ARTICLE Molecular diagnosis of 405 individuals with autism spectrum disorder

Noriko Miyake 1^{2} , Yoshinori Tsurusaki³, Ryoko Fukai⁴, Itaru Kushima^{5,6}, Nobuhiko Okamoto⁷, Kei Ohashi⁸, Kazuhiko Nakamura^{9,10}, Ryota Hashimoto¹¹, Yoko Hiraki¹², Shuraku Son¹³, Mitsuhiro Kato ¹⁴, Yasunari Sakai ¹⁵, Hitoshi Osaka ¹⁶, Kimiko Deguchi¹⁷, Toyojiro Matsuishi^{18,19}, Saoko Takeshita²⁰, Aviva Fattal-Valevski²¹, Nina Ekhilevitch²², Jun Tohyama²³, Patrick Yap ^{24,25}, Wee Teik Keng²⁶, Hiroshi Kobayashi²⁷, Keiyo Takubo²⁷, Takashi Okada^{5,28}, Shinji Saitoh ⁸, Yuka Yasuda¹¹, Toshiya Murai¹³, Kazuyuki Nakamura²⁹, Shouichi Ohga¹⁵, Ayumi Matsumoto¹⁶, Ken Inoue^{17,30}, Tomoko Saikusa^{18,19}, Tova Hershkovitz²², Yu Kobayashi²³, Mako Morikawa⁵, Aiko Ito³¹, Toshiro Hara³², Yota Uno⁵, Chizuru Seiwa³¹, Kanako Ishizuka⁵, Emi Shirahata³¹, Atsushi Fujita², Eriko Koshimizu ², Satoko Miyatake ^{2,33}, Atsushi Takata ³⁴, Takeshi Mizuguchi², Norio Ozaki ^{5,35} and Naomichi Matsumoto ²

© The Author(s), under exclusive licence to European Society of Human Genetics 2023

Autism spectrum disorder (ASD) is caused by combined genetic and environmental factors. Genetic heritability in ASD is estimated as 60–90%, and genetic investigations have revealed many monogenic factors. We analyzed 405 patients with ASD using familybased exome sequencing to detect disease-causing single-nucleotide variants (SNVs), small insertions and deletions (indels), and copy number variations (CNVs) for molecular diagnoses. All candidate variants were validated by Sanger sequencing or quantitative polymerase chain reaction and were evaluated using the American College of Medical Genetics and Genomics/Association for Molecular Pathology guidelines for molecular diagnosis. We identified 55 disease-causing SNVs/indels in 53 affected individuals and 13 disease-causing CNVs in 13 affected individuals, achieving a molecular diagnosis in 66 of 405 affected individuals (16.3%). Among the 55 disease-causing SNVs/indels, 51 occurred de novo, 2 were compound heterozygous (in one patient), and 2 were X-linked hemizygous variants inherited from unaffected mothers. The molecular diagnosis rate in females was significantly higher than that in males. We analyzed affected sibling cases of 24 quads and 2 quintets, but only one pair of siblings shared an identical pathogenic variant. Notably, there was a higher molecular diagnostic rate in simplex cases than in multiplex families. Our simulation indicated that the diagnostic yield is increasing by 0.63% (range 0–2.5%) per year. Based on our simple simulation, diagnostic yield is improving over time. Thus, periodical reevaluation of ES data should be strongly encouraged in undiagnosed ASD patients.

European Journal of Human Genetics; https://doi.org/10.1038/s41431-023-01335-7

INTRODUCTION

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction across multiple contexts, and restricted, repetitive patterns of behavior, interests, or activities, with onset during the first 3 years of life. The prevalence

Received: 11 December 2022 Revised: 8 February 2023 Accepted: 7 March 2023 Published online: 27 March 2023

¹Department of Human Genetics, National Center for Global Health and Medicine, Tokyo, Japan. ²Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. ³Faculty of Nutritional Science, Sagami Women's University, Sagamihara, Japan. ⁴Department of Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan. ⁵Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Japan. ⁶Medical Genomics Center, Nagoya University Hospital, Nagoya, Japan. ⁷Department of Medical Genetics, Osaka Women's and Children's Hospital, Osaka, Japan. ⁸Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan. 9Department of Neuropsychiatry, Hirosaki University Graduate School of Medicine, Hirosaki, Japan. ¹⁰Department of Psychiatry, Hamamatsu University School of Medicine, Hamamatsu, Japan. ¹¹Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Japan. ¹²Hiroshima Municipal Center for Child Health and Development, Hiroshima, Japan. ¹³Department of Psychiatry, Kyoto University Graduate School of Medicine, Kyoto, Japan. ¹⁴Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan. ¹⁵Department of Pediatrics, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.¹⁶Department of Pediatrics, Jichi Medical University, Shimotsuke, Japan.¹⁷Deguchi Pediatric Clinic, Omura, Japan. ¹⁸Departments of Pediatrics and Child Health, Kurume University School of Medicine, Kurume, Japan. ¹⁹Department of Pediatrics, St. Mary's Hospital, Kurume, Japan. ²⁰Department of Pediatrics, Yokohama City University Medical Center, Yokohama, Japan. ²¹Pediatric Neurology Institute, Dana-Dwek Children's Hospital, Tel Aviv Medical Center & Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. ²²The Genetics Institute, Rambam Health Care Campus, Haifa, Israel. ²³Department of Child Neurology, National Hospital Organization Nishiniiqata Chuo Hospital, Niigata, Japan. 24 Genetic Health Service New Zealand, Auckland, New Zealand. 25 Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand.²⁶Genetic Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia.²⁷Department of Stem Cell Biology, Research Institute, National Center for Global Health and Medicine, Tokyo, Japan.²⁸Department of Developmental Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Japan.²⁹Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan.³⁰Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan.³¹Department of Pediatrics, Yamagata Prefectural Rehabilitation Center for Children with Disabilities, Yamagata, Japan. 32 Fukuoka Children's Hospital, Fukuoka, Japan. 33 Department of Clinical Genetics, Yokohama City University Hospital, Yokohama, Japan. ³⁴Laboratory for Molecular Pathology of Psychiatric Disorders, RIKEN Center for Brain Science, Wako, Japan. ³⁵Institute for Glyco-core Research (iGCORE), Nagoya University, Nagoya, Japan. [⊠]email: nomiyake@ri.ncgm.go.jp; naomat@yokohama-cu.ac.jp

of ASD is estimated to be 1 in 54–100 children, and the male-to-female ratio has been reported as 4.2–4.3:1 [1, 2].

Both genetic and environmental factors have long been investigated as contributors to ASD. A genetic component of ASD was first suspected in twin studies; the concordance of ASD in identical twins (60-96% concordance) is much higher than that in dizygotic twins (0-36%) [3, 4]. Furthermore, the heritability of ASD is estimated to be 60-90% (reviewed by Ruzzo et al. [5]). In the past few decades, many genetic studies have revealed that a proportion of ASD cases can be explained by rare inherited or de novo single-nucleotide variants (SNVs), including small insertions and deletions (indels) and copy number variants (CNVs) [5-9], as well as by commonly inherited SNVs/indels and CNVs [4, 10]. It is important to distinguish between individual and population risks for ASD; common variants can explain a genetic contribution at the population level but not at the individual level, whereas rare and/or de novo variants can explain ASD at the individual level but not at the population level [11]. Indeed, both rare and common variants are thought to contribute to the pathomechanism of ASD, but it remains unclear how this actually occurs.

The diagnostic yield of ASD using chromosomal microarray analyses has been reported as 3.0-9.3% [12-14], and that using exome sequencing (ES) as 6.1-8.4% [12, 14]. There are relatively few reports of the molecular diagnosis of ASD using ES to detect both SNVs/indels and CNVs. Feliciano et al. [15] reported that the diagnostic yield of ES for detecting SNVs/indels and CNVs in ASD was 10.4% in 457 families. Although the diagnostic rate of ES depends on the cohort and target disease that is being assessed, its rate in ASD is lower than that in other genetic conditions, which have usually been reported as 25%–44% [16, 17]. However, many aberrant disease-causing genes have been identified over the years, and 10-15% of unresolved cases may obtain a molecular diagnosis by reanalysis (reviewed by Lee and Nelson [18]). Here, we investigated monogenic causes that may be clinically relevant for the molecular diagnosis of ASD, and simulated how the reanalysis of ASD cohorts might improve diagnostic rates (i.e., the identification of genetic causes in ASD patients).

MATERIALS AND METHODS

Study cohort

A total of 405 affected individuals (283 males and 122 females) from 377 families clinically diagnosed with ASD based on the Diagnostic and Statistical Manual of Mental Disorders-V (Table 1) and their unaffected parents were included in this multi-center cohort. Our ASD cohort consisted of 351 trios, 24 quads (two affected siblings in each), and two quintets (three affected siblings in each). Samples were collected after obtaining written informed consent. Part of this cohort (N = 261) has been previously analyzed [19]. However, the analytical approach and aim of the present study are very different to those of the previous investigation; this study evaluated the pathogenicity of the respective variants to reach a molecular diagnosis at the individual level considering the American College of Medical Genetics (ACMG)/Association for Molecular Pathology (AMP) guidelines [20], whereas the previous study was more explorative and sought to understand de novo variants in ASD in a population-based manner. The present study was approved by the Institutional Review Boards of Yokohama City University School of Medicine and the other collaborating hospitals.

Genetic analysis

DNA was extracted from human peripheral leukocytes or saliva. ES was performed as previously reported [21]. In brief, genomic DNA was sheared using the Covaris S2 system (Covaris) and genome partitioning was performed using SureSelect Human All Exon kits (Agilent Technology) according to the manufacturers' instructions. Prepared samples were run on a HiSeq 2000/2500 instrument (Illumina) with 101-bp paired-end reads. The reads were mapped to the human reference genome hg19 using Novoalign (Novocraft). The SNVs/indels were called using the Genome Analysis Toolkit 3 (UnifiedGenotyper) and annotated using ANNOVAR. We selected candidate SNVs/indels in all possible inheritance

 Table 1.
 Cohort overview and molecular diagnostic rate in this cohort.

| Classification | Numbers of affected individuals (% of total) |
|-----------------------|--|
| Total | 405 (100%) |
| Sex | |
| Male | 283 (70%) |
| Female | 122 (30%) |
| Family history | |
| Trios | 351 (86.7%) |
| Quads | 48 (24 families) (11.9%) |
| Quintets | 6 (2 families) (1.4%) |
| Molecularly diagnosed | |
| Total | 65 (16.0%) |
| SNVs | 53 (13.1%) |
| CNVs | 13 (3.2%) |
| Males | 37 (13.1%) |
| Females | 29 (23.8%) |
| | |

Molecularly diagnosed rates were calculated based on the number of affected individuals.

modes: autosomal dominant (de novo), autosomal recessive, X-linked dominant, and X-linked recessive (for male individuals only). Each filtering setting is described in the Supplementary text. The candidate variants were validated by Sanger sequencing. The pathogenicity of candidate variants was classified in accordance with the ACMG/AMP guidelines [20]. We used SIFT, Polyphen-2, MutationTaster, and CADD for the in silico prediction of missense variants. All variant descriptions were confirmed by Mutalyzer (2.0.35).

We also examined candidate CNVs in ES data using the eXome Hidden Markov Model (XHMM) [22]. We picked up the candidate CNV calls with a OSOME score of 60 or more. CNVs were excluded from the candidates when their regions overlapped with polymorphic structural variations (Database of Genomic Variants), overlapped with regions frequently observed in our in-house control data, or did not contain protein-coding genes. The rest of the candidate regions and any CNVs of more than 200 kb in length were manually evaluated in terms of whether they overlapped with dose-sensitive regions or genes with a score of 3 (Sufficient Evidence) or 2 (Emerging Evidence) for haploinsufficiency or triplosensitivity in ClinGen, were in regions known to be associated with CNV syndromes [DatabasE of genomiC variation and Phenotype in Humans using Ensembl Resources (DECIPHER) syndromes, n = 66], or were dosesensitive genes as determined using DECIPHER. The pathogenicity of each CNV was classified in accordance with ACMG/AMP guidelines [23]. The candidate CNVs were validated by quantitative polymerase chain reaction (PCR) in two different regions within the CNV-called regions, with normalization using two independent control regions (FBN1 and STXBP1). Primer sequences and PCR conditions are available on request.

Simulation of diagnostic yield by year

By referring to the Human Genome Mutation Database Professional and the Online Mendelian Inheritance in Man (OMIM) database, we noted the years in which many aberrant genes were first identified in neurodevelopmental disorders (including ASD). We then calculated the number of cases in which a molecular diagnosis would have been obtained (together with the diagnostic rate) if exome analysis were to have been performed each year.

Gene ontology (GO) analysis

We performed GO analysis using DAVID Bioinformatics Resources version 6.8 to determine the enrichment of genes in this ASD cohort. We input 52 genes (Supplementary Table 1) harboring disease-causing variants or associated with disease-causing CNVs in our cohort, and visualized the results using the GOseq package [24] of the R program. We also performed GO analysis with the enrichGO function from the ClusterProfiler package [25] of R program with Benjamini–Hochberg adjustment of *p* values. GO terms with the 25 lowest *p* values were visualized.

N. Miyake et al.

SNV/indel detection

Among the 405 affected individuals, 55 SNVs that can explain the clinical phenotype were detected in 53 individuals (13.1%), including 45 SNVs and 10 indels (Fig. 1A, Table 1, and Supplementary Table 2). All variants were confirmed by Sanger sequencing. Among them, 21 variants had not previously been reported (Supplementary Table 2). On the basis of the ACMG/AMP guidelines [20], 28 and 27 variants were classified as "Pathogenic" and "Likely pathogenic," respectively (Supplementary Table 2). The inheritance modes of the variants were classified as autosomal dominant (de novo, n = 39), autosomal recessive (n = 2, one individual with compound heterozygous variants, each inherited either from the father or the mother), X-linked dominant (de novo, n = 9), and X-linked recessive (de novo, n = 3; maternal inheritance, n = 2) (Fig. 1B, Supplementary Table 2).

One male patient (Individual ID: 8553) had two disease-causing variants: a de novo *BRAF* missense variant [NM_004333.6, c.722 C > T, p.(Thr241Met)] and a de novo *CAMK2B* missense variant [NM_001220.5, c.1991C > T, p.(Pro664Leu)]. He was diagnosed with *BRAF*-associated cardio-facio-cutaneous syndrome with acute encephalopathy [26]. This *BRAF* variant is classified as "Pathogenic" based on ACMG/AMP guidelines, and it was previously reported as pathogenic in multiple patients [27]. The *CAMK2B* variant is classified as "Likely pathogenic" based on ACMG/AMP guidelines, and its pathogenicity may contribute to some of the patient's clinical features (developmental delay, speech delay, and no reacquisition of language skills after acute encephalopathy) because pathogenic *CAMK2B* variants are found in mental retardation, autosomal dominant 54 (MIM#617799).

Another male patient (Individual ID: 22716) had a de novo splicing variant at an intronic region of *FOXP1* (NM_032682.6, c.1428 + 5 G > A). This variant is not registered in gnomAD, Exome Variant Server, or the Human Genetic Variation Database. It is not located at a canonical splicing site, but three different splice site prediction software tools gave a decreased score at the canonical splice donor site of intron 16 for the altered allele compared with the wild-type one (ESE finder, 10.49 to 7.05; NetGene2, 0.86 to



Fig. 1 Genetic architecture in our autism spectrum disorder (ASD) cohort. A Identified monogenic single-nucleotide variants (SNVs) including small insertions and deletions and copy number variations (CNVs) in our ASD cohort. B The number of individuals with pathogenic SNVs and CNVs leading to a molecular diagnosis under the respective genetic inheritance modes. C Size distribution of 25 candidate CNVs. D The diagnostic yields in male and female individuals with either SNVs or CNVs.

0.67; BDGP splice site predictor, 1.0 to 0.97). In addition, SpliceAl Lookup [28] predicted the splice acceptor loss 84 bp upstream from the variant at the pre-mRNA level (delta score 0.79) and the splice donor loss 5 bp upstream from the variant at the pre-mRNA level (delta score 0.52) in the variant allele; these predictions indicate the impaired splicing of exon 16. In addition, this base is highly conserved with a Genomic Evolutionary Rate Profiling score of 6.17. This patient's clinical features could be explained by the *FOXP1* variant, which causes mental retardation with language impairment with or without autistic features (MIM#613670). On the basis of ACMG/AMP guidelines, this variant can be classified as "Likely pathogenic."

We also found five variants of unknown significance (VUSs) in five genes that may potentially explain the patients' phenotypes (Supplementary Table 3). They were all missense variants in X-linked genes inherited from unaffected mothers. Among the five variants, three (*NLGN4X*, *TMLHE*, and *IL1RAPL1*) are thought to be X-linked recessive. These three variants match PM2 (extremely low frequency if recessive) and PP3 (multiple in silico prediction) and are classified as VUSs in accordance with ACMG/ AMP guidelines [20].

The other two genes—SRPK3 and PCDH11X—have not been conclusively shown to be related to diseases in OMIM. A missense SRPK3 variant [NM_014370.4, c.475 C > G, p.(His159Asp)] was previously reported in intellectual disability [29]. As for PCDH11X, at least three truncating variants, four gross deletions, and eight gross insertions have been reported previously in individuals with neurodevelopmental disorders [30], including autism and developmental dyslexia, according to the Human Gene Mutation Database (as of October 17, 2022). Meanwhile, at least 21 lossof-function PCDH11X variants are registered in gnomAD SVs v2.1, but only six loss-of-function hemizygous variants in 10 individuals and no homozygous loss-of-function variants were observed in control populations. The most common hemizygous loss-offunction variant [chrX: 91873457 C > T, p.(Arg1188*)] was observed in 5 of 183,192 alleles (minor allele frequency = 0.000027), implying that PCDH11X null variants produce variable neurodevelopmental symptoms including ASD. Our case (Individual ID 19010) had a novel missense variant, which remains a VUS in line with ACMG/AMP guidelines [20].

CNV detection

Among 405 affected individuals, 27 possible candidate CNVs were detected in 25 patients. Two of these 27 CNVs were unable to be confirmed by quantitative PCR (likely false positive). Another 4 of the remaining 25 CNVs in two affected individuals were unable to be validated because of the limited residual DNA samples (Supplementary Table 4). Finally, the remaining 21 CNVs were validated in 21 patients (5.2% of our cohort), including 10 deletions and 11 duplications. The sizes of the CNVs ranged from 289 to 21,972 kb, with XHMM QSOME scores of 92–99 (Fig. 1C, Supplementary Table 4).

Among the 21 CNVs, 13 led to a molecular diagnosis in 13 patients (3.2% of our cohort; five male and eight female patients) (Fig. 1A, Tables 1 and 2). Eight pathogenic CNVs occurred de novo, three were paternally inherited, and two were maternally inherited. Twelve disease-causing CNVs were classified as having autosomal dominant inheritance (including seven occurring de novo) and one as having X-linked dominant inheritance (Fig. 1B, Table 2). Seven CNVs were associated with microdeletion/duplication syndromes and the other six can be explained by the haploinsufficiency of single genes: BCL11A deletion related to Dias-Logan syndrome (MIM# 617101), NBEA deletion related to neurodevelopmental disorder with or without early-onset generalized epilepsy (MIM# 619157), SHANK3 deletion related to Phelan-McDermid syndrome (MIM#606232), NRXN1 deletion related to complex neurodevelopmental disorder [31, 32] (no MIM# is given), KDM6A deletion related to Kabuki syndrome 2 (MIM#300867), and EBF3 deletion related to hypotonia,

ataxia, and delayed development syndrome (MIM#617330) (Table 2, Supplementary Table 4).

The most common pathogenic CNVs in our ASD cohort were copy number gains of chromosome 15q11-q13 (n = 4), known as maternal 15q duplication syndrome associated with either maternal isodicentric 15q11.2-q13.1 supernumerary chromosome (tetrasomy for 15q11.2-q13.1) or maternal interstitial 15q11.2-q13.1 duplication (trisomy for 15q11.2-q13.1). Deletion of this region causes Prader–Willi syndrome or Angelman syndrome, depending on whether the deletion occurs in a paternally or maternally inherited chromosome. Regarding the copy numbers, two patients had trisomy and the other two possessed tetrasomy of this region (Supplementary Fig. 1). Three 15q11-q13 copy number gains occurred *de novo* (Individual IDs 15441, 17649, and 22728) and one was maternally inherited (Individual ID 15664 with trisomy).

Affected sibling analysis

In 24 quad and 2 quintet families, we examined two (in quad families) or three (in quintet families) affected siblings by ES to determine whether they shared an identical genetic cause. Surprisingly, only one quad family shared an indel variant [*IRF2BPL*, NM_024496.4:c.1484_1486delinsCGT, p.(Leu495_Pro496delinsPro-Ser)] in affected siblings (Individual IDs 16771 and 16774) (Supplementary Table 5). In this family, germline mosaicism in either of their parents was suspected based on the observation that there were no mutant alleles in >50 reads of ES in both parental samples.

Interestingly, in 2 of the 26 familial cases, one of the affected siblings had disease-causing variants but the other(s) did not (Supplementary table 5). The molecular diagnostic rate in familial cases was calculated to be 7.4% (4/54), which was lower than that in simplex cases (17.7%, 62/351) in this cohort, implying a possible difference in genetic architecture between simplex and multiplex families.

Autosomal versus sex chromosomal genes

We identified disease-causing SNVs in 32 (11.3%) and CNVs in 5 (1.8%) of 283 male patients, and disease-causing SNVs in 21 (17.2%) and CNVs in 8 (6.6%) of 122 female patients. Interestingly, the diagnostic rate was nearly twice as high in female patients than it was in male patients in our cohort (13.1% in male and 23.8% in female patients; p = 0.0075, chi-squared test) (Fig. 1D, Table 1). Of all of the disease-causing SNVs/indels and CNVs in our cohort, 31 and 8 disease-causing variants in male patients were autosomal and X-linked, respectively (ratio of X-linked/all variants = 8/39, 20.5%), while 22 and 7 disease-causing variants in female patients were autosomal and X-linked, respectively (ratio of X-linked/all variants = 7/29, 24.1%). The molecular diagnostic rate of X-linked variants in female individuals was almost twice that in male individuals [5.7% (7/122) in females and 2.8% (8/283) in males].

Diagnostic yield improvement over the years

Recent technological developments in comprehensive genomic sequencing and the identification of many aberrant genes, susceptibility genes, and chromosomal abnormalities associated with ASD and its associated disorders should have led to improvements in the diagnostic yield over the years. We therefore simulated the diagnostic yield of the exome data using known disease-related aberrant genes for each year since 1992 (Supplementary Fig. 2). Among 65 patients, only 10 could have been diagnosed before 2000. However, 19 patients could have been newly diagnosed from 2000 to 2010, and 38 patients from 2011 onwards. It has previously been reported that the number of known disease-causing variants has rapidly increased [18, 33], especially since 2012. A rapid increase in diagnostic yield in 2012 was also expected in our cohort (Supplementary Fig. 2). Thus, the periodic reanalysis of ASD patients for whom disease causation remains unresolved is highly recommended. The average percentage of newly resolved cases among unresolved cases was 0.63% (range 0%–2.5%) per year in our ASD cohort.

GO analysis of genes with disease-causing SNVs and CNVs in this cohort

GO analysis was performed using the 52 genes in which we identified SNVs or CNVs in this cohort using DAVID 6.8 with the default settings (Supplementary Table 1). The cluster that was most enriched by functional annotation clustering was that containing genes related to transcription and DNA binding (Enrichment score: 2.23, Supplementary Table 6). The second and fourth clusters were also related to transcription; namely, helicase and chromatin regulation (Enrichment score: 1.87) and repression of transcription (Enrichment score: 1.76), respectively. The enriched terms were "Mental retardation," "Disease mutation," "Phosphorylation," "Methylation," "Epilepsy," "Visual learning," "Chromosomal rearrangement," "Nucleus," and "Autism spectrum disorder" with false discovery rates of <0.01 (Fig. 2A, Supplementary Table 7). Another GO enrichment analysis with enrichGO also showed that the ontology terms were mainly associated with nervous system development, synapses, behavior, and memory (Fig. 2B, Supplementary Table 8).

DISCUSSION

We analyzed 405 individuals with ASD and identified pathogenic variants in 66 affected individuals (SNVs in 53 individuals, 13.1%, and CNVs in 13 individuals, 3.2%). We achieved a molecular diagnostic yield of 16.3% (66/405) in ASD patients by ES (Fig. 1A). The diagnostic yield in our cohort was much higher than those of previous reports (6.0–10.4%) [12, 14, 15], while the pathogenic CNV detection rate was similar to those of previous reports (3.0–9.3%) [12–14]. In families with affected individuals who received a molecular diagnosis, recurrent risk may be partially predicted. A molecular diagnosis may also provide a better understanding of the natural history of each patient and lead to appropriate medical care.

Some of our identified genes (*PBX1* and *ABCC8*) have not yet been established as ASD-related genes. In OMIM, "autistic behavior" is not listed as a phenotype of *PBX1* or *ABCC8* aberrations; however, pathogenic *PBX1* variants lead to speech and developmental delays, and pathogenic *ABCC8* variants can result in neurological phenotypes because of repeated episodes of hypoglycemia. ASD-related features can therefore be observed under these genetic conditions. Nonetheless, it remains possible that other substantial genetic factors may coexist for ASD.

We analyzed 26 multiplex families (24 quads and 2 quintets) and identified a shared pathogenic variant in only one quad (Supplementary Table 5). In two other quads, we identified disease-causing variants in only one of the two affected siblings (i.e., not in both siblings). We initially expected that some of the familial cases would be explained by recessive variants shared by affected siblings; however, all identified variants were de novo. It has been reported that recessive variants contribute to a small proportion of ASD cases (Lim et al. estimated autosomal recessive contribution as 3% [34]), or up to 30% of cases in consanguineous families [35]. However, our cohort did not include any consanguineous families.

Notably, the diagnostic yield of simplex families (17.7%, 62/351) was much higher than that of multiplex families (7.4%, 4/54) in the present study. Similarly, it has been reported that rare de novo protein truncating events are more frequently observed in simplex families than in multiplex families [5, 36, 37]. Together, these findings suggest different genetic architecture in simplex and multiplex families.

Monogenic causative variants were more commonly identified in female patients than in male patients in this study (23.8% vs. 13.1%, respectively). Similarly, a two-fold enrichment of de novo protein-truncating variants in highly constrained genes in ASD

| | ı leading | 7101) | #WI | ç | ome | vith or epilepsy | ation | ation | - | ation | disorder #) | (2) | ation | 17330) |
|-----------|--|---|--|---|--|---|--|--|--|--|--|--|--|--|
| | Molecular diagnosis (MIM#) (haploinsufficient gene deletior phenotype) | Dias-Logan syndrome (MIM# 61 (<i>BCL11A</i> deletion) | 22q11 duplication syndrome (M 608363) | 16p13.11 recurrent microdeletio syndrome (No MIM#) | 16p11.2 microduplication syndr (MM# 611913) | Neurodevelopmental disorder w without early-onset generalized (MIM# 619157) (NBEA deletion) | Chromosome 15q11-q13 duplics syndrome (MIM# 608636) | Chromosome 15q11-q13 duplics syndrome (MIM# 608636) | Phelan-McDermid syndrome (MIM#606232) (SHANK3 deletion) | Chromosome 15q11-q13 duplics syndrome (MIM# 608636) | (Complex neurodevelopmental due to <i>NRXN1</i> deletion, No MIM | Kabuki syndrome 2 (MIM#30086 (KDM6A deletion) | Chromosome 15q11-q13 duplics syndrome (MIM# 608636) | Hypotonia, ataxia, and delayed development syndrome (MIM#6 (<i>EBF3</i> deletion) |
| | qPCR | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed | Confirmed |
| | Class | Pathogenic (1.15) | Pathogenic (1) | Pathogenic (1) | Pathogenic (1) | Pathogenic (1) | Pathogenic (1.15) | Pathogenic (1) | Pathogenic (2.05) | Pathogenic (1.15) | Pathogenic (1.15) | Pathogenic (1.15) | Pathogenic (1.15) | Pathogenic (2.05) |
| | Inheritance | de novo | Maternal inherit | Paternal inherit | Paternal inherit | Paternal inherit | de novo | Maternal inherit | de novo | de novo | de novo | de novo | de novo | de novo |
| | Type | loss | gain | loss | gain | loss | gain | gain | loss | gain | loss | loss | gain | loss |
| | QSOME | 66 | 66 | 96 | 96 | 94 | 95 | 94 | 66 | 95 | 66 | 66 | 94 | 92 |
| | Size (kb) | 5131 | 1029 | 415 | 555 | 367 | 4775 | 4758 | 1205 | 4948 | 289 | 971 | 5020 | 5542 |
| for ASD. | Physical position (hg19) | 56,144,725–61,275,901 | 19,048,328–20,077,727 | 15,812,960–16,228,448 | 29,644,650–30,199,846 | 35,683,436–36,050,254 | 23,684,730–28,459,432 | 23,688,881–28,446,773 | 50,014,088–51,219,185 | 23,684,730–28,632,917 | 50966119–51255586 | 44,088,917–45,060,040 | 23,609,190–28,629,652 | 129,839,153–135,381,963 |
| ng CNVs f | Chr | 7 | 22 | 16 | 16 | 13 | 15 | 15 | 22 | 15 | 7 | × | 15 | 10 |
| se-causi | Sex | ш | Σ | Σ | ш | ш | Σ | Σ | Σ | ш | ш | ш | ш | ш |
| 2. Disea | ٩ | 6347 | 8385 | 9445 | 11159 | 11981 | 15441 | 15664 | 15680 | 17649 | 19022 | 22119 | 22728 | 26864 |
| Table | # | - | 7 | m | 4 | Ŋ | 9 | ~ | œ | 6 | 10 | = | 12 | 13 |



Fig. 2 Gene ontology (GO) enrichment analysis for our autism spectrum disorder (ASD) cohort. GO analysis was performed for 52 genes with pathogenic variants. **A** Visualization of GO enrichment analyzed with DAVID 6.8. Top 9 categories with a false discovery rate of <0.01 are shown. **B** Top 25 enriched GO terms analyzed with enrichGO are shown. Count: the number of hit genes within each category. Hits (%): proportion of the 52 input genes included in each GO term.

females versus males has been reported [9]. These observations are consistent with the "female protective effect" model in which fewer women are diagnosed despite having the same risk as men for developing ASD [9]. A gender bias also tends to be more prominent in groups with higher intelligence quotient scores (male bias is as high as 9:1 among cases with normal-to-highrange intelligence quotient scores [high-functioning] but as low as 1.6:1 among cases with intellectual disability), and it has been suggested that ASD may be masked by higher language skills, especially in high-functioning females (reviewed by Werling [38]). Fewer high-functioning females may therefore be diagnosed with ASD, resulting in a higher proportion of females with intellectual disabilities in the diagnosed population; consequently, more monogenic causes might be identified in female ASD patients.

The molecular diagnostic rate of X-linked variants in female individuals was approximately twice that of male individuals (5.7% vs. 2.8%, respectively) in the present study. In a neurodevelopmental disorder cohort, it was reported that de novo variants are enriched in some X-linked genes in female patients [39]. The most likely explanation for this phenomenon involves hemizygous male lethality [39]. Another possibility involves difficulties in defining the pathogenicity of variants in X-linked recessive traits especially for maternally inherited missense variants—when previously unreported as disease-causing. Thus, even if the variants are truly disease-causing, they might be classified as VUSs under the current ACMG/AMP guidelines. Further functional and/or genetic evidence may be needed for confirming pathogenicity.

In our GO analysis using DAVID 6.8, the most enriched cluster contained genes related to transcription and DNA binding. In previous studies [9, 40], gene expression regulation including chromatin regulation and transcription factors, neuronal communication including synaptic function, cytoskeleton, and others were enriched, also supporting the GO enrichment of "transcription" in our ASD cohort focusing on the monogenic causes of ASD.

Interestingly, another GO analysis using enrichGO showed that "Face development" was also enriched (p.adjust: 0.00082). This might reflect that some ASD patients are affected by syndromic ASD presenting with a characteristic face; in our cohort, *ANKRD11* aberration is known for KBG syndrome (MIM#148050), *CHD7* for

6

CHARGE syndrome (MIM#214800), *EP300* for Menke–Hennekam syndrome 2 (MIM#618333) or Rubinstein–Taybi syndrome 2 (MIM#613684), and *PTPN11* for Noonan syndrome 1 (MIM#163950) or LEOPARD syndrome 1 (MIM#151100). In addition, the enrichment of "GABA signaling pathway" (p.adjust: 0.0036) was also detected (Fig. 2B, Supplementary Table 8). The association between ASD and GABA has been previously reported and considered as a potential treatment target [41–43].

When analyzing the potential phenotypic expression of monogenic disorders including ASD, genomic imprinting must be considered. 15q11–q13 duplication is expected to be present in approximately 1 in 5000 individuals in the general population and its penetrance has been calculated as 54% [44]. Interestingly, penetrance differs depending on the parent from whom the abnormal copy originated and the type of chromosomal abnormality. It is 100% in maternal isodicentric 15q11.2–q13.1 supernumerary chromosome and almost 100% in maternal duplication, but is less than 50% in paternal abnormalities, with some cases even expressing a normal phenotype [45–47]. Thus, knowledge of the parental origin of such duplications (if inherited by the offspring) may be useful for predicting phenotype in genetic counseling.

Our analytical flow had some limitations. One is that we were likely to miss pathogenic SNVs/indels with incomplete penetrance in dominantly inherited traits because we selected candidate variants by picking up rare variants with a minor allele frequency of <0.1%; we filtered out all inherited variants for autosomal and X-linked dominant models. Furthermore, we focused on CNVs involving known disease genes or >200 kb CNVs that overlapped with known pathogenic regions. As a result, all of the detected CNVs were larger than 200 kb in size. As we have previously reported, XHMM is less powerful for detecting CNVs of less than 200 kb [48]. We might therefore have missed pathogenic CNVs smaller than 200 kb. In addition, we selected the pathogenic regions for which solid evidence was available based on the most recently updated public databases, including ClinGen and DECIPHER. Although these databases are very useful, not all genes have been curated in them yet. For example, we found an interesting case (Individual ID 8397) with a de novo 3.9 Mb duplication at 22q13, which partially overlaps with the critical region (chr22:51045516–51187844 based on hg19) for 22g13 deletion syndrome (Phelan-McDermid syndrome, MIM#606232) (Supplementary Table 4). Although there were 49 genes in the 3.9 Mb duplicated region of the patient, only 7 had been completely curated in ClinGen (Supplementary table 9). At present, there is no evidence that this region contains any triplosensitive genes based on ClinGen. Recently, however, the dosage sensitivity of all protein-coding genes has been reported [49]. Based on the recommended threshold (pHaplo score ≥ 0.86 and pTriplo score ≥0.94), candidate genes were identified in 8 of 12 CNVs unrelated to known diseases/syndromes (Supplementary Table 10). If more genetic evidence is accumulated, the diagnostic rate may thus be improved. In addition, variants in noncoding regions, mitochondrial dysfunction, and mosaic variants have been suggested to be involved in ASD [50-53]. Because the current ES analysis in the present cohort may have missed any of these, genome sequencing and/or deep sequencing may detect further disease-causing variants.

In conclusion, we performed a comprehensive analysis to detect SNVs/indels and CNVs using ES data, and achieved a molecular diagnosis in 66 of 405 affected individuals (16.3%). In addition, we demonstrated the effectiveness of reanalyzing ES data for unresolved cases with ASD. The higher diagnostic rates in simplex cases than in multiplex families in the present study support two genetic components, of monogenic and polygenic factors, in ASD genomic architecture. Because ASD is genetically and phenotypically heterogeneous, one medication is unlikely to successfully treat all patients. Disease-causing variants of monogenic diseases may have strong effects on phenotypes and thus signal potential treatment targets. We believe that a comprehensive analysis of the monogenic causes of ASD, to understand its pathomechanism, may be important for new drug discoveries.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are not publicly available due to privacy or ethical restrictions but are available from the corresponding author on reasonable request.

REFERENCES

- Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2016. MMWR Surveill Summ. 2020;69:1–12.
- Zeidan J, Fombonne E, Scorah J, Ibrahim A, Durkin MS, Saxena S, et al. Global prevalence of autism: a systematic review update. Autism Res. 2022;15:778–90.
- Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med. 1995;25:63–77.
- Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46:881–5.
- Ruzzo EK, Perez-Cano L, Jung JY, Wang LK, Kashef-Haghighi D, Hartl C, et al. Inherited and de novo genetic risk for autism impacts shared networks. Cell. 2019;178:850–66.e26.
- Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47:582–8.
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 2014;515:216–21.
- Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 2015;87:1215–33.
- Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Largescale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 2020;180:568–84.e23.
- Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
- 11. Dias CM, Walsh CA. Recent advances in understanding the genetic architecture of autism. Annu Rev Genomics Hum Genet. 2020;21:289–304.
- Martinez-Granero F, Blanco-Kelly F, Sanchez-Jimeno C, Avila-Fernandez A, Arteche A, Bustamante-Aragones A, et al. Comparison of the diagnostic yield of aCGH and genome-wide sequencing across different neurodevelopmental disorders. NPJ Genom Med. 2021;6:25.
- Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, et al. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics. 2010;125:e727–35.
- Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. JAMA. 2015;314:895–903.
- Feliciano P, Zhou X, Astrovskaya I, Turner TN, Wang T, Brueggeman L, et al. Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. NPJ Genom Med. 2019;4:19.
- Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014;312:1870–9.
- Narita K, Muramatsu H, Narumi S, Nakamura Y, Okuno Y, Suzuki K, et al. Wholeexome analysis of 177 pediatric patients with undiagnosed diseases. Sci Rep. 2022;12:14589.
- Lee H, Nelson SF. The frontiers of sequencing in undiagnosed neurodevelopmental diseases. Curr Opin Genet Dev. 2020;65:76–83.
- Takata A, Miyake N, Tsurusaki Y, Fukai R, Miyatake S, Koshimizu E, et al. Integrative analyses of de novo mutations provide deeper biological insights into autism spectrum disorder. Cell Rep. 2018;22:734–47.
- 20. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.
- Miyake N, Tsukaguchi H, Koshimizu E, Shono A, Matsunaga S, Shiina M, et al. Biallelic mutations in nuclear pore complex subunit NUP107 cause early-childhood-onset steroid-resistant nephrotic syndrome. Am J Hum Genet. 2015;97:555–66.
- 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.

- 23. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22:245–57.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNAseq: accounting for selection bias. Genome Biol. 2010;11:R14.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
- Okuzono S, Fukai R, Noda M, Miyake N, Lee S, Kaku N, et al. An acute encephalopathy with reduced diffusion in BRAF-associated cardio-facio-cutaneous syndrome. Brain Dev. 2019;41:378–81.
- Sarkozy A, Carta C, Moretti S, Zampino G, Digilio MC, Pantaleoni F, et al. Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. Hum Mutat. 2009;30:695–702.
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell 2019;176:535–48.e24.
- Niranjan TS, Skinner C, May M, Turner T, Rose R, Stevenson R, et al. Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. PLoS One. 2015;10:e0116454.
- Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB. Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia. Am J Med Genet B Neuropsychiatr Genet. 2013;162B:889–97.
- Zahir FR, Baross A, Delaney AD, Eydoux P, Fernandes ND, Pugh T, et al. A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1alpha. J Med Genet. 2008;45:239–43.
- Dabell MP, Rosenfeld JA, Bader P, Escobar LF, El-Khechen D, Vallee SE, et al. Investigation of NRXN1 deletions: clinical and molecular characterization. Am J Med Genet A. 2013;161A:717–31.
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. Am J Hum Genet. 2015;97:199–215.
- Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. 2013;77:235–42.
- Martin HC, Jones WD, McIntyre R, Sanchez-Andrade G, Sanderson M, Stephenson JD, et al. Quantifying the contribution of recessive coding variation to developmental disorders. Science. 2018;362:1161–4.
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316:445–9.
- Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, et al. Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. Am J Hum Genet. 2016;99:540–54.
- Werling DM. The role of sex-differential biology in risk for autism spectrum disorder. Biol Sex Differ. 2016;7:58.
- Turner TN, Wilfert AB, Bakken TE, Bernier RA, Pepper MR, Zhang Z, et al. Sexbased analysis of de novo variants in neurodevelopmental disorders. Am J Hum Genet. 2019;105:1274–85.
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014;515:209–15.
- Blatt GJ, Fatemi SH. Alterations in GABAergic biomarkers in the autism brain: research findings and clinical implications. Anat Rec (Hoboken). 2011;294:1646–52.
- Coghlan S, Horder J, Inkster B, Mendez MA, Murphy DG, Nutt DJ. GABA system dysfunction in autism and related disorders: from synapse to symptoms. Neurosci Biobehav Rev. 2012;36:2044–55.
- 43. Anagnostou E. Clinical trials in autism spectrum disorder: evidence, challenges and future directions. Curr Opin Neurol. 2018;31:119–25.
- Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, et al. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry. 2014;75:378–85.
- Cook EH Jr., Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, et al. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am J Hum Genet. 1997;60:928–34.
- Urraca N, Cleary J, Brewer V, Pivnick EK, McVicar K, Thibert RL, et al. The interstitial duplication 15q11.2-q13 syndrome includes autism, mild facial anomalies and a characteristic EEG signature. Autism Res. 2013;6:268–79.
- Browne CE, Dennis NR, Maher E, Long FL, Nicholson JC, Sillibourne J, et al. Inherited interstitial duplications of proximal 15q: genotype-phenotype correlations. Am J Hum Genet. 1997;61:1342–52.
- 48. 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;60:175–82.

- Collins RL, Glessner JT, Porcu E, Lepamets M, Brandon R, Lauricella C, et al. A crossdisorder dosage sensitivity map of the human genome. Cell 2022;185:3041–55 e25.
- Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, et al. Mutations in human accelerated regions disrupt cognition and social behavior. Cell. 2016;167:341–54.e12.
- 51. D'Gama AM. Somatic mosaicism and autism spectrum disorder. Genes (Basel). 2021;12:1699.
- Balachandar V, Rajagopalan K, Jayaramayya K, Jeevanandam M, Iyer M. Mitochondrial dysfunction: A hidden trigger of autism? Genes Dis. 2021;8:629–39.
- Sherman MA, Rodin RE, Genovese G, Dias C, Barton AR, Mukamel RE, et al. Large mosaic copy number variations confer autism risk. Nat Neurosci. 2021;24:197–203.

ACKNOWLEDGEMENTS

We thank the affected individuals and their families for participating in this study. We also thank Ms. Sayaka Sugimoto and Ms. Kaori Takabe from Yokohama City University Graduate School of Medicine for their technical assistance. This study makes use of data generated by the DECIPHER community. A full list of centers that contributed to the generation of the data is available from https://deciphergenomics.org/about/ stats and via e-mail from contact@deciphergenomics.org. Funding for the DECIPHER project was provided by Wellcome. Finally, we thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: NMi, NaM. Data curation: YT, RF, EK, SM, CO Investigation: NMi, YT, RF, IK, NO, KO, KN, RH, YH, SSo, MK, YS, HO, KD, TMa, ST, AF-V, NE, JT, PY, KWT, HK, KT, TO, SSa, YY, TMu, KN, SO, AM, KIn, TS, YK, MM, AI, TH, YU, CS, KIs, ES, AF, EK, SM, AT, TMi, NO, Visualization and Writing-original draft: NMi. Writing-review & editing: NMi, NaM.

FUNDING

This work was supported by AMED under grant numbers JP22ek0109486, JP22ek0109549, JP22ek0109493 (NMa), JP21wm0425007, and JP21dk0307103 (NO); JSPS KAKENHI under grant numbers JP19H03621 and 22H03047 (NMi), the Takeda Science Foundation (TM and NMa), and the NCGM Intramural Research Fund under grant number 21A1011 (NMi).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

This study was approved by the Institutional Review Board of Yokohama City University Faculty of Medicine. After obtaining written informed consent, peripheral blood leukocytes were collected from the patients and their parents.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-023-01335-7.

Correspondence and requests for materials should be addressed to Noriko Miyake or Naomichi Matsumoto.

Reprints and permission information is available at http://www.nature.com/ reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

8