Copy number variations in Japanese children with autism spectrum disorder

Yui Sakamoto[a], Shuji Shimoyama[b,c], Tomonori Furukawa[b], Masaki Adachi[c,d], Michio Takahashi[c,d], Tamaki Mikami[c], Michito Kuribayashi[c,d], Ayako Osato[a], Daiki Tsushima[c], Manabu Saito[a,c], Shinya Ueno[b,c] and Kazuhiko Nakamura[a,c]

Objective **Although autism spectrum disorder (ASD) occurs worldwide, most genomic studies on ASD were performed on those of Western ancestry. We hypothesized ASD-related copy number variations (CNVs) of Japanese individuals might be different from those of Western individuals.**

Methods **Subjects were recruited from the Hirosaki 5-year-old children's developmental health check-up (HFC) between 2013 and 2016 (ASD group; $n = 68$, control group; $n = 124$). This study conducted CNV analysis using genomic DNA from peripheral blood of 5-year-old Japanese children. Fisher's exact test was applied for profiling subjects and CNV loci.**

Results **Four ASD-related CNVs: deletion at 12p11.1, duplications at 4q13.2, 8p23.1 and 18q12.3 were detected ($P = 0.015$, 0.024, 0.009, 0.004, respectively). Specifically, the odds ratio of duplication at 18q12.3 was highest among the 4 CNVs (odds ratio, 8.13).**

Conclusions **Four CNVs: microdeletion at 12p11.1, microduplications at 4q13.2, 8p23.1 and 18q12.3 were detected as ASD-related CNVs in Japanese children in this study. Although these CNVs were consistent with several reports by Western countries at cytoband levels, these did not consistent at detailed genomic positions and sizes. Our data indicate the possibility that these CNVs are characteristic of Japanese children with ASD. We conclude that Japanese individuals with ASD may harbor CNVs different from those of Western individuals with ASD.** ***Psychiatr Genet*** **31: 79–87 Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc.**

Psychiatric Genetics 2021, **31**:79–87

Keywords: autism spectrum disorder, copy number variations, intellectual disability, Japanese, 12p11.1, 4q13.2, 8p23.1, 18q12.3

Departments of [a]Neuropsychiatry, [b]Neurophysiology, [c]Research Center for Child Mental Development and [d]Department of Clinical Psychological Science, Hirosaki University Graduate School of Medicine, Hirosaki, Aomori, Japan

Correspondence to Kazuhiko Nakamura MD, PhD, Department of Neuropsychiatry, Graduate School of Medicine, Hirosaki University, 5 Zaifu-cho, Hirosaki, Aomori 036-8562, Japan
Tel: +81 172 39 5065; fax: +81 172 39 5067;
e-mail: nakakazu@hirosaki-u.ac.jp

Received 18 August 2020 **Accepted** 15 January 2021

Introduction

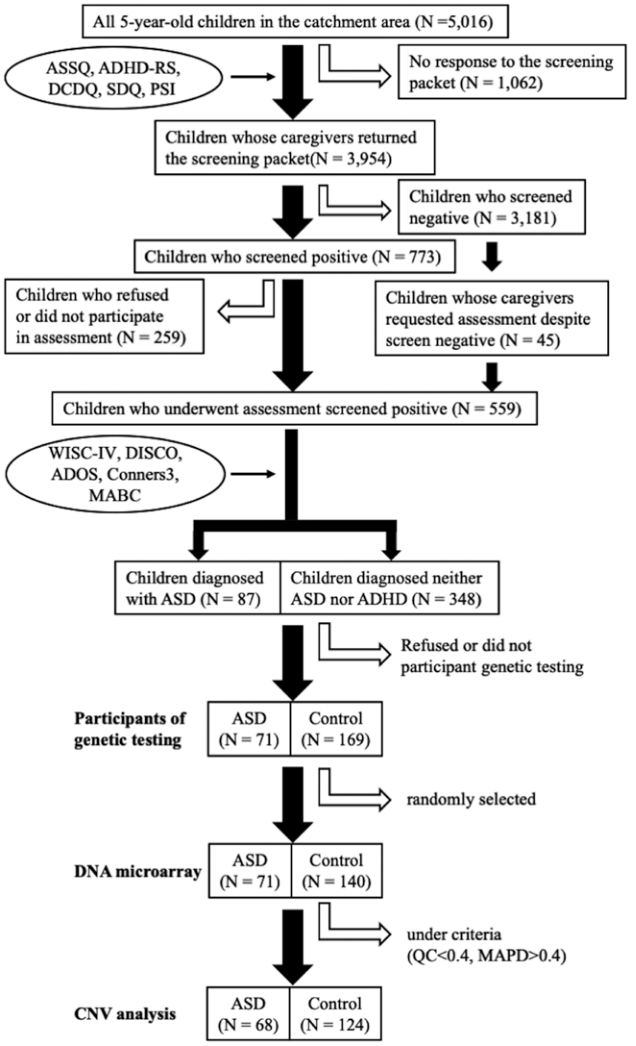
Autism spectrum disorder (ASD) is an innate neurodevelopmental disorder (NDD) characterized by impairments in social communication/interaction, repetitive and restrictive behavior (American Psychiatric Association, 2013). The US Centers for Disease Control and Prevention has reported that the overall ASD prevalence per 1000 children aged 4 years was 13.4 in 2010, 15.3 in 2012 and 17.0 in 2014, 15.6 in 2016 (Shaw *et al.*, 2020). We showed that adjusted ASD prevalence was 3.22% in Japanese 5-year-old children (Saito *et al.*, 2020). Many studies on ASD from various points of view, such as biochemistry, genetics and epidemiology, have been conducted; however, the etiology of ASD has not been clarified completely. Owing to unexplained etiology, therapeutics for ASD have not been established other than behavioral therapy.

The etiology of ASD is complicated due to its multifactorial nature, and both genetic factors and environmental factors are involved in the onset, phenotype and severity. This complexity is expressed as 'a complex dynamic of gene×environment effect' (Esposito *et al.*, 2017; Kikusui and Hiroi, 2017). Some twin studies showed that the concordance rate for ASD was 60–92% in monozygotic twins and 0–10% in dizygotic twins (Bailey *et al.*, 1995); it has also been suggested that approximately 50–60% of ASD etiologies are caused by genetic factors (Hallmayer, 2011; Gaugler *et al.*, 2014). In brief, this evidence indicates that genetic factors have a large impact on the etiology of ASD, conferring high heritability. Numerous studies suggest that chromosomal abnormalities such as copy number variation (CNV) and single-nucleotide polymorphism (SNP) contribute to an increased risk of ASD. According to a systematic review, the mean rate of cytogenetically

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.psychgenetics.com.

0955-8829 Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc. DOI: 10.1097/YPG.0000000000000276

Fig. 1



Flowchart of the HFC and CNV analysis. The schematic figure of this study was shown as described in Methods. HFC was carried out as reported previously (Saito *et al*., 2020). Random subjects were extracted; those who participated in the genetic test and those who fulfilled the criteria of microarray (quality check >0.4, MAPD <0.4) were subjected to CNV analysis. Finally, 68 children with ASD and 124 non-ASD/ADHD children were eligible for this study. ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; CNV, copy number variations; HFC, Hirosaki 5-year-old children developmental health check-up.

detectable chromosome abnormalities is 7.4% of ASD cases, with a range from 0–54% (Xu *et al*., 2004; Reddy, 2005).

The involvement of multiple genes, as opposed to a single gene, further complicates the etiology of ASD and makes it difficult to elucidate genetic effects. SFARI GENE (https://gene.sfari.org/. released on 20 June 2019) reports that 1089 genes and 2170 CNVs, including those with rare frequency, are related to ASD. Given that the number of human protein-coding genes is approximately 20 000 according to EMBL-EBI (https://www.ebi.ac.uk/) and NCBI (https://www.ncbi.nlm.nih.gov/), it is surprising that approximately 5% of all genes are related to ASD.

CNV is a structural variation in genomic sequence, including large multiplications and deletions ranging from 1 kilobase (kb) to several megabases (Mb) of DNA (Iafrate *et al*., 2004; Sebat *et al*., 2004). One of the representative CNVs in ASD is a maternal 15q11–q13 duplication that is found in 1–3% of individuals with the disorder (Cook *et al*., 1997). Indeed, this is the most common cytogenetic abnormality in ASD. It has also been indicated that 7–10% of individuals with sporadic ASD carry de novo CNVs (Sebat *et al*., 2004; Marshall *et al*., 2008). There are also many rare ASD-specific CNVs, for example, 2p16.3 deletion, 19q13.32 duplication and so on (Prasad *et al*., 2012). Not only CNVs including genes but also intergenic CNVs can cause changes in gene expression and function (Klopocki and Mundlos, 2011; Walker and Scherer, 2013); CNVs may confer phenotypic heterogeneity and different susceptibilities to disorders (Freeman *et al*., 2006). In addition, several studies have revealed differences in the number and size of CNVs between individuals with typical development and individuals with ASD (Sebat *et al*., 2007; Cho *et al*., 2009; Prasad *et al*., 2012). The number and size of CNVs may be related to genomic diversity and alter susceptibility to ASD because of unstable genomic structure. Thus, examining the number and size of CNVs is as important as identifying CNV loci.

Most genome-wide association studies have been carried out using samples of Western ancestry, including European and North American, even though ASD occurs worldwide (The Autism Genome Project Consortium, 2007; Pinto *et al*., 2010). Therefore, the genomic information associated with ASD thus far may differ from that of other ancestries. Asian areas, including Japan, are no exception, and there have been only a few studies regarding CNVs among Asian individuals with ASD. Moreover, several studies have reported some differences in CNVs among different ethnicities; as an example, Melanesian and Papuan individuals harbor more CNVs than other populations, with Kalash individuals harboring fewer CNVs than the average (Itsara *et al*., 2009). We hypothesized that ASD-related CNVs of Japanese individuals may be different from those of Western individuals.

To examine differences in ASD-related CNVs between those of Japanese ancestry and Western ancestry, this study conducted CNV analysis using genomic DNA from peripheral blood of Japanese children.

Methods

Ethics approval

This study was approved by the Research Ethics Committee of Hirosaki University Graduate School of Medicine (approval number: 2013-293). All subjects were recruited with informed consent. The information security policies of Hirosaki city were followed to protect the personal data of the participants.

Study design of health check-up study

The flow chart shows our study design (Fig. 1). Subjects were recruited from the Hirosaki 5-year-old children developmental health check-up (HFC) between 2013 and 2016. The HFC has been held annually in Hirosaki, a city located in Aomori prefecture in Japan, and targeted all 5-year-old children in the city. The city is approximately 524.20 km^2 in size, and the total population was 175 777 as of 2016. The HFC consisted of two stages, the initial screening stage and the comprehensive assessment stage, as reported previously (Adachi *et al.*, 2018; Saito *et al.*, 2020).

In the initial screening stage, we used the following measures to screen for ASD: the Japanese version of autism spectrum screening questionnaire (ASSQ) (Ehlers and Gillberg, 1993; Adachi *et al.*, 2018), the Japanese version of the strengths and difficulties questionnaire (SDQ) (Goodman, 1997; Matsuishi *et al.*, 2008) and the Japanese version of parenting stress index (PSI) (Abidin, 1995; Narama *et al.*, 1999). We also used attention deficit hyperactivity disorder (ADHD)-rating scale-IV (ADHD-RS-IV) (DuPaul *et al.*, 1998; Takayanagi *et al.*, 2016) as a screening tool for ADHD, and developmental coordination disorder questionnaire (DCDQ) (Nakai *et al.*, 2011; Wilson et al., 2000, 2009) as a screening tool for developmental coordination disorder (DCD).

We next invited screening-positive children to a comprehensive assessment at the medical check-up in Hirosaki city; screening-negative children also participated in the assessment as applicants if their caregiver requested an evaluation. We used diagnostic interview for social and communication disorders (DISCO) to collect information about behavior related to NDD (Wing *et al.*, 2002). Psychologists used the Japanese version of the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV), to assess cognitive function (Wechsler and Japanese WISC-IV Publication Committee, 2010). Children with ASD or suspected ASD were administered autism diagnostic observation schedule (Gotham *et al.*, 2009).

The multidisciplinary team consisted of child and adolescent psychiatrists, psychologists, pediatricians and occupational therapists diagnosed NDD based on screening tools and diagnostic assessment. Children with ASD were diagnosed using DSM-5 criteria (American Psychiatric Association, 2013). Children whose IQ was below 70 were diagnosed with intellectual disability.

Between 2013 and 2016, the 5016 children born in Hirosaki were eligible for the HFC. Caregivers of 3954 children (participation rate 78.8%) completed and returned the screening packet (ASSQ, ADHD-RS, DCDQ, SDQ and PSI). Preschool teachers completed the teacher-rated SDQ. Among children who returned the screening packet, 773 children (19.5%) were screen-positive using HFC original screening criteria (Saito *et al.*, 2020) and invited to the comprehensive assessment. Of these screen-positive children, caregivers of 259 children either refused to consent or did not bring their children for the assessment. For screen-negative children ($N = 3181$), caregivers of 45 children requested the assessment. Therefore, 559 children underwent the comprehensive in-person assessment. Of the 559 children, 87 children (60 boys and 27 girls) were confirmed to have ASDs (Saito *et al.*, 2020).

Subjects for copy number variation analysis

This CNV study targeted Japanese children who participated in the HFC and provided a voluntary blood sample. The subjects of blood sampling were only those who consented to this experiment. Of the 87 children diagnosed with ASD, 68 children (45 boys and 23 girls, mean age = 65.1 ± 3.1 months) were recruited for the ASD group. Of 348 children with neither ASD nor ADHD, 169 children provided a voluntary blood sample. Of 169 children, 140 children were randomly selected for CNV analysis. There were no twins or siblings in this study. The profiles of the subjects were blinded before CNV analysis and opened after the experiment to conduct statistical analysis. After quality control, 124 children (62 boys and 62 girls, mean age = 65.0 ± 3.0) were recruited for the control group. In both groups, children with other psychiatric disorders other than DCD and intellectual disability were excluded.

Extraction of genomic DNA

Peripheral blood from subjects was collected for extraction and purification from genomic DNA using a QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany) in accordance with the manufacturer's instructions. The quantity and quality (ratio of absorbance at 260 and 280 nm) of genomic DNA were estimated with a Microvolume Spectrophotometer Q5000 (TOMY SEIKO, Tokyo, Japan) and all samples showed a ratio between 1.8 and 2.0.

DNA microarray

Fifty nanograms of genomic DNA was used for a genotyping array using Genome-Wide Human SNP Array 6.0 (ThermoFisher Scientific, Waltham, Massachusetts, USA) in accordance with the manufacturer's instructions. This product contains approximately 1.8 million genetic markers that cover the whole genome, including more than 906 600 SNPs and more than 946 000 probes for the detection of CNVs. Genomic DNA (250 ng) was digested with Sty and Nsp. Fragmented and labeled DNA was hybridized at 50 °C and 60 rpm for 16 h using GeneChip Hybridization Oven 640 (Thermo Fisher Scientific). After washing and staining with GeneChip Fluidics Station 450 (Thermo Fisher Scientific), the array cartridges were scanned with a GeneChip Scanner 3000 7G System. After scanning the array cartridge, quality check was performed by Genotyping Console software. We ascertained whether all the array data fulfilled the following criteria: quality check >0.4 and median of the absolute values of all pairwise differences (MAPD) <0.4, as recommended quality check metrics according

Table 1 The profiles of the subjects

	Age (months)	Number of intellectual disability cases (%)
ASD ($N = 68$)	65.0 ± 3.0	23 (33.8%)*
Boys ($n = 45$)	64.7 ± 2.8	13 (28.9%)
Girls ($n = 23$)	65.5 ± 3.3	10 (43.5%)*
Control ($N = 124$)	65.1 ± 3.1	19 (15.3%)
Boys ($n = 62$)	65.0 ± 3.3	10 (16.1%)
Girls ($n = 62$)	65.2 ± 2.9	9 (14.5%)

Among the subjects who participated in HFC, CNV analysis was performed against 68 children with ASD and 124 children with neither ASD nor ADHD. The average age (month), the breakdown of boys and girls, and the presence or absence of intellectual disabilities were shown. Subjects with intellectual disabilities were included in this study. Asterisks indicate significant differences between control subjects of same sex or total (*$P < 0.01$); means ± SD.
ADHD, attention deficit hyperactivity disorder; ASD, autism spectrum disorder; CNV, copy number variations; HFC, Hirosaki 5-year-old children developmental health check-up.

Table 2 Number of CNVs (>1 kbp)

	Total	*P* value
ASD ($N = 68$)	37.4 ± 2.3	0.466
Boys ($n = 45$)	38.0 ± 2.8	0.528
Girls ($n = 23$)	36.8 ± 2.1	0.342
Control ($N = 124$)	37.8 ± 1.9	–
Boys ($n = 62$)	38.2 ± 2.0	–
Girls ($n = 62$)	37.4 ± 1.8	–

The numbers of CNVs (>1kbp) was shown. The *P* value indicates the result when compared with control subjects of same sex or total.
CNV, copy number variations.

to the manufacturer's instructions. Some array data were excluded from CNV analysis because they did not fulfill the criteria. All of the DNA microarray data that fulfilled the criteria were deposited at Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE144918.

Estimation of log R Ratio and B allele frequency

Both log R Ratio (LRR) and B allele frequency (BAF) were estimated using PennCNV. The LRR was the normalized values of the signal intensity for each CNV and SNP probe in array chips. After the normalization, copy numbers were estimated by signal intensity around 0 was set to two copies. Higher and lower values than 0 mean duplication and deletion, respectively. BAF refers to the fraction of reads supporting nonreference alleles at a given SNV position and estimated the patterns of allelic duplication or deletion. In the case of normal copy (copy number = 2), the values of BAF around 0, 0.5 and 1 indicate AA, AB and BB genotypes, respectively.

Copy number variation analysis

CNV analysis refers to the human genome version of GRCh37/hg19 because it contains probe information for Genome-Wide Human SNP Array 6.0 and more information than the latest version (GRCh38/hg38). For the estimation of copy number, PennCNV (Wang *et al.*, 2007) and Partek Genomics Suite 7.0 (Partek, St. Louis, Missouri, USA) were used with a hidden Markov model. Copy number = 2 was determined as the standard copy number, copy number = 3 or more was determined as duplication, and copy number = 1 or less was determined as deletion. We calculated the average number and size of CNVs (>1 kbp) in each group.

Copy number variation and gene loci

Based on the results of CNV analysis, CNV loci were confirmed by AutDB (http://autism.mindspec.org/autdb/html) (Pereanu *et al.*, 2018), SFARI GENE (https://gene.sfari.org/), ENCODE at UCSC Genome Browser (http://genome.ucsc.edu/ENCODE/) and Ensembl genome browser (http://www.ensembl.org/index.html). The lists of genes located within the CNV loci were confirmed by ENCODE at the UCSC Genome Browser.

Statistics

SPSS (version 25.0) was used for statistical analyses. Fisher's exact test was applied for profiling subjects and CNV loci. A *P* value of less than 0.05 was considered statistically significant.

Results

The profile of the subjects

The schematic figure of this study is shown in Fig. 1. The profile of the subjects in this study regarding age, sex and the comorbidity rate of intellectual disability are shown in Table 1. Since participants were recruited from the HFC, there was no difference in age between the ASD group and the control group (65.0 ± 3.0 and 65.1 ± 3.1, respectively). The comorbidity rate of intellectual disability was significantly higher in the ASD group than in the control group (33.8 and 15.3%, respectively; $\chi^2 = 8.79588$, $P = 0.0056$).

The average number of copy number variation

There was no significant difference in the average number of CNVs between the ASD group and the control group (Table 2, 37.4 ± 2.3 and 37.8 ± 1.9, respectively), as shown in Table 2. There were also no sex differences in the average number of CNVs.

Characteristic copy number variations in Japanese children with autism spectrum disorder

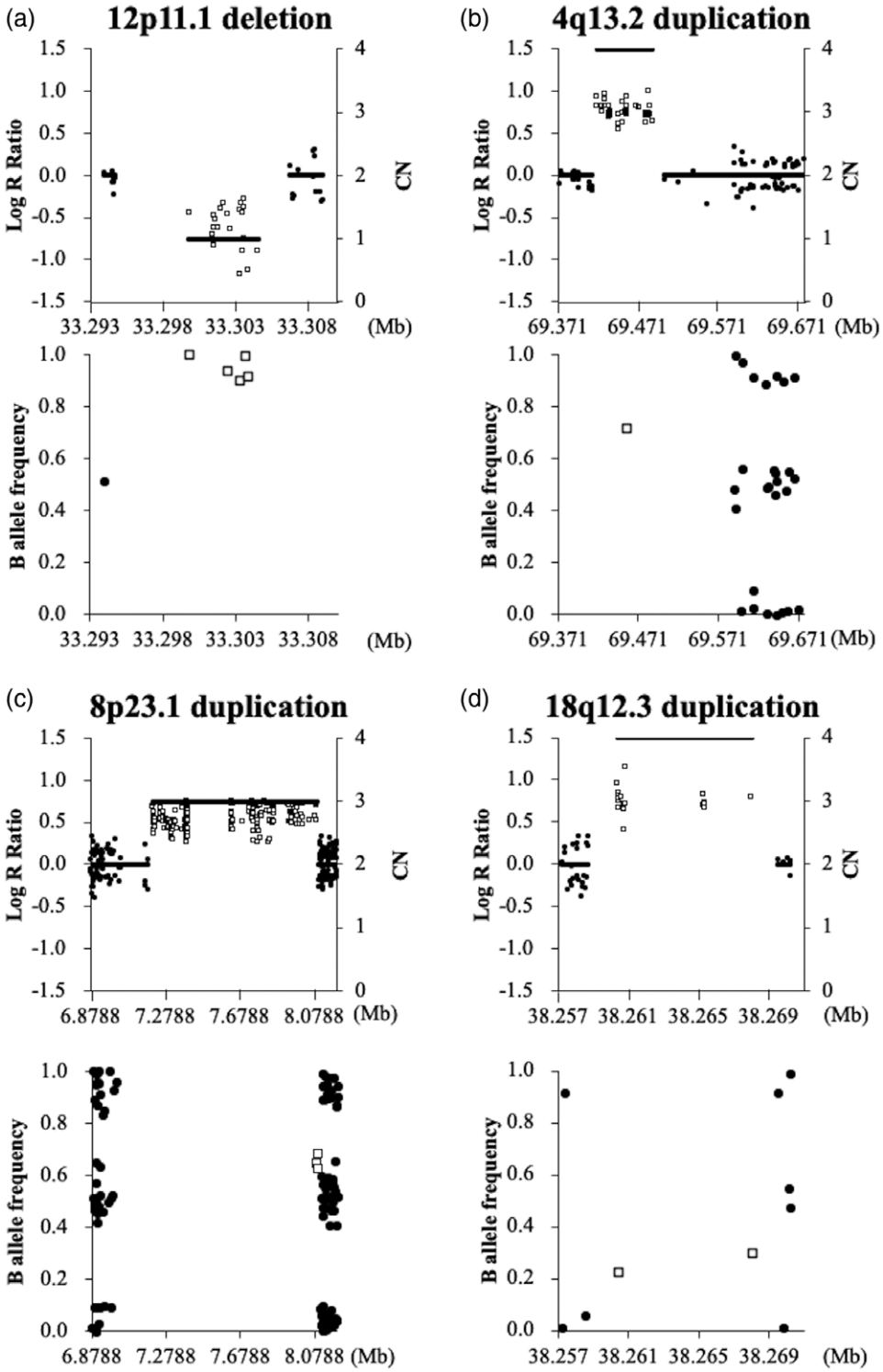
Four characteristic CNV loci among the Japanese children with ASD compared to the control group were identified: deletion at 12p11.1, duplications at 4q13.2, 8p23.1 and 18q12.3 (Table 3, Supplementary Figure A–D, Supplemental digital content 1, *http://links.lww.com/PG/A250*, Fisher's exact test $P = 0.015$, 0.024, 0.009, 0.004, respectively. odds ratios = 2.91, 3.11, 6.15, 8.13, respectively). The odds ratio of duplication at 18q12.3 was highest among the four CNVs (odds ratio = 8.13). Average sizes of deletion at 12p11.1, duplications at 4q13.2, 8p23.1 and 18q12.3 in the ASD group were 5621, 67 503, 487 929 and 5322 bp, respectively; all of these were microdeletion

Table 3 **Characteristic region of CNVs in children with ASD**

		Position		Average size (bp)		Copy number (*n*)									Possession (*n*, %)						
						ASD					Control				ASD		Control		*P* value		
Chromosome	Loci	Start	End	ASD	Control	0	1	2	3	4<	0	1	2	3	4<	Deletion	Duplication	Deletion	Duplication	Deletion	Duplication	Odds ratio
chr 12	p11.1	33 301 395 33 301 395	33 304 596 33 307 387	5621	5739	4	11	53	0	0	2	9	113	0	0	15 (22.1%)*	0 (0%)	11 (8.9%)	0 (0%)	0.015	–	2.91
chr 4	q13.2	69 420 012 69 420 012 69 420 012	69 484 122 69 485 992 69 489 348	67 503	68 076	0	0	56	0	12	0	0	116	0	8	0 (0%)	12 (17.6%)*	0 (0%)	8 (6.5%)	–	0.024	3.11
chr 8	p23.1	7 214 599 7 214 599 7 214 599 7 214 599 7 214 599 7 214 599 7 214 599 7 214 599	7 747 712 7 759 606 7 783 229 7 787 975 7 799 403 7 809 919 7 238 032 8 072 316 8 094 763	487 020	658 078	0	0	59	9	11	0	2	119	3	0	0 (0%)	9 (13.2%)*	2 (1.6%)	3 (2.4%)	–	0.009	6.15
chr 18	q12.3	38 260 408 38 260 408	38 265 402 38 268 024	5322	5169	8	0	52	1	7	11	2	109	0	2	8 (11.8%)	8 (11.8%)*	13 (10.5%)	2 (1.6%)	0.626	0.004	8.13

The detailed information of characteristic CNVs in children with ASD. The *P* values were calculated by Fisher's exact test.
ASD, autism spectrum disorder; CNVs, copy number variations.

Fig. 2



Illustrations of LRR and BAF values for each CNV in children with ASD. Representative values of LRR, BAF in 12p11.1 (a), 4q13.2 (b), 8p23.1 (c) and 18q12.3 (d) were shown with scatter plot. Copy number state was shown with line plot. The values of LRR and BAF in the CNV region were indicated by square, and their surrounding area were indicated by circle. The LRR values centered around zero mean copy number = 2. The BAF values around 0, 0.5 and 1.0 mean AA, AB and BB genotypes, respectively. BAF, B allele frequency; CNV, copy number variations; LRR, log R ratio.

Table 4 The list of genes within each CNV

Chromosome	Loci	Genes					Reference
chr 12	p11.1			–			Christian *et al.* (2008) Chen *et al.* (2017)
chr 4	q13.2			*UGT2B17*			Celestino-Soper (2011)
chr 8	p23.1	*ZNF705G*	*DEF4B*	*DEFB103B*	*SPAG11B*	*DEFB104B*	Autism Genome Project Consortium (2007)
		DEFB106B	*DEFB105B*	*DEFB107B*	*PRR23D1*	*FAM90A7P*	Marshall *et al.* (2008)
		FAM90A10P *DEFB104A* *USP17L*	*PRR23D2* *SPAG11A* *USP17L3*	*DEFB107A* *DEFB4A* *MIR548I3*	*DEFB105A* *ZNF705B* *FAM85B*	*DEFB106A* *FAM66E* *FAM86B3P*	Pinto *et al.* (2010) Prasad *et al.* (2012)
chr 18	q12.3			–			Autism Genome Project Consortium (2007) Prasad *et al.* (2012)

12p11.1 and 18q12.3 have no genes within detected CNV. 4q13.2 and 8p23.1 has several genes within detected CNV, but these genes have not been reported to be associated with ASD. References show previous studies that reported characteristic CNVs in patients with ASD at the cytoband level.
ASD, autism spectrum disorder; CNVs, copy number variations.

or microduplication. The representative data of the LRR and BAF of each CNVs in children with ASD are shown in Fig. 2.

Discussion

This study detected four ASD-related CNVs at 12p11.1, 4q13.2, 8p23.1 and 18q12.3. One out of four CNVs was microdeletion, and three out of four CNVs were microduplications.

Although these CNVs were consistent with several reports by Western countries at cytoband levels, these were not consistent at the detailed genomic positions and sizes as shown in Table 3 (Celestino-Soper *et al.*, 2011; Chen, 2017; Christian *et al.*, 2008; Marshall *et al.*, 2008; Pinto *et al.*, 2010; Prasad *et al.*, 2012; The Autism Genome Project Consortium, 2007). Many studies have reported that some microdeletion/microduplication caused characteristic clinical features, and these are called the microdeletion/microduplication syndrome (Weise *et al.*, 2012). For example, microdeletion at 22q11.2 is called Di George syndrome or CATCH 22. Here, we report that the characteristic CNVs possibly associated with pathogenesis were found with Japanese children with ASD.

We used ENCODE at the UCSC Genome Browser and SFARI GENE to assess whether genes that contained in the detected CNV loci are reported as ASD-related genes (Table 4, Supplementary Figure A–D, Supplemental digital content 1, *http://links.lww.com/PG/A250*). Table 4 is a list of genes contained in the four characteristic CNVs in Japanese children with ASD. Although duplications at 4q13.2 and 8p23.1 contained genes, none of them are classified as ASD-related genes in SFARI GENE. Deletion at 12p11.1 and duplication at 18q12.3 did not contain any gene, these were intergenic CNVs.

According to the guideline of interpreting the clinical relevance of CNV from the American College of Medical Genetics (Kearney *et al.*, 2011), CNVs are classified into pathogenic, uncertain clinical significance and benign. Moreover, CNVs classified into uncertain clinical significance further classified into likely pathogenic, likely benign and no subclassification. The CNVs at 4q13.2 and 18q12.3 we detected are classified into 'uncertain clinical significance; no subclassification', and CNVs at 12p11.1 and 18q12.3 are classified into 'uncertain clinical significance; likely benign'.

The mechanism of action of these intergenic CNVs, deletion at 12p11.1 and duplication at 18q12.3, in the pathogenesis of ASD could be (1) through altering the necessary copy number or positional context of key DNA sequence elements required for regulating the proper expression of nearby genes (Klopocki and Mundlos, 2011), (2) affecting still undiscovered genes or noncoding RNAs residing in the CNV regions and (3) disrupting uncharacterized isoforms of the adjacent annotated genes (Walker and Scherer, 2013).

We also searched frequencies of these four CNVs among Asian ancestry using gnomAD v2.1.1 (https://gnomad.broadinstitute.org/). The one deletion at 12p11.1 whose position (12:33296559-33307366) did not exactly match with our CNV but located nearby our CNV had been registered, and its frequency was 0.1229 among East Asian. Similarly, two duplications at 18q12.3 whose position (18:38221619-38291063, 18:38224748-38361013) did not exactly match with our CNV but located nearby our CNV had been registered; however, there were no East Asian individuals with these CNVs. The duplications at 4q13.2 and 8p23.1 whose position matched with our CNVs had not been registered in gnomAD.

In this study, we could not conclude that these CNVs are specific to Japanese individuals with ASD. Regarding three of the four CNVs we detected, many studies have reported associations between ASD and duplication of 4q13.2 (Celestino-Soper *et al.*, 2011), deletion-duplication of 8p23 (The Autism Genome Project Consortium, 2007; Marshall *et al.*, 2008; Pinto *et al.*, 2010), and duplication of 18q12.3 (The Autism Genome Project Consortium, 2007) (Table 4). These data suggest that these three CNVs

are common to Japanese ancestry with ASD as well as Western ancestry with ASD.

It is difficult to utilize four CNVs we detected for a diagnostic algorithm because 4.0–12.1% of the control group also harbored these CNVs (Table 3). Although genomic information such as CNVs may not be useful for diagnosing ASD because of the complexity, large differences between individuals and cost, these data may be helpful in elucidating the etiology of ASD.

Although we did not find a significant difference in the number of CNVs between the ASD group and the control group, some studies have reported that individuals with ASD have a significantly small average number of CNVs compared to control subjects (Prasad *et al.*, 2012). On the other hand, other studies have reported that individuals with ASD carry a large average number of CNVs compared to control subjects (Sebat *et al.*, 2007; Cho *et al.*, 2009). These opposite results may be caused by differences in sample size, analytic methods and reference genome version. Furthermore, another study showed that the number of CNVs depends on ancestry (Itsara *et al.*, 2009), and the difference in ancestry might be responsible for the opposite results and our data. Considering that we have not reached a consensus on the number of ASD-related CNVs, the current data are not contradictory to prior studies.

The strength of this study is its precision diagnosis through an appropriate process and focus on 5-year-old children. We could exclude acquired effects after birth and focus on the genetic background because this study was community-based and we targeted 5-year-old children. Our results suggest characteristics of CNVs specific to Japanese children with ASD.

This study has some limitations. First, the sample size was small. There is a possibility that characteristic CNVs reported by other studies were not detected in this study due to our small sample size. In other words, the four CNVs that we detected despite the small sample size may be meaningful. Second, some subjects were comorbid with other NDDs; therefore, we could not rule out the effects of them. In HFC, of all children with ASD, 88.5% of children were found to have at least one co-occurring NDD (i.e. one or more among ADHD, DCD, intellectual disability and/or borderline intellectual functioning). Common co-occurring conditions include ADHD 50.6%, DCD 63.2 %, intellectual disability 36.8% and borderline intellectual functioning 20.7% (Saito *et al.*, 2020). Other studies have also reported the co-occurring rate of intellectual disability in ASD ranges between 25 and 70% (Chakrabarti and Fombonne, 2001; Yeargin-Allsopp *et al.*, 2003). Particularly, we have not excluded children with intellectual disability, and 33.8% of children in the ASD group were comorbid with intellectual disability in this study. Considering these high co-occurring rates, it is impossible to completely separate the effects of other NDDs.

Conclusion

Four CNVs: microdeletion at 12p11.1, microduplications at 4q13.2, 8p23.1 and 18q12.3 were detected as ASD-related CNVs in Japanese children in this study. At the cytoband level, these CNVs were consistent with some studies on Western countries; however, these were not consistent at detailed locations and sizes. Our data indicate the possibility that these CNVs are characteristic of Japanese children with ASD. We conclude that Japanese individuals with ASD may harbor CNVs different from those of Western individuals with ASD.

Acknowledgements

The authors would like to thank subjects and their families for participating in this study. The authors also wish to express their gratitude to Sachiko Kamikawa for a technical assistant. This study was supported by the Hirosaki Institute of Neuroscience in Japan. We also would like to thank Springer Nature Author Services (https://authorservices.springernature.com/) for English language editing and ENCODE for permission to use images.

This study was supported by the following grants: JSPS KAKENHI [grant numbers: 15H04889 (K.N.) and 16K10239 (M.S.)], the Karoji Fund for Medical Research (A) (K.N.), the Uehara Memorial Foundation (K.N.), and Hirosaki University Institutional Research Grant (K.N. and S.U.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the authors' organization and JSPS.

Conflicts of interest

There are no conflicts of interest.

References

Abidin RR (1995). *The Parenting Stress Index*. 3rd ed. Florida: Psychological Assessment Resources, Inc.

Adachi M, Takahashi M, Takayanagi N, Yoshida S, Yasuda S, Tanaka M, *et al.* (2018). Correction: adaptation of the autism spectrum screening questionnaire (ASSQ) to preschool children. *PLoS One* **13**:e0203254.

American Psychiatric Association, editor (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Washington, DC: American Psychiatric Association.

Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**:63–77.

Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG, *et al.* (2011). Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. *Hum Mol Genet* **20**:4360–4370.

Chen C-H (2017). High resolution analysis of rare copy number variants in patients with autism spectrum disorder from Taiwan. *Sci Rep* **7**:11919.

Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, *et al.* (2008). Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry* **63**:1111–1117.

Chakrabarti S, Fombonne E (2001). Pervasive developmental disorders in preschool children. *JAMA* **285**:3093–3099.

Cho SC, Yim SH, Yoo HK, Kim MY, Jung GY, Shin GW, *et al.* (2009). Copy number variations associated with idiopathic autism identified by whole-genome microarray-based comparative genomic hybridization. *Psychiatr Genet* **19**:177–185.

Cook, EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, *et al.* (1997). Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am J Hum Genet* **60**:928–934.

DuPaul GJ, Power TJ, McGoey KE, Ikeda MJ, Anastopoulos AD (1998). Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *J Psychoeduc Assess* **16**:55–68.

Ehlers S, Gillberg C (1993). The epidemiology of asperger syndrome: a total population study. *J Child Psychol Psychiatry* **34**:1327–1350.

Esposito G, Hiroi N, Scattoni ML (2017). Cry, baby, cry: expression of distress as a biomarker and modulator in autism spectrum disorder. *Int J Neuropsychopharmacol* **20**:498–503.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, *et al.* (2006). Copy number variation: new insights in genome diversity. *Genome Res* **16**:949–961.

Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, *et al.* (2014). Most genetic risk for autism resides with common variation. *Nat Genet* **46**:881–885.

Goodman R (1997). The strengths and difficulties questionnaire: a research note. *J Child Psychol Psychiatry* **38**:581–586.

Gotham K, Pickles A, Lord C (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord* **39**:693–705.

Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, *et al.* (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* **68**:1095–1102.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, *et al.* (2004). Detection of large-scale variation in the human genome. *Nat Genet* **36**:949–951.

Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, *et al.* (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**:148–161.

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST; Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* **13**:680–685.

Kikusui T, Hiroi N (2017). A self-generated environmental factor as a potential contributor to atypical early social communication in Autism. *Neuropsychopharmacology* **42**:378.

Klopocki E, Mundlos S (2011). Copy-number variations, noncoding sequences, and human phenotypes. *Annu Rev Genomics Hum Genet* **12**:53–72.

Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, *et al.* (2008). Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**:477–488.

Matsuishi T, Nagano M, Araki Y, Tanaka Y, Iwasaki M, Yamashita Y, *et al.* (2008). Scale properties of the Japanese version of the Strengths and Difficulties Questionnaire (SDQ): A study of infant and school children in community samples. *Brain and Development* **30**:410–415.

Nakai A, Miyachi T, Okada R, Tani I, Nakajima S, Onishi M, *et al.* (2011). Evaluation of the Japanese version of the developmental coordination disorder questionnaire as a screening tool for clumsiness of Japanese children. *Res Dev Disabil* **32**:1615–1622.

Narama M, Kanematsu Y, Araki A, Maru M, Nakamura N, Takeda J, *et al.* (1999). Validity and reliability of Japanse version of the parenting stress index [in Japanese]. *J Child Health* **58**:610–616.

Pereanu W, Larsen EC, Das I, Estévez MA, Sarkar AA, Spring-Pearson S, *et al.* (2018). AutDB: a platform to decode the genetic architecture of autism. *Nucleic Acids Res* **46**:D1049–D1054.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, *et al.* (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**:368–372.

Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, *et al.* (2012). A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 (Bethesda)* **2**:1665–1685.

Reddy KS (2005). Cytogenetic abnormalities and fragile-X syndrome in autism spectrum disorder. *BMC Med Genet* **6**:3.

Saito M, Hirota T, Sakamoto Y, Adachi M, Takahashi M, Osato-Kaneda A, *et al.* (2020). Prevalence and cumulative incidence of autism spectrum disorders and the patterns of co-occurring neurodevelopmental disorders in a total population sample of 5-year-old children. *Mol Autism* **11**:35.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, *et al.* (2004). Large-scale copy number polymorphism in the human genome. *Science* **305**:525–528.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, *et al.* (2007). Strong association of de novo copy number mutations with autism. *Science* **316**:445–449.

Shaw KA, Maenner MJ, Baio J, Washington A, Christensen DL, Wiggins LD, *et al.* (2020). Early identification of autism spectrum disorder among children aged 4 years – early autism and developmental disabilities monitoring network, six sites, United States, 2016. *MMWR Surveill Summ* **69**:1–11.

Takayanagi N, Yoshida S, Yasuda S, Adachi M, Kaneda-Osato A, Tanaka M, *et al.* (2016). Psychometric properties of the Japanese ADHD-RS in preschool children. *Res Dev Disabil* **55**:268–278.

The Autism Genome Project Consortium (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* **39**:319–328.

Walker S, Scherer SW (2013). Identification of candidate intergenic risk loci in autism spectrum disorder. *BMC Genomics* **14**:499.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, *et al.* (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**:1665–1674.

Wechsler D; Japanese WISC-IV Publication Committee (2010). *Japanese version of the Wechsler intelligence scale for children-fourth edition (WISC-IV)*. Tokyo: Nihon Bunka Kagakusha.

Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, Hardekopf D, *et al.* (2012). Microdeletion and microduplication syndromes. *J Histochem Cytochem* **60**:346–358.

Wilson BN, Crawford SG, Green D, Roberts G, Aylott A, Kaplan BJ (2009). Psychometric properties of the revised developmental coordination disorder questionnaire. *Phys Occup Ther Pediatr* **29**:182–202.

Wilson BN, Kaplan BJ, Crawford SG, Campbell A, Dewey D (2000). Reliability and validity of a parent questionnaire on childhood motor skills. *Am J Occup Ther* **54**:484–493.

Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M (2002). The diagnostic interview for social and communication disorders: background, inter-rater reliability and clinical use. *J Child Psychol Psychiatry* **43**:307–325.

Xu J, Zwaigenbaum L, Szatmari P, Scherer S (2004). Molecular cytogenetics of autism. *Curr Genomics* **5**:347–364.

Yeargin-Allsopp M, Rice C, Karapurkar T (2003). Autism prevalence in the United States. *Developmental and Behavioral Pediatrics* **24**:6.