced translocation der(5)t(5;14) and a 16p11.2 duplication). These diagnoses were re-confirmed by chromosomal microarray testing and/or fluorescence *in situ* hybridization. The NGS-first approach generated no false-negative or false-positive results for pathogenic CNVs, indicating its high sensitivity and specificity in detecting pathogenic CNVs. The results of this study show the possible clinical utility of pathogenic CNV screening using disease-related exome panel analysis and XHMM.

Human Genome Variation (2016) **3**, 16025; doi:10.1038/hgv.2016.25; published online 18 August 2016

INTRODUCTION

Chromosomal microarray testing is a powerful tool for detecting genomic copy number aberrations. Combinatorial use of fluorescent dyes and oligonucleotide probes has enabled the detection of invisible small chromosomal deletions or duplications. Pathogenic copy number variations (CNVs) were detected in 17% of patients with developmental delays of unknown etiology, using chromosomal microarray testing of the whole human genome as a clinical diagnostic tool.[1] In 2010, this chromosomal microarray testing was recommended as a first-tier diagnostic tool for patients with developmental delays of unknown etiology.[2] Currently, this method is widely used to detect disease-causing CNVs worldwide.

Next-generation sequencing (NGS) is another prevalent method that can be used to obtain a large number of short-read nucleotide sequences with high power. NGS can be used as a clinical diagnostic tool in patients suspected of having single gene disorders. Thus, whole-exome sequencing (WES) has become a standard application for the detection of gene mutations responsible for human disease, especially single-nucleotide variations (SNVs) and small insertions/deletions (indels).[3]

Currently, researchers are unable to reach a consensus on the most effective clinical diagnostic tool for the first-tier examination of patients with developmental delays of unknown etiology. Efforts are being made to use WES data to detect CNVs to overcome this dilemma, and some studies have suggested the predominant use of an exome-first approach.[4–9] However, the possible use of lower density exome data extracted by specifically targeting only disease-related genes for the screening of CNVs needs to be confirmed.

In this study, clinical exome sequencing was performed for 61 patients with developmental delays of unknown etiology and 24 possible disease-causing SNVs were detected; some of the novel SNVs were reported elsewhere.[10–14] The remaining samples, which showed no possible candidate variants, might contain pathogenic CNVs. Thus, these samples were analyzed using NGS data extracted with a targeted re-sequencing approach. Here, we will discuss the possible use of NGS data for the detection of pathogenic CNVs as a NGS-first approach.

MATERIAL AND METHODS

Samples and grouping

This study was performed in accordance with the Helsinki declaration. In addition, the requisite permissions were obtained from the ethical committee of Tokyo Women's Medical University.

A total of 61 patients with developmental delays of unknown etiology were enrolled in this study. All patients were clinically evaluated by neuropediatricians at the institutions. All patients were subjected to conventional G-banding prior to this study. Written informed consent was obtained from the families of all patients after comprehensive genetic counseling regarding the appropriate methods of dealing with genetic information and possible incidental findings. Blood samples were collected from all patients and their parents. Samples from 34/61 patients were previously analyzed by chromosomal microarray testing, which did not identify definite disease-causing CNVs. Therefore, samples from these patients were analyzed by NGS and enrolled in the preliminary study (designated the microarray-first approach) to validate the eXome Hidden Markov Model (XHMM) system constructed in this study. Thirty samples that did not show CNVs > 200 kb were used for normalization. The remaining four patients possessed five CNVs that were over 200 kb in length; however, these CNVs were considered non-pathogenic and were used as positive controls. The remaining 27 patients were first analyzed by NGS and then subsequently enrolled in a non-biased and blinded study (designated the NGS-first approach). All samples tested by the NGS-first approach were retrospectively re-analyzed by chromosomal microarray testing. Consequently, all 61 patients were analyzed by both microarray and NGS (Table 1).

A total of 52 samples from healthy parents were used for a family-based trio-analysis. Four parental samples for patients #2 and #3 were analyzed using the microarray-first approach; the parental samples of patient #3 were also analyzed by NGS. The remaining 48 samples obtained from the healthy parents were analyzed by NGS. The sample obtained from patient #6's mother was re-analyzed by microarray. The numbers of samples analyzed by each method are summarized in Table 1.

Genetic analysis methods
Chromosomal microarray testing was performed using the Agilent Human Genome 60 K Array (Agilent Technologies, Santa Clara, CA, USA) as previously described.[15] Particular attention was paid to CNVs > 200 kb in length because these CNVs were most likely to be detected and most CNVs < 200 kb were familial and benign.[9] CNVs excluding TruSight One panel (Illumina, San Diego, CA, USA) genes cannot be detected by XHMM; therefore, five CNVs including TruSight One panel genes were selected as positive controls for the XHMM even though they were considered non-pathogenic CNVs. Upon the identification of suspected disease-causing CNVs, the parental samples were analyzed by chromosomal microarray testing and/or fluorescence *in situ* hybridization (FISH) using human bacterial artificial chromosomes as probes as

previously described.[15,16] Metaphase spreads were prepared from peripheral blood lymphocytes using standard methods.

Targeted exome sequencing was performed using the TruSight One sequencing panel (Illumina) comprising 125,396 probes to capture 11,946,514-bp targeted exon regions. These exon regions consisted of 4,813 genes that might be associated with known clinical phenotypes based on the Online Mendelian Inheritance in Man database (OMIM; http://www.omim.org/). The sequence library was constructed using 50 ng of genomic DNA, and the 151-bp paired-end reads were sequenced using the MiSeq next-generation sequencer (Illumina) as previously described.[10–14] The extracted data were mapped to a reference genome using the BWA Enrichment v1.0 cloud software (Illumina). GRCh37/hg19 was used as the reference sequence in this study. On average, we obtained 2.6 Gb of targeted aligned sequences and a mean coverage depth of 105.3 (target coverage at $20 \times$: 95.1%). The extracted variants were annotated and filtered using the Variant Studio software (Illumina). Candidate SNVs that could be related to the clinical features of patients with developmental delays were further analyzed through a family-based approach using parental samples.

The XHMM was used to quantify copy numbers throughout the genome.[4,7,9] Briefly, BAM files generated from the GATK data-processing pipeline were accumulated and the depth of coverage was calculated. The mean coverage of the targeted regions was obtained and normalized using principal component analysis. The principal component analysis-normalized depths were filtered and z-scores were obtained for these depths. The calculated data were validated using negative samples (30 BAM files from patients who showed no significant CNVs in previous chromosomal microarray testing and 15 BAM files from their healthy parents; Table 1). BAM files from patients #1, #2, #3 and #4 and from the mother of patient #3 were used as the positive controls. Twenty-seven BAM files from patients and 34 BAM files from their healthy parents subjected to the NGS-first approach were used for the XHMM analysis in addition to the BAM files subjected to the microarray-first approach (Table 1). The final CNV results were visualized in the primary genomic structure as a graph.

RESULTS
Microarray-first approach
The microarray-first approach was used for 34 patients. Ten CNVs > 200 kb in size were detected in eight patients. Among these, five CNVs did not include genes targeted by the TruSight One panel (indicating non-pathogenic CNVs). Therefore, the remaining five CNVs detected in four patients were listed as positive controls for the XHMM (Table 2).

Two patients (patients #2 and #3) showed relatively large CNVs; these CNVs were initially suspected to be disease-causing CNVs. However, both CNVs were confirmed to be inherited from healthy parents and thus were considered non-pathogenic CNVs. Briefly, the microarray result in patient #2 was arr 10q11.22q11.23 $(46,964,973-51,595,050) \times 3$ (Figure 1a). The parental samples of patient #2 were analyzed by microarray and the 10q11.22q11.23 duplication was identified in the father. This result suggested that this CNV was familial and not related to the developmental delay observed only in patient #2, although recurrent occurrence of reciprocal deletions/duplications of this region was previously reported.[17] The microarray result in patient #3 was arr 10q22.3q23.2 $(81,697,501–88,517,433) \times 1$ (Figure 1b). The 10q22.3q23.2 deletion in patient #3 was confirmed by FISH (Figure 2a). The parental FISH analyses showed a deletion in the mother, indicating a parental origin. The 10q22.3q23.2 deletion in the mother was retrospectively re-confirmed by microarray (Table 2). DNA samples from patients #2 and #3 and patient #3's parental samples were further analyzed by NGS; however, this

Table 1. Number of samples and detected CNVs

	Patients	Parents
Microarray-first approach		
Microarray only	0 (0)	2 (1)
Microarray, then NGS	34 (4)	2 (1)
NGS only	0 (0)	14 (0)
NGS-first approach		
NGS only	0 (0)	33 (0)
NGS, then microarray	27 (2)	1 (1)
Total	61 (6)	52 (3)

Abbreviations: CNV, copy number variation; NGS, next-generation sequencing.
Number of subjects analyzed by each approach is shown. Parentheses indicate the number of detected CNVs.

Table 2. Comparison of the CNV data extracted from the microarray and XHMM

Chromosome regions	Loss/gain	Microarray			XHMM			Number of genes[a]	
		Start	Stop	Size	Start	Stop	Size	Included in TruSight One	Included in the region
Microarray-first approach									
Patient #1 8p23.1	Gain	3,710,810	5,922,013	2,211,203	3,855,414	4,495,090	639,676	1	1
Patient #2 10q11.22q11.23	Gain	47,011,584	51,805,020	4,793,436	48,381,894	51,562,407	3,180,513	5	38
(father) 10q11.22q11.23	Gain	47,148,490	51,579,159	4,430,669	NA	NA	NA		
Patient #3 10q22.3q23.2	Loss	81,697,501	88,517,433	6,819,932	81,697,600	88,492,743	6,795,143	8	21
(mother) 10q22.3q23.2	Loss	81,697,501	88,517,433	6,819,932	81,697,600	88,492,743	6,795,143	8	21
Patient #4 20p13	Gain	3,182,144	3,724,665	542,521	3,193,804	3,660,228	466,424	3	6
20p13	Gain	4,534,383	4,937,261	402,878	Not detected			2	5
NGS-first approach									
Patient #5 5p15.33p15.2	Loss	151,737$	12,748,960	12,597,223	223,586$	10,465,066	10,241,480	11	47
14q32.12q32.33	Gain	93,705,209	107,089,189#	13,383,980	94,750,291	106,322,333#	11,572,042	20	121
Patient #6 16p11.2	Gain	29,820,221	30,105,987	285,766	29,802,070	30,102,524	300,454	7	21
(mother) 16p11.2	Gain	29,820,221	30,105,987	285,766	29,802,070	30,102,524	300,454	7	21

Abbreviations: CNV, copy number variation; NA; not analyzed; XHMM, eXome Hidden Markov Model.
Theoretically, $ should be 1 and # should be 107,349,540, because these CNVs include the terminal region.
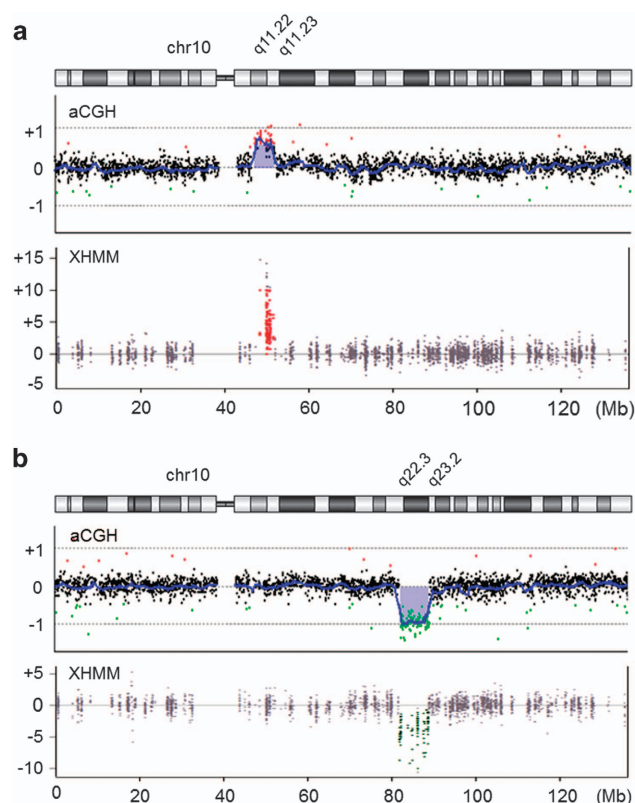[a]Number of genes detected by microarray analysis.

Figure 1. Genomic copy number variants detected by the microarray-first approach. (**a**) A gain of 10q11.22q11.23 is shown by microarray (upper) and XHMM (bottom). (**b**) A loss of 10q22.3q23.2 is shown by microarray (upper) and XHMM (bottom). Horizontal axes indicate physical positions of chromosome 10. Vertical axes indicate signal $\log_2$ ratio for microarray (upper) and z-scores for XHMM (bottom). The results of microarray are visualized in Gene View, created by the Agilent Genomic Workbench v.6.5 (Agilent Technologies). XHMM, eXome Hidden Markov Model.

analysis yielded no possible disease-causing SNVs. The XHMM analysis of these samples detected CNVs that were almost identical to those identified by the microarray (Figure 1a and b, Table 2).

Although one CNV initially identified by the microarray in patient #4 was successfully detected by XHMM, the other CNV in patient #4 was not detected. The CNV identified in patient #1 was detected by XHMM; however, the calculated size (640 kb) was significantly different from the size calculated by the microarray (2.2 Mb; Table 2).

NGS analysis of the 34 patients yielded 16 possible candidate SNVs, some of which were previously reported.[10–14]

NGS-first approach

Twenty-seven patients were analyzed by the NGS-first approach; possible disease-causing SNVs were identified in eight of these patients. All 27 BAM files extracted via NGS were analyzed by XHMM, and two samples from patients #5 and #6, in whom no possible disease-causing SNVs were identified through the NGS analysis, showed possible CNVs larger than 200 kb. Patient #5 showed a genomic copy number loss at 5p and a gain at 14q from which an unbalanced translocation between 5p and 14q was suspected (Figure 3a). These findings were re-confirmed by microarray as arr 5p15.33p15.2(1-12,748,960) × 1,14q32.12q32.33 (93,705,209–107,349,540) × 3 (Figure 3a). FISH confirmed the unbalanced translocation der(5)t(5;14)(p15.2;q32.12) (Figure 2b). The parental examination showed no translocation, indicating a *de novo* occurrence in the proband. Patient #6 showed a genomic copy number gain in the 16p11.2 region (Figure 3b). Her mother showed the same result (Figure 3b), indicating a possible familial duplication of 16p11.2. The microarray result in this mother and daughter was arr 16p11.2(29,820,221–30,105,987) × 3 (Figure 3c). This result was almost identical to the XHMM result.

Samples from the remaining 25 patients that did not show any CNVs in the XHMM analysis were re-analyzed by microarray, but no CNVs larger than 200 kb were detected. This finding indicates that the presence or absence of pathogenic CNVs was not misdiagnosed by XHMM (no false-negative or false-positive detection of pathogenic CNVs); thus, the detection ratio for the pathogenic CNVs was 100%.
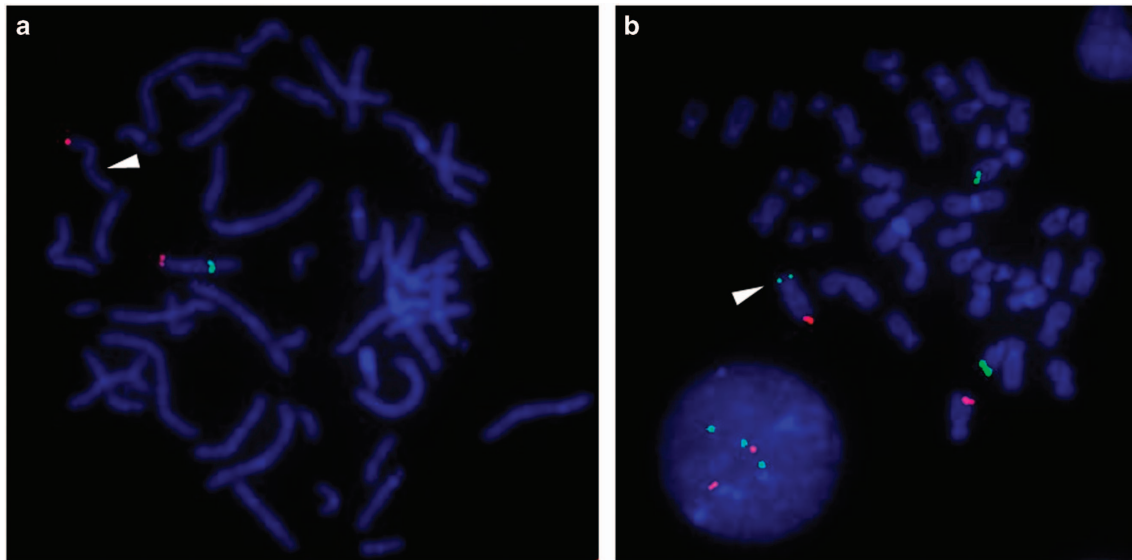
Figure 2. Results of fluorescence *in situ* hybridization (FISH) analyses. (**a**) Loss of the green signal (a white arrowhead) labeled for RP11–185K11 (10q23.1: 84,628,854–84,778,657) indicates a deletion in patient #3. Red signals labeled for RP11-387K19 (10p15.3: 322,071–162,974) are markers of chromosome 10. (**b**) An unbalanced translocation between 5p and 14q is confirmed in patient #5 by an additional green signal (a white arrowhead) labeled for RP11-379F22 (14q32.33: 106,920,250–107,014,205) on chromosome 5, indicated by a red signal labeled for RP11-260C12 (5q35.2: 174,674,775–174,854,980).
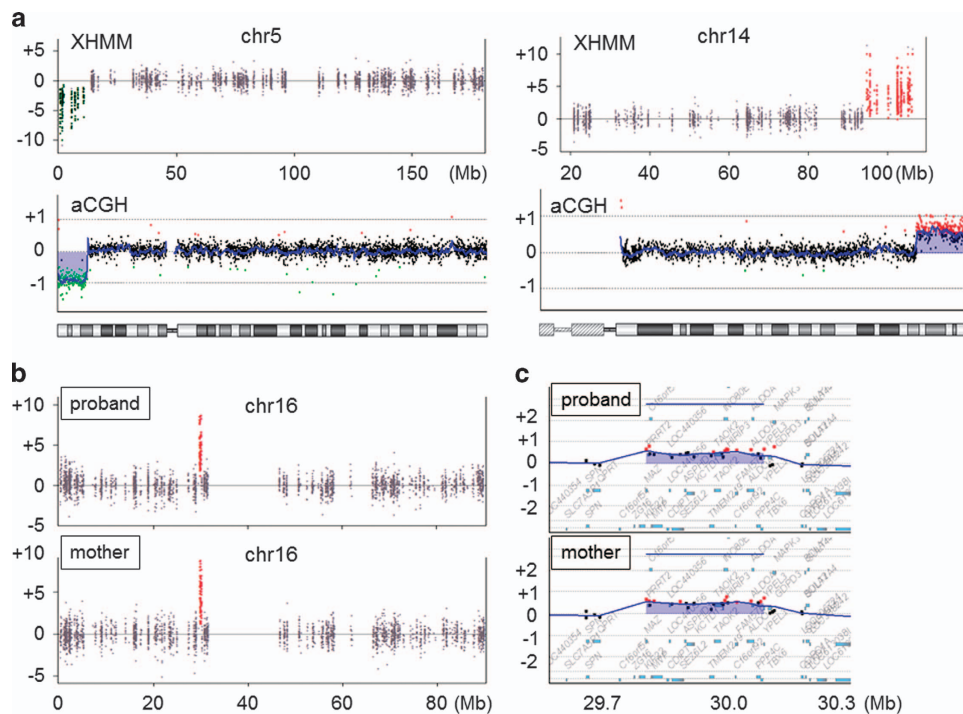

Figure 3. Genomic copy number variants detected by the XHMM-first approach. (**a**) XHMM analysis shows genomic copy number loss and gain in the terminal region of 5p (left) and 14q (right), respectively (upper). Similar patterns are re-confirmed by microarray (lower). (**b**) XHMM shows a gain in 16p11.2 in both the proband and her mother. (**c**) Duplicated 16p11.2 region, observed in both the proband and her mother, is expanded and visualized in the Gene View (Agilent Genomic Workbench; Agilent Technologies). Horizontal axes indicate physical positions of the chromosomes. Vertical axes indicate signal $\log_2$ ratio of the microarray and z-scores for XHMM. XHMM, eXome Hidden Markov Model.

Patient reports

The clinical features of two patients (patients #2 and #3) who showed relatively large CNVs in the microarray-first approach and two patients (patients #5 and #6) who were diagnosed as having chromosomal aberrations by the NGS-first approach are described below.

A 16-month-old girl (patient #2) was born with a birth weight of 2594 g, a length of 49 cm, and an occipitofrontal circumference of 31 cm, indicating microcephaly. The patient presented mild developmental delay and her brain magnetic resonance imaging showed reduced brain volume.

A boy aged 5 years and 7 months (patient #3) was delivered by Cesarean section with a birth weight of 2430 g. He started to walk at 2 years; however, he was incapable of forming meaningful words even at this age, indicating a severe developmental delay.

A 1-year-old girl (patient #5) was born with a birth weight of 1,282 g, a length of 38 cm, and an occipitofrontal circumference of 29 cm. From early infancy, she showed severe wheezing due to laryngomalacia. This patient presented multiple congenital anomalies, including a high-arched plate, micrognathia and large thumbs. Based on these findings, she was suspected of having Rubinstein-Taybi syndrome.[18] Congenital cataracts and bilateral moderate deafness were also noted. The severe laryngomalacia necessitated a tracheostomy. At 8 months of age, this patient was 55.5 cm (−5.8 s.d.) long, weighed 4508 g (−4.8 s.d.), and presented an occipitofrontal circumference of 38.5 cm (−2.7 s.d.), indicating a severe growth delay. She showed a severe developmental delay, with a developmental quotient of 28 evaluated by the Enjoji developmental test. Conventional G-banding showed a normal female karyotype of 46,XX. Retrospective evaluation facilitated the identification of a rounded face; however, the severe wheezing drowned out the mewing cry representative of cri du chat syndrome.[19]

A 10-year-old girl (patient #6) was born at 41 weeks of gestation with a birth weight of 3,240 g and an occipitofrontal circumference of 33 cm. She showed normal development in early infancy. However, at 3 years and 10 months, she suffered from partial seizures on her right side. Subsequently, complex partial seizures, secondary generalized seizures, and atypical-absence seizures were observed. After the occurrence of the seizures, the patient presented a decline in her language ability. Brain magnetic resonance imaging at 4 years showed no abnormality. The interictal electroencephalogram showed continuous slow waves and spikes. The seizures were controlled and no EEG abnormalities were observed after prescription of anti-epileptic drugs. At 10 years, the Wechsler Intelligence Scale for Children-III measured her intelligence quotient (IQ) as 68, indicating a mild intellectual disability. At present, the patient shows no abnormal neurological findings or abnormal behavior.

DISCUSSION

In this study, a system was constructed to detect CNVs using BAM files extracted from the GATK pipeline after targeted re-sequencing, which was initially aimed to identify disease-causing SNVs in patients with developmental delays of unknown etiology. The CNVs detection was accomplished using XHMM as a statistical tool. The term 'whole-exome' generally covers a region of approximately 58 Mb. In comparison, the TruSight One sequencing panel (Illumina) used in this study targeted regions of ~ 12 Mb; therefore, the region targeted by this panel was approximately 21% of the 'whole-exome' sequence. This limitation raised a concern as to whether the exome data extracted from this sparse target region could be sufficient to screen disease-causing CNVs.

The TruSight One sequencing panel (Illumina) was constructed for the detection of disease-causing SNVs. For this purpose, this panel targets approximately 5,000 genes associated with known clinical disease phenotypes based on the Online Mendelian Inheritance in Man (OMIM; http://www.omim.org/). Indeed, we identified 24 possible disease-causing SNVs in 61 patients with developmental delays of unknown etiology, some of which were reported elsewhere.[10–14] Therefore, the genes selected by this panel were strongly relevant to Mendelian diseases. This finding led to the consideration that CNVs in the genetic regions targeted by the TruSight One sequencing panel (Illumina) could have disease-causing traits in most cases.

In the preliminary study (microarray-first approach), we used five positive-control CNVs in four samples (these CNVs were previously identified by the microarray-first approach in these samples) (Table 2). Among these samples, four CNVs were successfully detected (detection ratio 80%), whereas the remaining CNV (in patient #4) was not detected by XHMM. This undetected CNV region (chr20:4,534,383–4,937,261) included two TruSight One genes (PRNP and PRND); however, these two genes consisted of only two exons each. The CNV identified by the microarray in patient #1 was detected by XHMM; however, the calculated size of this CNV was significantly different. Only 1 gene (CSMD1) in this CNV region (chr8:3,710,810–5,922,013) was targeted by the TruSight One panel, and only the first 5 exons of the total of 70 exons of CSMD1 were included in the CNV region. Therefore, the small range of the targeted exons in the CNV regions could be attributed to their lack of detection by XHMM. In comparison, three CNVs were successfully identified because a sufficient number of exons in the CNV regions were targeted by the TruSight One panel (Illumina). Although the lengths of the CNVs identified by both microarray analysis and XHMM were not identical to one another, the accuracy of the analysis depended on the number or density of the exons included in the TruSight One panel within the region. Consequently, this preliminary microarray-first approach showed a high detection ratio by XHMM, which encouraged us to perform the subsequent NGS-first approach.

Disease-causing chromosomal abnormalities were successfully identified in two patients using the non-biased and blinded serious study with the NGS-first approach. The ~ 300 kb-long 16p11.2 duplication was successfully identified in patient #6. Despite its short length, this region is gene-rich and contains seven of the genes included in the TruSight One panel. This finding may explain why XHMM clearly revealed this small duplication in both patient #6 and her mother. The z-score data calculated by XHMM did not reflect the exact copy number and could not be used to determine whether the suggested gain of this region was due to a duplication or triplication. Re-confirmation by microarray was necessary for the final diagnosis. Recently, the 16p11.2 duplication was reported as a risk factor for Rolandic epilepsy.[20] Patient #6 showed continuous slow waves and spikes, which might be related to the 16p11.2 duplication in this patient.

Patient #5 was initially suspected of having a single gene disorder (Rubinstein-Taybi syndrome) based on her clinical features. The samples obtained from this patient were preferentially subjected to NGS analysis as opposed to chromosomal microarray testing. The NGS data showed no disease-causing SNVs; however, the XHMM results suggested a possible unbalanced translocation with der(5)t(5;14), which was confirmed by the FISH analysis (this unbalanced translocation was misdiagnosed during conventional karyotyping because of similar G-band patterns in both telemetric regions). A retrospective clinical evaluation of patient #5 suggested the lack of any contradiction; thus, the majority of the clinical features in patient #5 were compatible with those often observed in patients with the 5p- syndrome, although many features were modified by the 14q duplication.[19]

The 10q22.3q23.2 deletion in patient #3 was initially considered a benign CNV because the same deletion was identified in the healthy mother. Therefore, the existence of an unmasked mutation in the homologous allele in the 10q22.3q23.2 region was suspected. However, this hypothesis was negated by the subsequent NGS analyses. van Bon et al.[21] reported two cases with familial 10q22.3q23.3 deletions. Finally, we concluded that this deletion might be rather pathogenic in patient #3.

The TruSight One panel was developed with the aim of clinical sequencing. Consequently, this panel targets clinically relevant SNVs. In this study, BAM files extracted through NGS using the TruSight One panel were used to detect clinically relevant CNVs. XHMM was used for this purpose because it focused on

identifying rare CNVs with a population frequency < 5%.[4] A pre-normalization step is performed in XHMM to increase the homogeneity of the samples prior to the principal component analysis, which removes extremely variable targets unrelated to CNVs.[8,9] Thus, CNVs detected via XHMM are prone to be pathogenic. The ability of the NGS-first approach to detect pathogenic CNVs was comparable to that of chromosomal microarray testing using a 60k platform, which was consistent with the results of previous reports on WES.[5,6]

In conclusion, pathogenic CNVs were successfully identified by XHMM using BAM files extracted through disease-related exome panel sequencing. The NGS-first approach yielded no false-negative or false-positive values, indicating the high sensitivity and high specificity of this method for the detection of pathogenic CNVs. Although the sample size was not sufficient to reach a final conclusion, the results of this study indicated that XHMM combined with the targeted exome data covering a 12 Mb region could be used as a first-tier screening approach for the identification of both SNVs and CNVs. The identification of possible pathogenic CNVs by the NGS-first approach should be re-confirmed by microarray as the final diagnosis.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

1 Yamamoto T, Shimojima K, Shimada S, Yokochi K, Yoshitomi S, Yanagihara K et al. Clinical impacts of genomic copy number gains at Xq28. Hum Genome Var 2014; 1: 14001.

2 Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010; 86: 749–764.

3 Yamamoto T, Shimojima K, Sangu N, Komoike Y, Ishii A, Abe S et al. Single nucleotide variations in CLCN6 identified in patients with benign partial epilepsies in infancy and/or febrile seizures. PLoS ONE 2015; 10: e0118946.

4 Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet 2012; 91: 597–607.

5 de Ligt J, Boone PM, Pfundt R, Vissers LE, Richmond T, Geoghegan J et al. Detection of clinically relevant copy number variants with whole-exome sequencing. Hum Mutat 2013; 34: 1439–1448.

6 Hayes JL, Tzika A, Thygesen H, Berri S, Wood HM, Hewitt S et al. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. Genomics 2013; 102: 174–181.

7 McCallum KJ, Wang JP. Quantifying copy number variations using a hidden Markov model with inhomogeneous emission distributions. Biostatistics 2013; 14: 600–611.

8 Kadalayil L, Rafiq S, Rose-Zerilli MJ, Pengelly RJ, Parker H, Oscier D et al. Exome sequence read depth methods for identifying copy number changes. Brief Bioinform 2015; 16: 380–392.

9 Miyatake S, Koshimizu E, Fujita A, Fukai R, Imagawa E, Ohba C et al. Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov Model: an 'exome-first' approach. J Hum Genet 2015; 60: 175–182.

10 Yamamoto T, Shimojima K. A novel MED12 mutation associated with non-specific X-linked intellectual disability. Hum Genome Var 2015; 2: 15018.

11 Yamamoto T, Shimojima K, Shibata T, Akiyama M, Oka M, Akiyama T et al. Novel PLA2G6 mutations associated with an exonic deletion due to non-allelic homologous recombination in a patient with infantile neuroaxonal dystrophy. Hum Genome Var 2015; 2: 15048.

12 Shimojima K, Okamoto N, Yamamoto T. A novel TUBB3 mutation in a sporadic patient with asymmetric cortical dysplasia. Am J Med Genet A 2016; 170: 1076–1079.

13 Yamamoto T, Igarashi N, Shimojima K, Sangu N, Sakamoto Y, Shimoji K et al. Use of targeted next-generation sequencing for molecular diagnosis of craniosynostosis: Identification of a novel de novo mutation of EFNB1. Congenit Anom (Kyoto) 2016; 56: 91–93.

14 Oka M, Shimojima K, Yamamoto T, Hanaoka Y, Sato S, Yasuhara T et al. A novel HYLS1 homozygous mutation in living siblings with Joubert syndrome. Clin Genet 2016; 89: 739–743.

15 Yamamoto T, Wilsdon A, Joss S, Isidor B, Erlandsson A, Suri M et al. An emerging phenotype of Xq22 microdeletions in females with severe intellectual disability, hypotonia and behavioral abnormalities. J Hum Genet 2014; 59: 300–306.

16 Yamamoto T, Shimojima K, Nishizawa T, Matsuo M, Ito M, Imai K. Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. Am J Med Genet A 2011; 155A: 113–119.

17 Stankiewicz P, Kulkarni S, Dharmadhikari AV, Sampath S, Bhatt SS, Shaikh TH et al. Recurrent deletions and reciprocal duplications of 10q11.21q11.23 including CHAT and SLC18A3 are likely mediated by complex low-copy repeats. Hum Mutat 2012; 33: 165–179.

18 Yamamoto T, Kurosawa K, Masuno M, Okuzumi S, Kondo S, Miyama S et al. Congenital anomaly of cervical vertebrae is a major complication of Rubinstein-Taybi syndrome. Am J Med Genet A 2005; 135: 130–133.

19 Cerruti Mainardi, P. Cri du Chat syndrome. Orphanet J Rare Dis 2006; 1: 33.

20 Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM et al. 16p11.2 600 kb Duplications confer risk for typical and atypical Rolandic epilepsy. Hum Mol Genet 2014; 23: 6069–6080.

21 van Bon BW, Balciuniene J, Fruhman G, Nagamani SC, Broome DL, Cameron E et al. The phenotype of recurrent 10q22q23 deletions and duplications. Eur J Hum Genet 2011; 19: 400–408.