Clinical spectrum of *SCN2A* mutations expanding to Ohtahara syndrome

Kazuyuki Nakamura, MD
Mitsuhiro Kato, MD, PhD
Hitoshi Osaka, MD, PhD
Sumimasa Yamashita, MD, PhD
Eiji Nakagawa, MD, PhD
Kazuhiro Haginoya, MD, PhD
Jun Tohyama, MD, PhD
Mitsuko Okuda, MD, PhD
Takahito Wada, MD, PhD
Shuichi Shimakawa, MD, PhD
Katsumi Imai, MD
Saoko Takeshita, MD
Hisako Ishiwata, MD
Dorit Lev, MD
Tally Lerman-Sagie, MD
David E. Cervantes-Barragán, MD
Camilo E. Villarroel, MD
Masaharu Ohfu, MD, PhD
Karin Writzl, MD, PhD
Barbara Gnidovec Stražišar, MD, PhD
Shinichi Hirabayashi, MD, PhD
David Chitayat, MD
Diane Myles Reid, MSc
Kiyomi Nishiyama, PhD
Hirofumi Kodera, PhD
Mitsuko Nakashima, MD, PhD
Yoshinori Tsurusaki, PhD
Noriko Miyake, MD, PhD
Kiyoshi Hayasaka, MD, PhD
Naomichi Matsumoto, MD, PhD
Hirotomo Saitsu, MD, PhD

Correspondence to Dr. Matsumoto: naomat@yokohama-cu.ac.jp or Dr. Saitsu: hsaitsu@yokohama-cu.ac.jp

Supplemental data at www.neurology.org

ABSTRACT

Objective: We aimed to investigate the possible association between *SCN2A* mutations and early-onset epileptic encephalopathies (EOEEs).

Methods: We recruited a total of 328 patients with EOEE, including 67 patients with Ohtahara syndrome (OS) and 150 with West syndrome. *SCN2A* mutations were examined using high resolution melt analysis or whole exome sequencing.

Results: We found 14 novel *SCN2A* missense mutations in 15 patients: 9 of 67 OS cases (13.4%), 1 of 150 West syndrome cases (0.67%), and 5 of 111 with unclassified EOEEs (4.5%). Twelve of the 14 mutations were confirmed as de novo, and all mutations were absent in 212 control exomes. A de novo mosaic mutation (c.3976G>C) with a mutant allele frequency of 18% was detected in one patient. One mutation (c.634A>G) was found in transcript variant 3, which is a neonatal isoform. All 9 mutations in patients with OS were located in linker regions between 2 transmembrane segments. In 7 of the 9 patients with OS, EEG findings transitioned from suppression-burst pattern to hypsarrhythmia. All 15 of the patients with novel *SCN2A* missense mutations had intractable seizures; 3 of them were seizure-free at the last medical examination. All patients showed severe developmental delay.

Conclusions: Our study confirmed that *SCN2A* mutations are an important genetic cause of OS. Given the wide clinical spectrum associated with *SCN2A* mutations, genetic testing for *SCN2A* should be considered for children with different epileptic conditions. ***Neurology*® 2013;81:992–998**

GLOSSARY

BFNIS = benign familial neonatal-infantile seizures; **DS** = Dravet syndrome; **EOEE** = early-onset epileptic encephalopathy; **HRM** = high resolution melt; **OS** = Ohtahara syndrome; **WES** = whole exome sequencing; **WS** = West syndrome.

Early-onset epileptic encephalopathies (EOEEs) include Ohtahara syndrome (OS), early myoclonic epileptic encephalopathy, West syndrome (WS), Dravet syndrome (DS), and other diseases.[1,2] OS is characterized by an early onset of spasms, mainly in the neonatal period, intractable seizures, and a suppression-burst pattern on EEG.[3] WS is characterized by spasms, an EEG finding termed hypsarrhythmia, and arrest of psychomotor development.[4] De novo mutations in *STXBP1*, *KCNQ2*, *CDKL5*, *ARX*, and *SPTAN1* are known causes of OS and WS.[5–11]

Voltage-gated sodium channels consist of 1 α subunit and 1 or 2 β subunits. The α subunit forms a pore structure, and is composed of 4 domains (domains I–IV), each containing 5 hydrophobic segments (S1, S2, S3, S5, S6) and 1 positively charged segment (S4).[12] The voltage-gated sodium channel repertoire in humans includes 9 α subunits ($Na_v1.1$–$Na_v1.9$). $Na_v1.1$, $Na_v1.2$, $Na_v1.3$, and $Na_v1.6$, encoded by *SCN1A*, *SCN2A*, *SCN3A*, and *SCN8A*,

From the Department of Human Genetics (K. Nakamura, K. Nishiyama, H.K., M.N., Y.T., N. Miyake, N. Matsumoto, H.S.), Yokohama City University Graduate School of Medicine, Yokohama; Department of Pediatrics (K. Nakamura, M.K., K. Hayasaka), Yamagata University Faculty of Medicine, Yamagata; Division of Neurology (H.O., S.Y., M. Okuda, T.W.), Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama; Department of Child Neurology (E.N.), National Center Hospital, National Center of Neurology and Psychiatry, Tokyo; Department of Pediatric Neurology (K. Haginoya), Takuto Rehabilitation Center for Children, Sendai; Department of Pediatrics (J.T.), Epilepsy Center, Nishi-Niigata Chuo National Hospital, Niigata; Department of Pediatrics (S.S.), Osaka Medical College Hospital, Osaka; National Epilepsy Center (K.I.), Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka; Department of Pediatrics (S.T.), Yokohama City University Medical Center, Yokohama; Department of Pediatrics (H.I.), Tokyo Metropolitan Bokuto Hospital, Tokyo, Japan; Metabolic Neurogenetic Clinic (D.L., T.L.-S.), Wolfson Medical Center, Holon, Israel; Department of Human Genetics (D.E.C.-B., C.E.V.), National Institute of Pediatrics, Mexico City, Mexico; Division of Child Neurology (M. Ohfu), Okinawa Nanbu Medical Center and Children's Medical Center, Okinawa, Japan; Institute of Medical Genetics (K.W.), University Medical Center Ljubljana; Department of Child, Adolescent and Developmental Neurology (B.G.S.), University Children's Hospital, Ljubljana, Slovenia; Department of Neurology (S.H.), Nagano Children's Hospital, Nagano, Japan; Department of Obstetrics and Gynecology (D.C.), The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto; and Division of Clinical and Metabolic Genetics (D.C., D.M.R.), The Hospital for Sick Children, University of Toronto, Canada.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

respectively, are highly expressed in the human brain,[13] and mutations can cause epilepsies and psychiatric disorders.[14–17]

Most *SCN2A* mutations cause benign phenotypes such as benign familial neonatal-infantile seizures (BFNIS),[18] and are usually inherited from an affected parent. In contrast, several de novo *SCN2A* mutations have been reported to cause more severe phenotypes,[14,19–22] suggesting a possible involvement in severe epileptic encephalopathies.

To elucidate the genetic basis of OS, we performed exome sequencing and found a de novo *SCN2A* mutation. Subsequently, we screened patients with EOEE for *SCN2A* mutations, and found that de novo *SCN2A* mutations contributed to the development of severe epileptic encephalopathies.

METHODS Patients. A total of 328 patients with EOEE (67 patients with OS, 150 with WS, and 111 with unclassified EOEE) were analyzed for *SCN2A* mutations. The 67 patients with OS consisted of 44 Japanese and 23 non-Japanese (from other countries). The diagnosis was made based on clinical features and characteristic patterns on EEG. In 257 patients, mutations in *STXBP1* and *KCNQ2* had been excluded by high resolution melt (HRM) analysis in advance. Mutation analysis was performed by HRM analysis or direct sequencing, and 41 of the 328 patients (32 with OS, 4 with WS, and 5 with unclassified EOEEs) were also analyzed by whole exome sequencing (WES). We obtained detailed clinical data from all patients with *SCN2A* mutations, including brain MRI/CT and EEG findings.

Mutation screening. Genomic DNA was obtained from peripheral blood leukocytes by standard methods. DNA for mutation screening was amplified using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Japan, Tokyo, Japan). Mutation screening of exons 2 to 27 (including exon 6A) covering the *SCN2A* coding regions (transcript variant 1, NM_021007.2) and exon 6N of transcript variant 3 (NM_001040143.1) was performed by HRM analysis or by direct sequencing of part of exon 27. The DNA of patient 10 from nails and hairs was isolated using ISOHAIR (Nippon Gene, Tokyo, Japan), and DNA in saliva was isolated using Oragene (DNA Genotek, Kanata, Canada). Real-time PCR and subsequent HRM analysis were performed using a Light Cycler 480 (Roche Diagnostics, Otsu, Japan). Samples showing an aberrant melting curve in the HRM analysis were sequenced. PCR primers and conditions are shown in table e-1 on the *Neurology*® Web site at www.neurology.org. All novel mutations were verified using original genomic DNA, and searched for in the variant database of our 212 in-house control exomes.

Whole exome sequencing. DNAs were captured using the SureSelect^{XT} Human All Exon v4 Kit (Agilent Technologies, Santa Clara, CA) and sequenced with 4 samples per lane on an Illumina HiSeq 2000 (Illumina, San Diego, CA) with 101–base pair paired-end reads. Image analysis and base calling were performed by sequence control software with real-time analysis and CASAVA software v1.8 (Illumina). Reads were aligned to GRCh37 with Novoalign (Novocraft Technologies, Selangor, Malaysia), and duplicate reads were marked using Picard (http://picard.sourceforge.net/index.shtml) and excluded from downstream analysis. Local realignments around indels and base quality score recalibrations were performed using the Genome Analysis Toolkit (GATK).[23] Single-nucleotide variants and small indels were identified using the GATK, and were annotated using ANNOVAR.[24] All mutations detected by WES were confirmed by Sanger sequencing.

Parentage testing. For the family showing de novo mutations, parentage was confirmed by microsatellite analysis using ABI Prism Linkage Mapping Set version 2.5, MD10 (Life Technologies, Carlsbad, CA). We chose 12 probes for screening (D6S422, D7S493, D8S285, D9S161, D10S208, D11S987, D12S345, D16S503, D17S921, D18S53, D19S220, and D20S196). Appropriate biological parentage was confirmed if 3 or more informative markers were compatible and other markers showed no discrepancy.

Standard protocol approvals, registrations, and patient consents. The experimental protocols were approved by the Institutional Review Boards for Ethical Issues of Yokohama City University School of Medicine and Yamagata University Faculty of Medicine. Informed consent was obtained from the families of all patients.

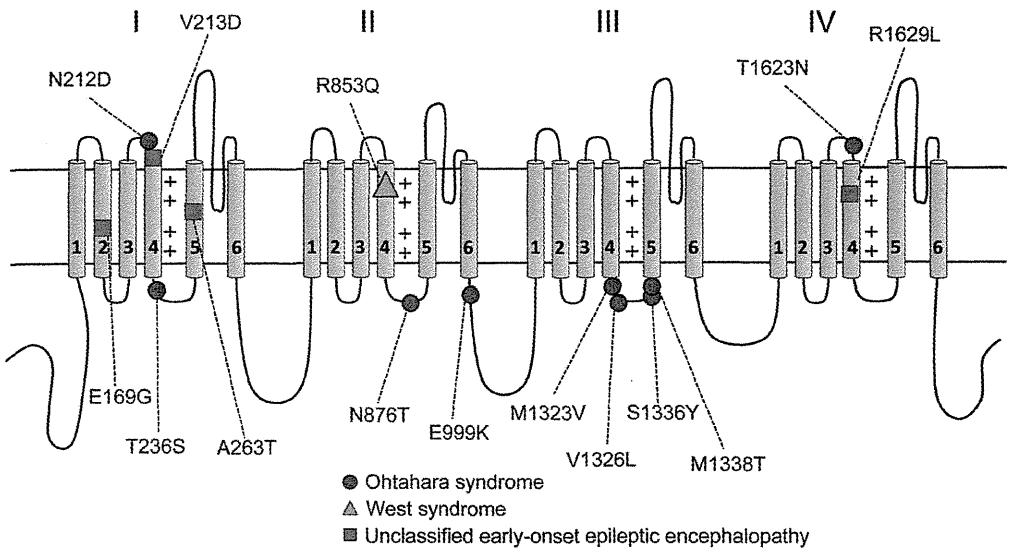
RESULTS Identification of *SCN2A* mutations. We previously performed WES of 12 patients with OS, including patient 142 of the current study.[7,25] To systematically find de novo mutations, we additionally analyzed the parents of patient 142 by WES. We found 2 de novo nonsynonymous mutations in patient 142: c.4868C>A (p.T1623N) in *SCN2A* and c.538A>T (p.I180F) in *FHOD1*. Because de novo *SCN2A* mutations have been reported to cause intractable epilepsy,[19,21] the *SCN2A* mutation was highly likely to have caused the OS. In fact, we found a total of 14 mutations in 15 patients: 9 of 67 OS cases (13.4%), 1 of 150 WS cases (0.67%), and 5 of 111 with unclassified EOEEs (4.5%) (table 1). One mutation (c.638T>A) occurred recurrently in 2 patients, and the mutation was confirmed as a de novo event in 1 of the 2 patients. All other mutations except for 2 (c.2995G>A and c.4886G>T; parents were unavailable) occurred de novo. Interestingly, a de novo mosaic mutation (c.3976G>C) was detected by WES in patient 10. The mutant allele frequency was 18% (37/205 alleles) based on read counts of WES. We confirmed mosaicism in hair, nail, and saliva DNA samples (figure e-1). One mutation (c.634A>G [p.N212D]) was found in the transcript variant 3, which is a neonatal isoform. All of the 15 mutations are missense changes that have not been previously reported. They cannot be found in the 6,500 exomes sequenced by the National Heart, Lung, and Blood Institute exome project or among our 212 in-house control exomes. Sorting Intolerant From Tolerant (SIFT), Polyphen2, and Mutation Taster predicted that all mutations would be highly damaging to the structure of $Na_v1.2$ (table e-2). All 9 mutations in patients with OS were located in linker regions between transmembrane segments (figure 1). Six of these mutations were located in a linker between S4 and S5, 2 were located between S3 and S4, and one near the end of S6. One

Table 1 Clinical features of *SCN2A* mutations

Patient	Sex	Dx	Mutation, inheritance	Age at onset	Initial epileptic attacks	Initial EEG at onset	Transition of seizure and EEG (age)	Response to treatment	HC-SD (age)	DD	MRI findings (age)
10	F	OS (→WS)	c.3976G>C, p.V1326L (mosaic), de novo	8 d	Spasms	SB	TS, myoclonus (2 y); hypsarrhythmia (4 mo); multifocal spikes (3 y, 2 mo)	Intractable	−0.4 (0 d)	Severe	Cerebral atrophy, delayed myelination (2 y, 2 mo)
18	F	EOEE	c.638T>A, p.V213D, N/A	3 mo	Focal seizure	Focal spikes	TS (7 y)	Intractable	N/A	Severe	Mild cerebellar and cerebral atrophy (6 mo); delayed myelination (9 mo)
99	F	WS	c.2558G>A, p.R853Q, de novo	10 mo	Spasms	Hypsarrhythmia	Multifocal spikes, MISF (1 y, 6 mo); no epileptic discharges (6 y, 2 mo)	Seizure-free after LTG at 6 y, 2 mo	N/A	Severe	Cerebral atrophy (frontal, temporal), cerebellar atrophy, thin CC (4 y, 4 mo)
142	M	OS (→WS)	c.4868C>A, p.T1623N, de novo	1 d	TS	SB	Multifocal spikes (2 mo); spasms (3 mo); hypsarrhythmia (3 mo); mainly slow wave (1 y, 4 mo)	Intractable	−1.1 (1 y, 4 mo)	Severe	Thin CC (3 mo); cerebral atrophy (1 y, 4 mo)
146	F	OS	c.707C>G, p.T236S, de novo	0 d	Focal seizure	SB	MISF (5 mo)	Seizure-free after ZNS, LEV, PHT, VGB	N/A	Severe	Thin CC, white matter atrophy, T2 hyperintensity within globi pallidi (1 y, 3 mo)
185	M	OS (→WS)	c.4007C>A, p.S1336Y, de novo	1 d	Myoclonic, TS	SB	Modified hypsarrhythmia (1 y, 3 mo)	Intractable	−0.4 (1 d)	Severe, died at 5 y	Normal (1 mo)
207	M	OS (→WS)	c.3967A>G, p.M1323V, de novo	13 d	GTCS	SB	Hypsarrhythmia (2.5 mo); multifocal polyspikes (3 mo); TS (5 mo); multifocal spikes (8 y)	Intractable	N/A	Severe	Mild cortical atrophy (7 y)
230	F	OS	c.4013T>C, p.M1338T, de novo	7 d	Spasms	SB	Focal seizure (3 mo); multifocal spikes (11 mo)	Intractable	−2.0 (5 y, 3 mo)	Severe	Normal (1 mo); cerebral atrophy (4 y, 9 mo)
251	F	EOEE	c.638T>A, p.V213D, de novo	1.5 mo	TS	Multifocal spikes	Multifocal spikes and diffuse S-W (14 y)	Intractable	−0.8 (4 mo)	Severe	Delayed myelination, cerebral and cerebellar atrophy (1 y); thin CC (14 y)
252	M	OS (→WS)	c.634A>G (variant 3), p.N212D, de novo	14 d	Pedaling	SB	TS, eyelid myoclonia (3 mo); spasms (4 mo); hypsarrhythmia (3 mo); TS (9 mo)	Intractable	N/A	Severe	Cerebral atrophy (16 d); severe cerebral atrophy, delayed myelination (4 y)
254	F	EOEE	c.506A>G, p.E169G, de novo	6 mo	TS	Multifocal spikes	Febrile seizure (6 mo); myoclonic seizure, focal seizure, TS (1 y)	Intractable	−1.7 (5 y, 8 mo)	Severe	Cerebral atrophy, thin CC (5 y)
255	F	EOEE	c.4886G>T, p.R1629L, N/A	3 d	Focal seizure, myoclonus	Focal spikes	Burst of spikes, rhythmic slow wave (3 mo)	Intractable	−0.4 (4 mo)	Severe	Thin CC and hypoplasia of white matter, enlarged lateral ventricle (3 mo)
271	M	OS (→WS)	c.2627A>C, p.N876T, de novo	8 d	TS, eye deviation, mouth automatism	SB	Spasms (5 mo); hypsarrhythmia (5 mo); focal seizure (12 mo)	Intractable	−0.8 (2 mo)	Severe	Mild cerebral atrophy, delayed myelination (10 mo)
305	F	EOEE	c.787G>A, p.A263T, de novo	3 d	CS and TS	Multifocal spikes	Modified hypsarrhythmia (2 mo); mainly slow waves (10 mo)	Seizure-free after LTG at 6 mo	−0.1 (0 d)	Severe	Thin CC (2.5 mo)
322	M	OS (→WS)	c.2995G>A, p.E999K, N/A	6 d	CS	SB	Hypsarrhythmia (3 mo); spasms (4 mo)	Intractable	−0.6 (3 mo)	Severe	No abnormality (2 mo)

Abbreviations: CC = corpus callosum; CS = clonic seizure; DD = developmental delay; Dx = diagnosis; EOEE = early-onset epileptic encephalopathy; GTCS = generalized tonic-clonic seizure; HC-SD = head circumference SD; LEV = levetiracetam; LTG = lamotrigine; MISF = multiple independent spike foci; N/A = not available; OS = Ohtahara syndrome; PHT = phenytoin; SB = suppression-burst pattern; S-W = spike and slow wave; TS = tonic seizure; VGB = vigabatrin; WS = West syndrome; ZNS = zonisamide.

Figure 1 Structure of the human $Na_v1.2$ channel with localization of *SCN2A* mutations



Red circles = Ohtahara syndrome; yellow triangle = West syndrome; blue squares = unclassified early-onset epileptic encephalopathy.

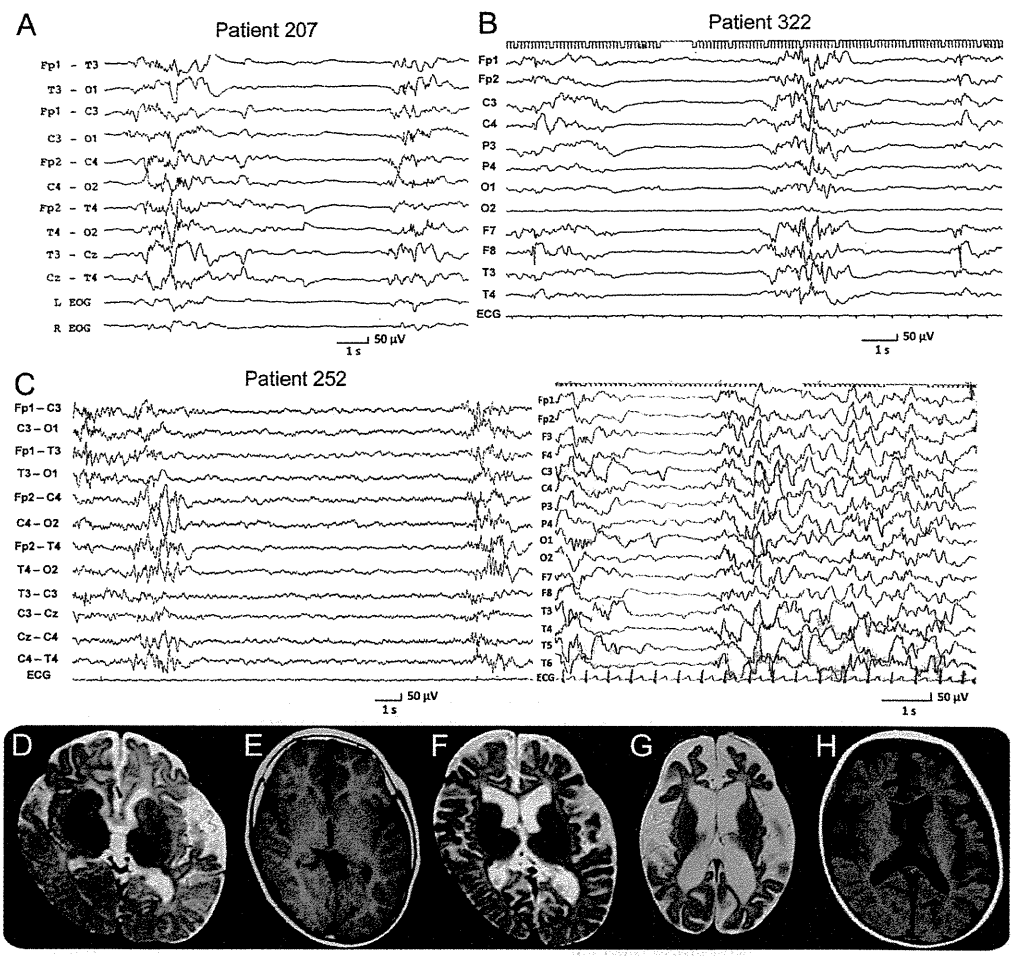
mutation in a patient with WS was located in a positively charged segment (S4) of domain II. Other mutations found in patients with unclassified EOEE were located in S4 of domain IV (1 mutation), a linker between S3 and S4 of domain I (1 mutation), and in S2 and S5 of domain I (2 mutations).

Clinical features of *SCN2A* mutations. The clinical information and EEG and MRI findings of patients with *SCN2A* mutations are summarized in table 1 and figure 2. Case reports (in appendix e-1) and further EEG (in figure e-2) and MRI (in figure e-3) findings are shown in supplemental data. There were 9 patients with OS (6 Japanese, 1 Canadian, 1 Mexican, and 1 Israeli), 1 with WS (Japanese), and 5 with unclassified EOEE (4 Japanese and 1 Slovenian). The ratio of females to males was 9:6. The mean age at onset was 45.3 days after birth (5.7 days in patients with OS; 65 days in unclassified EOEE; 10 months in a patient with WS). The initial epileptic attacks were tonic seizures in 8 patients, spasms in 3 patients, clonic seizures in 3 patients, and generalized tonic-clonic seizures in 1 patient. Three patients (patients 99, 254, and 305) had dystonia, and patient 99 also had chorea and ballismus. EEG at epilepsy onset showed suppression-burst pattern in 9 patients and focal or multifocal spikes in 6 patients. Seven of 15 patients with *SCN2A* mutations received adrenocorticotropic hormone therapy, leading to a transient reduction of seizures followed by recurrence. Epileptic seizures disappeared and EEG findings improved in 2 patients with WS and unclassified EOEE after administration of lamotrigine (patients 99 and 305), and in 1 patient with OS after combination therapy with zonisamide, levetiracetam, phenytoin, and vigabatrin (patient 146). Twelve of the 15 patients could not be controlled with antiepileptic drugs. All 15 patients had severe developmental delay, and 1 patient with unclassified EOEE (patient 185) died at the age of 5 years. Brain MRI showed cerebral atrophy in 10 patients, thin corpus callosum in 7 patients, delayed myelination in 5 patients, and cerebellar atrophy in 3 patients. MRI was normal in 2 patients in the early infantile period (1 and 2 months of age). Of the 9 patients with OS, 6 showed spasms, and all patients showed suppression-burst pattern on EEG at epilepsy onset. Seven patients with OS developed WS, and 1 case (patient 230) was suspected to do so because of spasms and developmental delay; however, no EEG was done to verify this.

DISCUSSION We identified 14 novel *SCN2A* mutations in 15 patients with EOEE. Two of the mutations could not be confirmed as de novo, but all mutations showed high scores for predicted negative effects on protein function, and were not found in our in-house control exome data. Of note, 9 of the 15 patients were diagnosed with OS, and all patients showed severe developmental delay. Previously, *SCN2A* mutations have been reported in a wide spectrum of disorders, including autism, severe nonsyndromic sporadic intellectual disability, benign seizure syndromes such as BFNIS, benign familial infantile seizure, and generalized epilepsy with febrile seizures plus, and severe epileptic disorders such as DS, infantile spasms, and AERRPS (acute encephalitis with refractory, repetitive partial seizures).[18–22,26–33]

Most of the disorders caused by *SCN1A* mutations involve DS or generalized epilepsy with febrile seizures

Figure 2 EEG and brain MRI of patients with *SCN2A* mutations



(A, B) Suppression-burst pattern on interictal EEG (A, patient 207 at an age of 18 days; B, patient 322 at an age of 7 days). (C) Suppression-burst pattern on interictal EEG of patient 252 at an age of 14 days (left). Transition from Ohtahara syndrome to West syndrome with modified hypsarrhythmia at an age of 3 months (right). (D–H) Brain MRI axial images at the level of the basal ganglia (D, T2-weighted image of patient 10 at 2 years and 2 months; E, T1-weighted image of patient 207 at 7 years; F, T2-weighted image of patient 251 at 1 year; G, T2-weighted image of patient 252 at 10 months; H, T1-weighted image of patient 271 at 10 months) showing cortical atrophy and delayed myelination (D, F–H) and cortical atrophy of the left temporal region (E).

plus, and onset of DS is usually in the infantile period.[17] By contrast, the onset of OS is neonatal and that of BFNIS is neonatal or early infantile. The phenotypic differences between *SCN2A* and *SCN1A* mutations might be explained by the different expression patterns of the channels. $Na_v1.2$ (SCN2A) and $Na_v1.6$ (SCN8A) are the major sodium channel α subunits in excitatory neurons. $Na_v1.2$ channels are expressed early in development and $Na_v1.6$ channels gradually replace $Na_v1.2$ channels in a population of neurons during maturation.[31] $Na_v1.2$ is expressed predominantly at terminals, unmyelinated axons, and in proximal axon initial segments, where action potentials are initiated, whereas $Na_v1.1$ (SCN1A) is expressed predominantly in neurons releasing γ-aminobutyric acid, and is localized to the soma and proximal processes of the neuron.[13,14,34,35] Therefore, *SCN1A* mutations may mainly lead to a reduction of sodium-channel activity in inhibitory neurons, increasing net excitability. However, *SCN2A* mutations likely also affect excitatory neurons and proximal axon initial segments especially during the early developmental period. Therefore, the effect of *SCN2A* mutations on the net excitability of neurons may result in earlier onset of epileptic disorders.

Investigation of the genotype-phenotype correlation of *SCN2A* mutations may contribute to our understanding of the pathophysiologic mechanisms of seizures caused by *SCN2A* mutations. *SCN2A* mutations causing BFNIS tend to be located in transmembrane regions (7/11 mutations), whereas mutations causing intractable epilepsies were located outside transmembrane domains, such as upstream of domain

I or between 2 domains.[18,19,21,26,28] In this study, all 9 mutations causing OS were found in linkers between transmembrane regions of $Na_v1.2$, especially between S4 and S5 of domain III. The S4-S5 linker of domain III interacts with the linker of domains III and IV, which is critical for fast inactivation of the sodium channel,[12,36] suggesting that alteration of this linker region may specifically affect the function of $Na_v1.2$. Similarly, the mutation causing WS was located in a positively charged segment (S4) of domain II, where no mutations have been previously reported. Interestingly, patient 10 showing OS had a low-level mosaicism of the p.V1326L mutation, which was confirmed in DNA samples from blood leukocytes, hair follicle, nail, and saliva. There are no previous reports of mosaicism for *SCN2A* mutations, but one case of mosaic 2q24 duplication including *SCN2A* and *SCN3A* has been reported.[37] The mosaic duplication was detected in 40% to 51% of blood cells, and the phenotype partially overlapped with both DS and BFNIS.[37] In cases of *SCN1A* mutations in familial DS, mosaic carriers with approximately 25% mutated alleles had experienced simple febrile seizure, whereas a carrier with approximately 12.5% mutated alleles was asymptomatic.[38] Patient 10 showed a very severe OS phenotype despite the low rate of mosaicism (18%). However, we did not determine whether the same mosaic rate was maintained in brain tissue. The location of the p.V1326L mutation in a linker of S4 and S5 of domain III might support the hypothesis that alteration of linker regions, especially between S4 and S5 of domain III, uniquely affects the function of $Na_v1.2$ in association with OS.

Patient 252 with OS had a de novo mutation in exon 6N, which is a neonatal isoform (transcript of variant 3). *SCN1A*, *SCN2A*, *SCN3A*, and *SCN8A* are subject to alternative splicing of exon 6N and 6A. In the developing mouse brain, *Scn2a* shows higher or equal amounts of the 6N isoform compared with the 6A isoform at birth, but 6N is gradually replaced by 6A during postnatal development.[39] The channels of the neonatal isoform are less excitable than the adult isoform.[40] In patient 252, the age of seizure onset was 14 days, which was relatively late compared with the other cases of OS. The expression of abnormal neonatal isoforms would be expected to gradually decrease, leading to an alleviation of neuronal hyperexcitability. In contrast, his symptoms, such as the transition to WS, development delay, and cerebral atrophy, worsened with increasing age. Further studies are required to clarify how mutations of the 6N isoform are involved in pathogenesis of EOEE.

To date, many disease-causing genes contributing to EOEE phenotypes have been identified, including *ARX*, *CDKL5*, *STXBP1*, *SPTAN1*, *KCNQ2*, *SCN1A*, and *SCN2A*.[5–11,17,20,21] In addition, mutations in one gene can cause a wide phenotypic spectrum as revealed by previous reports and the results of this study.[6,7,25] Therefore, rapid genetic diagnosis of EOEEs by screening for all known disease-causing genes with Sanger sequencing is very difficult. Recently, targeted exome sequencing has been developed and applied as a diagnostic tool for epileptic disorders.[33] Targeted exome sequencing as well as WES would be quite useful for genetic testing for EOEEs.

AUTHOR CONTRIBUTIONS

Kazuyuki Nakamura and Mitsuhiro Kato: study concept and design, analysis of the clinical data, interpretation of the data, and drafting/revising of the manuscript. Hitoshi Osaka, Sumimasa Yamashita, Eiji Nakagawa, Kazuhiro Haginoya, Jun Tohyama, Mitsuko Okuda, Takahito Wada, Shuichi Shimakawa, Katsumi Imai, Saoko Takeshita, Hisako Ishiwata, Dorit Lev, Tally Lerman-Sagie, David E. Cervantes-Barragán, Camilo E. Villarroel, Masaharu Ohfu, Karin Writzl, Barbara Gnidovec Stražišar, Shinichi Hirabayashi, David Chitayat, and Diane Myles Reid: analysis of the clinical data and sample collection. Kiyomi Nishiyama, Hirofumi Kodera, Mitsuko Nakashima, Yoshinori Tsurusaki, and Noriko Miyake: analysis of the genetic data. Kiyoshi Hayasaka: analysis of the clinical data and sample collection. Naomichi Matsumoto: study concept and design, interpretation of the data, and drafting/revising of the manuscript. Hirotomo Saitsu: study concept and design, analysis of the genetic data, interpretation of the data, and drafting/revising of the manuscript.

ACKNOWLEDGMENT

The authors thank the patients and their families for their participation in this study, and Aya Narita for her technical assistance.

STUDY FUNDING

Supported by the Ministry of Health, Labour and Welfare of Japan (24133701, 11103577, 11103340, 10103235), a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (24591500), a Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science (10013428, 12020465), the Takeda Science Foundation, the Japan Science and Technology Agency, the Strategic Research Program for Brain Sciences (11105137), and a Grant-in-Aid for Scientific Research on Innovative Areas (Transcription Cycle) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (12024421).

DISCLOSURE

K. Nakamura, M. Kato, H. Osaka, S. Yamashita, E. Nakagawa, and K. Haginoya report no disclosures. J. Tohyama is an investigator in clinical trials sponsored by Meiji Seika Pharma Co. Ltd., Novartis Pharma K.K., and UCB Japan Co. Ltd. M. Okuda, T. Wada, S. Shimakawa, K. Imai, S. Takeshita, H. Ishiwata, D. Lev, T. Lerman-Sagie, D. Cervantes-Barragán, C. Villarroel, M. Ohfu, K. Writzl, B. Gnidovec Stražišar, S. Hirabayashi, D. Chitayat, D. Myles Reid, K. Nishiyama, H. Kodera, M. Nakashima, and Y. Tsurusaki report no disclosures. N. Miyake is funded by research grants from the Ministry of Health, Labour and Welfare of Japan, a Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science, and a Research Grant from the Takeda Science Foundation. K. Hayasaka reports no disclosures. N. Matsumoto is supported by grants from the Ministry of Health, Labour and Welfare of Japan, a Grant-in-Aid for Scientific Research (A) from the Japan Society for the Promotion of Science, the Takeda Science Foundation, the Japan Science and Technology Agency, the Strategic Research Program for Brain Sciences, a Grant-in-Aid for Scientific Research on Innovative Areas (Transcription Cycle) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. H. Saitsu is funded by research grants from the Ministry of Health, Labour and Welfare of Japan, and a Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science. Go to Neurology.org for full disclosures.

Received March 14, 2013. Accepted in final form June 6, 2013.

REFERENCES

1. Berg AT, Berkovic SF, Brodie MJ, et al. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia 2010;51:676–685.
2. Mastrangelo M, Leuzzi V. Genes of early-onset epileptic encephalopathies: from genotype to phenotype. Pediatr Neurol 2012;46:24–31.
3. Ohtahara S, Yamatogi Y. Ohtahara syndrome: with special reference to its developmental aspects for differentiating from early myoclonic encephalopathy. Epilepsy Res 2006;70(suppl 1):S58–S67.
4. Kato M. A new paradigm for West syndrome based on molecular and cell biology. Epilepsy Res 2006;70(suppl 1):S87–S95.
5. Saitsu H, Kato M, Mizuguchi T, et al. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. Nat Genet 2008;40:782–788.
6. Weckhuysen S, Mandelstam S, Suls A, et al. *KCNQ2* encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. Ann Neurol 2012;71:15–25.
7. Saitsu H, Kato M, Koide A, et al. Whole exome sequencing identifies *KCNQ2* mutations in Ohtahara syndrome. Ann Neurol 2012;72:298–300.
8. Weaving LS, Christodoulou J, Williamson SL, et al. Mutations of *CDKL5* cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. Am J Hum Genet 2004;75;1079–1093.
9. Kato M, Saitoh S, Kamei A, et al. A longer polyalanine expansion mutation in the *ARX* gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome). Am J Hum Genet 2007;81:361–366.
10. Kato M, Das S, Petras K, Sawaishi Y, Dobyns WB. Polyalanine expansion of *ARX* associated with cryptogenic West syndrome. Neurology 2003;61:267–276.
11. Saitsu H, Tohyama J, Kumada T, et al. Dominant-negative mutations in alpha-II spectrin cause West syndrome with severe cerebral hypomyelination, spastic quadriplegia, and developmental delay. Am J Hum Genet 2010;86:881–891.
12. Catterall WA. From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. Neuron 2000;26:13–25.
13. Whitaker WR, Faull RL, Waldvogel HJ, Plumpton CJ, Emson PC, Clare JJ. Comparative distribution of voltage-gated sodium channel proteins in human brain. Brain Res Mol Brain Res 2001;88:37–53.
14. Oliva M, Berkovic SF, Petrou S. Sodium channels and the neurobiology of epilepsy. Epilepsia 2012;53:1849–1859.
15. Zuberi SM, Brunklaus A, Birch R, Reavey E, Duncan J, Forbes GH. Genotype-phenotype associations in *SCN1A*-related epilepsies. Neurology 2011;76:594–600.
16. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;485:246–250.
17. Guerrini R. Dravet syndrome: the main issues. Eur J Paediatr Neurol 2012;16(suppl 1):S1–S4.
18. Heron SE, Crossland KM, Andermann E, et al. Sodium-channel defects in benign familial neonatal-infantile seizures. Lancet 2002;360:851–852.
19. Kamiya K, Kaneda M, Sugawara T, et al. A nonsense mutation of the sodium channel gene *SCN2A* in a patient with intractable epilepsy and mental decline. J Neurosci 2004;24:2690–2698.
20. Shi X, Yasumoto S, Nakagawa E, Fukasawa T, Uchiya S, Hirose S. Missense mutation of the sodium channel gene *SCN2A* causes Dravet syndrome. Brain Dev 2009;31:758–762.
21. Ogiwara I, Ito K, Sawaishi Y, et al. De novo mutations of voltage-gated sodium channel alphaII gene *SCN2A* in intractable epilepsies. Neurology 2009;73:1046–1053.
22. Kobayashi K, Ohzono H, Shinohara M, et al. Acute encephalopathy with a novel point mutation in the *SCN2A* gene. Epilepsy Res 2012;102:109–112.
23. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–498.
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
25. Saitsu H, Kato M, Osaka H, et al. *CASK* aberrations in male patients with Ohtahara syndrome and cerebellar hypoplasia. Epilepsia 2012;53:1441–1449.
26. Herlenius E, Heron SE, Grinton BE, et al. *SCN2A* mutations and benign familial neonatal-infantile seizures: the phenotypic spectrum. Epilepsia 2007;48:1138–1142.
27. Sugawara T, Tsurubuchi Y, Agarwala KL, et al. A missense mutation of the Na^+ channel alpha II subunit gene Na(v)1.2 in a patient with febrile and afebrile seizures causes channel dysfunction. Proc Natl Acad Sci USA 2001;98:6384–6389.
28. Berkovic SF, Heron SE, Giordano L, et al. Benign familial neonatal-infantile seizures: characterization of a new sodium channelopathy. Ann Neurol 2004;55:550–557.
29. Ito M, Shirasaka Y, Hirose S, Sugawara T, Yamakawa K. Seizure phenotypes of a family with missense mutations in *SCN2A*. Pediatr Neurol 2004;31:150–152.
30. Striano P, Bordo L, Lispi ML, et al. A novel *SCN2A* mutation in family with benign familial infantile seizures. Epilepsia 2006;47:218–220.
31. Liao Y, Deprez L, Maljevic S, et al. Molecular correlates of age-dependent seizures in an inherited neonatal-infantile epilepsy. Brain 2010;133:1403–1414.
32. Need AC, Shashi V, Hitomi Y, et al. Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet 2012;49:353–361.
33. Lemke JR, Riesch E, Scheurenbrand T, et al. Targeted next generation sequencing as a diagnostic tool in epileptic disorders. Epilepsia 2012;53:1387–1398.
34. Furuyama T, Morita Y, Inagaki S, Takagi H. Distribution of I, II and III subtypes of voltage-sensitive Na^+ channel mRNA in the rat brain. Brain Res Mol Brain Res 1993;17:169–173.
35. Ogiwara I, Miyamoto H, Morita N, et al. Nav1.1 localizes to axons of parvalbumin-positive inhibitory interneurons: a circuit basis for epileptic seizures in mice carrying an *Scn1a* gene mutation. J Neurosci 2007;27:5903–5914.
36. Smith MR, Goldin AL. Interaction between the sodium channel inactivation linker and domain III S4-S5. Biophys J 1997;73:1885–1895.
37. Vecchi M, Cassina M, Casarin A, et al. Infantile epilepsy associated with mosaic 2q24 duplication including *SCN2A* and *SCN3A*. Seizure 2011;20:813–816.
38. Shi YW, Yu MJ, Long YS, et al. Mosaic *SCN1A* mutations in familial partial epilepsy with antecedent febrile seizures. Genes Brain Behav 2012;11:170–176.
39. Gazina EV, Richards KL, Mokhtar MB, Thomas EA, Reid CA, Petrou S. Differential expression of exon 5 splice variants of sodium channel alpha subunit mRNAs in the developing mouse brain. Neuroscience 2010;166:195–200.
40. Xu R, Thomas EA, Jenkins M, et al. A childhood epilepsy mutation reveals a role for developmentally regulated splicing of a sodium channel. Mol Cell Neurosci 2007;35:292–301.

OPEN ACCESS Freely available online

PLOS ONE

Performance Comparison of Bench-Top Next Generation Sequencers Using Microdroplet PCR-Based Enrichment for Targeted Sequencing in Patients with Autism Spectrum Disorder

Eriko Koshimizu[1☯], Satoko Miyatake[1☯], Nobuhiko Okamoto[2], Mitsuko Nakashima[1], Yoshinori Tsurusaki[1], Noriko Miyake[1], Hirotomo Saitsu[1], Naomichi Matsumoto[1]*

1 Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan, **2** Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan

Abstract

Next-generation sequencing (NGS) combined with enrichment of target genes enables highly efficient and low-cost sequencing of multiple genes for genetic diseases. The aim of this study was to validate the accuracy and sensitivity of our method for comprehensive mutation detection in autism spectrum disorder (ASD). We assessed the performance of the bench-top Ion Torrent PGM and Illumina MiSeq platforms as optimized solutions for mutation detection, using microdroplet PCR-based enrichment of 62 ASD associated genes. Ten patients with known mutations were sequenced using NGS to validate the sensitivity of our method. The overall read quality was better with MiSeq, largely because of the increased indel-related error associated with PGM. The sensitivity of SNV detection was similar between the two platforms, suggesting they are both suitable for SNV detection in the human genome. Next, we used these methods to analyze 28 patients with ASD, and identified 22 novel variants in genes associated with ASD, with one mutation detected by MiSeq only. Thus, our results support the combination of target gene enrichment and NGS as a valuable molecular method for investigating rare variants in ASD.

Citation: Koshimizu E, Miyatake S, Okamoto N, Nakashima M, Tsurusaki Y, et al. (2013) Performance Comparison of Bench-Top Next Generation Sequencers Using Microdroplet PCR-Based Enrichment for Targeted Sequencing in Patients with Autism Spectrum Disorder. PLoS ONE 8(9): e74167. doi:10.1371/journal.pone.0074167

Editor: Takeo Yoshikawa, Rikagaku Kenkyūsho Brain Science Institute, Japan

Received June 5, 2013; **Accepted** July 29, 2013; **Published** September 16, 2013

Funding: Research grants from the Ministry of Health, Labour and Welfare: the Japan Science and Technology Agency: and the Strategic Research Program for Brain Sciences, Grant-in-Aid for Scientific Research on Innovative Areas (Transcription cycle) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science, Grant-in-Aid for Young Scientist from Japan Society for the Promotion of Science, and Grant from the Takeda Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: naomat@yokohama-cu.ac.jp

☯ These authors contributed equally to this work.

Introduction

Recent advances in next generation sequencing (NGS) technologies combined with efficient gene enrichment, allows the comprehensive resequencing of multiple known causative or associated genes in highly heterogeneous diseases. In addition, these technologies make it possible to perform resequencing more inexpensively and rapidly than the conventional Sanger method. Higher sequencing accuracy due to the deeper achievable coverage with the aid of improved bioinformatic analysis is expected as well. Different bench-top next generation DNA sequencers are currently available for target resequencing. Each NGS machine adopts specific technologies, thus the property and/or quality of sequence reads is likely different. However there is little comparative evidence on the data quality between sequencers used in human gene analysis.

Autism spectrum disorder (ASD) is a complex disorder with several hundred associated loci, following a polygenic mode of inheritance [1]. It is relatively common, with a prevalence of 1.1% [2], and is typically a child-onset disorder characterized by impaired social interactions, communication deficits, and restricted and repetitive behaviors [3]. It is known to be highly heritable, yet the majority of its heritability is so far unresolved [4]. Previous studies suggest a genetic contribution, consisting of both common and rare alleles, accounts for a portion of ASD risk, with a heritability of 38–90% [4–8]. Considering the frequency and socio-economic impact of ASD, verification of the actual heritability of ASD is of importance. Common single-nucleotide variants (SNVs) have been reported as a major source of ASD risk, with the heritability exceeding 40% [7]. However, their impact on ASD development is relatively small in each case, with an estimated odds ratio (OR) <1.2 [9]. Conversely, rare variants occurring *de novo* or inherited are assumed to affect ASD risk as well [1,10–13]. Recent work revealed a larger effect of *de novo* SNVs, although they accounted for only a small portion of overall ASD risk, with an estimated 10% contribution to ASD risk [10–13]. Recently, an additive 5% contribution to ASD risk was reported in rare complete knockouts, derived from inheriting rare

recessive variations [14]. To further explore the missing ASD risk heritability, a promising approach would be to comprehensively identify rare variants that have additive gene effects or show a multigenic epistatic contribution.

Here we have developed a rapid, cost-effective and comprehensive analysis workflow for detecting rare variants in ASD patients. We screened 62 known ASD associated genes using microdroplet PCR-based technology, together with the Ion Torrent Personal Genome Machine (PGM) and MiSeq platforms. To validate the systems, we sequenced 10 positive controls with other diseases and 28 ASD patients. Sequencing data produced by the two sequencers were compared, demonstrating successful identification of positive control variants and novel SNVs associated with ASD.

Materials and Methods

Ethics statement

Written informed consents were obtained from all patients or their parents. Experimental protocols were approved by the Committee for Ethical Issue at Yokohama City University School of Medicine.

Patients

A total of 28 ASD patients, diagnosed according to DSM IV-TR criteria, and 10 patients with other identified diseases with known mutations in one of the target genes, were used for this study. DNA was obtained from peripheral blood leukocytes.

RainDance library preparation and DNA enrichment

The RainDance ASDSeq™ Research Screening Panel was provided by RainDance Technologies™ (Lexington, MA, USA). The RainDance ASDSeq™ panel is a genetic screening tool that offers >92% coverage of 62 genes containing known mutations associated with ASD. The library contains 2349 amplicons ranging in size from 167 to 600 bp and covering a 1034 kb region. Coverage includes all exons for each gene plus 50 bp up- and downstream of each exon, to capture intron/exon splice junctions, as well as 1 kb of both the 5′ promoter region and 3′ UTRs. The panel includes both autosomal and X-linked genes.

A total of 2.5 μg of genomic DNA was used for DNA enrichment. The primer library and a template mix, including 1.5 μg of fragmented genomic DNA and all the PCR reaction components except the primers, were loaded on the RainDance for PCR droplet preparation, according to the manufacturer's instructions. Samples were run on the RDT 1000 machine and PCR droplets were generated. The PCR droplets were amplified under the following conditions: 94°C for 2 min, then 54 cycles of 94°C for 30 sec, 54°C for 30 sec and 68°C for 60 sec, followed by 68°C for 10 min and 4°C for holding. After amplification, the PCR droplets were broken to release the amplicons. The amplicons were purified and quantified using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The ends of the DNA fragments were repaired at 25°C for 30 min using New England BioLabs End Repair Module (New England BioLabs, Ipswich, MA, USA), followed by purification using Qiagen MinElute columns (Qiagen, Valencia, CA, USA). The PCR fragments were concatenated at 20°C for 30 min using NEB Quick Ligation Kit (New England BioLabs). The ligated products were purified using the Qiagen MinElute columns and fragmented using a Covaris S2 machine (duty cycle 10%, intensity 5, cycle/burst 200, total time per treatment 430 s).

Sequencing using ion torrent PGM and data processing

Library preparation was carried out using the Ion Plus Fragment Library Kit, with 50 ng of amplicons. Adapter ligation, nick repair and amplification were performed as described in the Ion Torrent protocol (Ion Plus Fragment Library Kit; Part Number 4471989 Rev. B; Life Technologies, Grand Island, NY, USA). The Agilent 2100 Bioanalyzer (Agilent Technologies) and associated High Sensitivity DNA kit (Agilent Technologies) were used to determine quality and concentration of the libraries. Emulsion PCR and enrichment steps were carried out using the Ion OneTouch™ Template Kit (Life Technologies) and associated protocol (Part Number 4472430 Rev. C). Sequencing of the amplicon libraries was carried out on the Ion Torrent PGM system using 316 or 318 chips, and barcoding with Ion Xpress™ Barcode Adapters 1–16 Kit (Life Technologies). The Ion Sequencing Kit v2 (Life Technologies) was used for all sequencing reactions (expected read length was 100 bp), following the recommended protocol (Part Number 4469714 Rev. B). After sequencing, reads were mapped to hg19 using Torrent Mapping Alignment Program (TMAP). TMAP is a customized mapping tools for sequencing data generated by PGM, ignoring the indel calls around homopolymer stretch to reduce the hundreds of false negative calls. Torrent Suite 2.0 and/or 3.2 were used for all analyses. Coverage depth was calculated using Torrent Coverage Analysis. SNVs and small insertions/deletions (indels) were identified using the Torrent Variant Caller. Common variants (MAF ≥1%) registered in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/) without a flag as clinically associated, or ones in the lower versions of dbSNP, were filtered out. Filter-passed variants were annotated using ANNOVAR [15] and a custom pipeline. In order to compare the ability of mutation detection, reads of positive controls were aligned to GRCh37 with Novoalign v3.00 (Novocraft Technologies, Selangor, Malaysia) with the parameters for PGM and Local realignments around indels and base quality score recalibration were performed using the Genome Analysis Toolkit (GATK) v1.5–21 [16]. SNVs and small indels were identified using the GATK UnifiedGenotyper.

Sequencing using MiSeq and data processing

The same amplicons were sequenced on the Illumina MiSeq sequencer, using the SureSelect^{XT} Reagents (Agilent Technologies) protocol, with 50 ng input material. Each multiplex library pool was sequenced on an Illumina MiSeq for 150 cycles from each end, plus a 6 base-index sequence read, using the MiSeq Reagent Kit (Illumina, San Diego, CA, USA). Image analysis and base calling were performed using sequence control software with real-time analysis, and Consensus Assessment of Sequence and Variation (CASAVA) software v1.8 (Illumina). Reads were aligned to GRCh37 with Novoalign v2.08 (Novocraft Technologies), and Local realignments around indels and base quality score recalibration were performed using the GATK v1.5–21 [16]. SNVs and small indels were identified using the GATK UnifiedGenotyper, and filtered according to the Broad Institute's best-practice guidelines v3. Common variants (MAF ≥1%) registered in dbSNP135 (http://www.ncbi.nlm.nih.gov/projects/SNP/) without a flag as clinically associated, or ones in the lower versions of dbSNP, were filtered out. Filter passed variants were annotated using ANNOVAR [15] and a custom pipeline.

Quality validation of sequence reads

For quality comparison, we combined sequencing data from four random samples obtained by either PGM or MiSeq and evaluated the average quality of data from multiple samples. Box plots for base-call quality of combined runs from each sequencer

were generated using fastqc software (Babraham Bioinformatics, Cambridge, UK). To count the number of single nucleotide polymorphisms (SNPs) and short indels in our combined sequencing data, we used samtools mpileup command with the minimum mapping quality assignment option. We excluded calls with either a depth ≤10 or genotype quality ≤30.

Validation of novel variants

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org/www/SIFT_BLink_submit.html), MutationTaster (http://www.mutationtaster.org/) and Genomic Evolutionary Rate Profiling (GERP) [17] were used to evaluate SNVs in terms of sequence conservation, chemical change and likelihood of pathogenicity. The Human Gene Mutation Database (Biobases, Wolfenbuettel, Germany; (https://portal.biobase-international.com/hgmd/pro/start.php) was used for determining if variants were previously reported.

Sanger confirmation of variants detected by next-generation sequencing

Possible pathological variants were confirmed by Sanger sequencing using an ABI 3500×l or ABI 3100 autosequencer (Life Technologies), according to the manufacturer's protocol. Sequencing data was analyzed using sequence analysis software version 5.1.1 (Applied Biosystems, Foster City, CA, USA) and Sequencher 4.10-build 5828 (GeneCodes Corporation, Ann Arbor, MI, USA).

Statistical analysis

All statistical analyses were carried out using SPSS Statistics 19 (IBM, NY, USA). The carrier frequency of each novel SNV was compared between ASD patients and in-house 212 normal Japanese controls using Fisher's exact test. $p<0.05$ was considered statistically significant.

Results

Sequencing yields and targeting efficiency

The targeted NGS panel was designed to amplify all exons of the 62 known ASD associated genes (Table S1). To validate the performance of RainDance sample enrichment and our chosen NGS systems, ten positive controls, each with a mutation in either *NSD1* (c.3958C>T, c.5177C>T, c.5179G>C, c.6499T>C), *MECP2* (c.243_244insC, c.316C>T), *CASK* (c.277_288del), *SCN1A* (c.342_344delinsAGGAGTT, c.4313T>A) or *CDKL5* (c.145G>A) were used. Our workflow strategy is summarized (Table 1). NGS after target enrichment yielded an average of 295.97 (PGM-TMAP), 201.73 (PGM-Novoalign) and 469.42 (MiSeq) Mb of sequence, in which 96.8% (PGM-TMAP), 78.8% (PGM-Novoalign) and 75% of reads were mapped to the genome, and 26.7% (PGM-TMAP), 28.3% (PGM-Novoalign) and 22.7% were mapped to the targeted regions, by PGM and MiSeq, respectively (Table 2). The percentage of mapped bases was greater in PGM-TMAP than in PGM-Novoalign, while the ones in PGM-Novoalign and MiSeq were similar. On-target rate was also similar and generally low in these data. The total coverage of all targeted bases was on average for PGM (TMAP), 93.7% at 10× and 85.9% at 20×, with a mean read depth of 63×, and for MiSeq, 96.8% at 10× and 93.2% at 20×, with a mean read depth of 95× (Table 2). The complete coverage information on the differences between PGM and MiSeq is presented in Table 2. The mean depth of coverage on genes across all samples ranged from 21× for *PTCHD1* to 237× for *NHS*, with an average of 95× by MiSeq. Despite the high mean read depth and target region coverage, several exons including exon 15 of *NIPBL*, exon 43 of *RELN*, exon 2 of *BRAF*, exon 7 of *PTEN*, exon 10 of *SLC6A4*, exon11 of *SHANK3*, exon 43 of *DMD*, exon 8 of *CASK*, exon 36 of *MED12* and exon 2 of *L1CAM*, had no mapped reads from either sequencer. These unmapped regions may be due to sequence complexity, problematic library synthesis necessitating the use of a concatenation step for sample preparation, or unusual GC content of the fragments for the enrichment system. Exon 11 of *SHANK3* has a very high GC content (80%), while exons 2, 43, 15, and 43, of *BRAF*, *DMD*, *NIPBL*, and *RELN*, respectively, have a very low GC content (<35%), and consequently no mapped reads in the NGS data.

Comparison of sequencing quality

The mean base-call quality score obtained from MiSeq was high through entire reads, with a score >30 (Figure S1A, B). The dispersion of scores among reads at specific positions was relatively small. Conversely, the mean base-call quality score obtained from PGM was >25 at the beginning of reads, but gradually decreased to around 20, at approximately base position 100. The dispersion of scores among reads was larger than those obtained using MiSeq. n addition, read lengths produced by each sequencer were different. With MiSeq, all reads had the expected length of 151 bases, whereas with PGM, read lengths were widely distributed from 60 to 150 bp long, although the expected read length was 100 bp (Figure S1C).

Overall, it appeared that the MiSeq output sequences had a higher base-call quality, but it was difficult to compare the scores derived from each sequencer, as PGM and MiSeq adopt different scoring systems for evaluating base-call quality. MiSeq uses Phred [18], while PGM uses a unique Phred-like system consisting of six predictors whose quality values are correlated with the probability of a base miscall. Therefore we compared the mapping quality of each read from both sequencers, as both sequencers adopt the same scoring system for mapping quality [19]. We summed up the total number of reads with a mapping quality >40 and reads <40, and found 94.5% (MiSeq) and 71.2% (PGM) of aligned reads had a mapping score >40 (Figure S1D).

Next we compared the number of indel calls detected by PGM and MiSeq, in the combined data from four individuals randomly chosen (Table S2). With PGM, 9685 SNPs or indels were called, with 5544 indels calls (57.2%). The frequency of indels was calculated as 1.34 per 1 kb per sample. With MiSeq, 3818 SNPs or indels were called, with 395 calls (10.3%) being indels. The frequency of indels was calculated as 0.096 per 1 kb per sample. After filtering the SNP and indel call with a mapping quality >40, and comparing again, 5288 indels out of 7574 total calls (69.8%) were detected with PGM, while 386 indels out of 3553 total calls (10.9%) were detected with MiSeq, leading to an expected frequency of 1.27 indels per 1 kb per sample (PGM) versus 0.093 indels per 1 kb per sample (MiSeq).

Confirmation of variant detection

The ability of PGM and MiSeq to efficiently detect various mutations, including point mutations and small indels, was tested using previously Sanger-confirmed mutations in variant-positive samples (Table 3). The variant-positive samples included all types of variants, including missense, small insertion, small deletion and small indel variants, in the genes *SCN1A*, *NSD1*, *MECP2*, *CDKL5* and *CASK* (Table 3). Some of the insertion and indel variants detected by NGS are shown (Figure S2A, B). All confirmed variants had a coverage of at least 8× reads, and a mutant allele percentage of 33–62% for heterozygous or 83–100% for hemizygous variants (Table 3). The mutation detection rate was

Table 1. Strategy for validation of RainDance sample enrichment and NGS methods.

	PGM	MiSeq
Number of samples	10	10
Sample enrichment	RDT1000*	RDT1000*
Sequence generated	100 bp single-end** (316 chip/318 chip)	150 bp pair-end (Miseq Reagent Kit)
Mapping	TMAP v2.0.1/Novoalign	Novoalign
SNP/indel identification	Variant caller/GATK	GATK
Annotation	ANNOVAR	ANNOVAR

*The sequencing library used was the RainDance ASDSeq™Research Screening Panel.
**PGM provided the protocol for paired-end sequencing in the end of 2011, only for optional.
doi:10.1371/journal.pone.0074167.t001

either 70% (PGM using standard analysis software of TMAP and Variant Caller) or 100% (MiSeq). With PGM, the variant located near the homopolymer could not be detected because of PGM's high frequency of homopolymer sequencing errors [20,21]. When using TSv3.2 for PGM data analysis, one out of four mutations not identified by TSv2.0, were additionally detected. In order to analyze on the same analytical platforms, sequence data of PGM were also processed using Novoalign for mapping and GATK for variant calling. The mutation detection rate differed significantly between platforms (TMAP-Variant Caller and Novoalign-GATK) (Table 3). Respective PGM data, displayed in the Integrative Genomics Viewer (IGV) [22], showed an increase in sequence mismatch patterns at amplicon ends.

Validation of the RainDance ASD panel for detecting novel mutations in ASD patients

RainDance targeted resequencing was obtained on a total of 28 ASD patients, with a mean total sequence length of 273 or 446 Mb, and an average read depth of approximately 65× or 115×, for PGM and MiSeq, respectively (Table 4). After filtering by dbSNP135, a total of 98 (PGM) and 62 (MiSeq) variants were discovered following RainDance target enrichment. Of these, 62 (PGM) and 46 (MiSeq) were nonsynonymous SNVs (Table S3). Under a rare variant hypothesis, variants were filtered to exclude common variants (MAF ≥1%), using the Exome Variant Server from the NHLBI Exome Sequencing Project and an internal dataset of 212 control exomes from the Japanese population. Although c.878C>T (p.S293F) in *SLC6A4* was detected in 4/212 control exomes (MAF = 0.01%), we chose not to remove this SNV, since it has been functionally proven to disrupt serotonin transporter activity [23]. We validated a total of 57 (PGM) and 30 (MiSeq) SNVs. These SNVs were confirmed by Sanger sequencing, with 21 (PGM) and 22 (MiSeq) shown to be true positives (Table S3). In contrast, after filtering to exclude common variants, no indel mutations were detected by either PGM or MiSeq. All 21 SNVs detected by PGM were also detected by MiSeq. We analyzed the ability of each platform to detect variants and found that both platform was able to identify true variants, but PGM produced many false variant calls. The true positive call rates in the entire coding region were 36.8% (PGM) and 73.3% (MiSeq) (Table S3). We inspected each false positive calls in PGM and MiSeq using IGV to evaluate what kind of errors they were. In PGM, 27/36 calls (75%) had low depth, 21/36 calls (58.3%) had calls at respective read end, 14/36 calls (38.8%) were located near homopolymers, and 1/36 calls (2.7%) had PGM specific low quality error. In MiSeq, 5/8 calls (62.5%) had calls at respective read end and 3/8 calls (37.5%) had MiSeq specific errors. (Table S3).

Table 2. Comparison between PGM and MiSeq sequencing performance in 10 positive controls.

	PGM		MiSeq
	TMAP	Novoalign	
Average total number of bases (Mb)	295.97	201.73	469.42
Average read length (base)	116	116	150
% mapped on human genome	96.8%	78.8%	75%
% on target regions	26.7%	28.3%	22.7%
Mean depth of coverage	63	57	95
% of target regions at >10-fold coverage	93.7%	92.1%	96.8%
% of target regions at >20-fold coverage	85.9%	82.0%	93.2%

doi:10.1371/journal.pone.0074167.t002

Candidate rare SNVs associated with ASD

We identified 22 rare SNVs in 28 patients with ASD (Table 5). Clinical features of the patients with these rare SNVs were demonstrated (Table S4). We considered some to be disease causing, as they are the same mutations previously reported in patients with different diseases that accompany autistic features, namely, c.4612G>A (p.V1538I) in *SCN1A*, identified in a patient with Dravet syndrome [24], and c.878C>T (p.S293F) in *SLC6A4*, identified in a patient with serotonin transporter deficiency [23]. The c.7880G>A (p.R2627Q) mutation identified in *CHD7* was not the same mutation, but was found at the same position, as the one detected in a patient with CHARGE syndrome [25]. Of these three patients, parent samples were only available for the patient with the *SLC6A4* mutation, and the mutation was shown to be inherited from a mother with no autistic features.

Eighteen of the identified SNVs were not observed in 212 in-house Japanese control exomes, suggesting they may be strong candidates for ASD associated SNVs. The remaining four SNVs were also observed in control exomes; however, with a lower frequency than patients with ASD, leading to an OR of 1.93–25.32. In particular, c.56C>T (p.A19V) was detected significantly more frequently in patients with ASD than in controls (OR, 25.32; 95% confidence interval (CI), 2.54–252.76). The remaining SNVs did not reach statistical significance, likely due to the limited number of patients analyzed.

Based on web-based prediction software, 72.7% of the detected SNVs (16/22) were deemed pathogenic by either PolyPhen-2

Table 3. Validation of our chosen NGS methods for mutation detection.

					Detected by			Coverage			Mutant allele (%)		
Sample	Sex	Chr	Gene	Mutation	PGM[1)]	PGM[2)]	MiSeq	PGM[1)]	PGM[2)]	MiSeq	PGM[1)]	PGM[2)]	MiSeq
1	F	2	*SCN1A*	c.342_344delinsAGGAGTT	−	−	+	13	n.a.	91	n.a.	n.a.	44
2	F	2	*SCN1A*	c.4313T>A (p.M1438K)	+	+	+	31	42	48	33	31	38
3	M	5	*NSD1*	c.3958C>T (p.R1320X)	+	−	+	34	n.a.	50	62	n.a.	40
4	M	5	*NSD1*	c.5177C>T (p.P1726L)	+	−	+	37	n.a.	93	38	n.a.	46
5	M	5	*NSD1*	c.5179G>C (p.A1725P)	+	−	+	55	n.a.	62	47	n.a.	50
6	M	5	*NSD1*	c.6499T>C (p.C2167R)	+	−	+	77	n.a.	223	46	n.a.	54
7	F	X	*MECP2*	c.243_244insC	−	−	+	18	n.a.	123	n.a.	n.a.	41
8	F	X	*MECP2*	c.316C>T (p.R106W)	+	−	+	60	n.a.	76	42	n.a.	47
9	M	X	*CDKL5*	c.145G>A (p.E49K)	−	−	+	8	n.a.	46	n.a.	n.a.	100
10	M	X	*CASK*	c.227_228del	(+)	+	+	35	47	112	83	81	97

F, Female; M, Male; Chr, Chromosome; +, Detected; −, Not detected; (+), Mutation only detected by TSv3.2, and not by TSv2.0.; n.a., Not applicable;
[1)]Reads were mapped by TMAP and SNVs and indels were identified using the Torrent Variant Caller.
[2)]Reads were mapped by Novoalign v3.00 and SNVs and indels were identified using the GATK v1.5–21.
doi:10.1371/journal.pone.0074167.t003

(36.3%; 8/22 SNVs), SIFT (50%; 11/22 SNVs), or MutationTaster (13.5%; 3/22 SNVs). We annotated positions with their conservation as scored with the GERP. Mutations at highly conserved positions would be predicted to be functionally important (45.5%; 10/22 SNVs).

Five out of 28 patients had multiple SNVs (Table S5). Following the multigenic contribution theory in ASD [4], these could be associated with the onset or the severity of this disease.

Discussion

We have developed an efficient workflow for detecting rare SNVs/indels in ASD associated genes using bench-top next generation sequencers with target gene enrichment. The evaluation and comparison of NGS devices are of recent interest to us. In this study we chose to compare the Ion Torrent PGM and Illumina MiSeq, which are currently the most popular NGS. The characteristics of the two devices are shown (Table S6). In this study, we compared the sequence yield and quality of these two NGS platforms, and showed a practical use for targeted resequencing of human genes.

Our comparison of two bench-top sequencers showed their yields were both greater than expected; however, the quality of sequence reads varied: better than expected through entire reads in MiSeq, while barely exceeding the minimum expected quality value with large discrete reads in PGM. Comparing the mapping quality of the two sequencers, which was calculated based on the same algorithm, the percentage of reads with a mapping quality ≤40 was markedly more in PGM than in MiSeq. Considering their target regions were the same, this difference reflects the difference of overall read quality from the two sequencers. Focusing on indel calls, we found an excess with PGM, compared to MiSeq. The number of MiSeq indel calls is reasonable, compared to the estimated error rate (0.11 to 0.08 per 1 kb) in conventional capillary sequencing of the human genome [26]. Even with filtering of the reads for low genotyping quality and depth, the excess indel calls in PGM did not decrease. As previously reported, excess indel calls or a lower read quality are considered to be largely due to homopolymers [20,27]. This unique characteristics of PGM was reflected in the difference of mapped rates for PGM-generated data when using different mapping tools, TMAP or Novoalign. As shown in Table 2, the mapped rates of bases between PGM-generated data and MiSeq-generated data using Novoalign were similar, being reasonable since these two data were derived from the same sample libraries, while the one for PGM-generated data using TMAP was better. We assume this is because TMAP consider homopolymer-associated indel errors on mapping and could map more reads which standard mapping tools such as Novoalign could not. The difference in the mapped rates for PGM-generated data might affect the mutation detection rate. Based on the difference in mutation detection rates of positive controls in PGM-generated data with different pipelines (Table 3), custom mapping and the SNP/indel detection software, TMAP and Variant Caller, are necessary for the PGM workflow to reduce mapping errors without compromising detection sensitivity. Otherwise the number of false positive indel calls would be greatly increased.

Table 4. Comparison between PGM and MiSeq sequencing performance in 28 ASD patients.

	PGM	MiSeq
Average of total number of bases (Mb)	273.06	445.99
% on target regions	30.20%	25.60%
Mean depth of coverage	65	115
% of target regions at >10-fold coverage	92.70%	95.50%

doi:10.1371/journal.pone.0074167.t004

Generally, target gene enrichment using the RDT machine worked well, but there were some disadvantages, including a relatively low on-target rate as shown in Table 2, and occasional sample enrichment failure. This may be partially due to the genomic complexity or a biased GC content of target regions. Alternatively, it may be due to the screening panel itself, which does not employ a tailed primer system using PCR amplification primers, therefore necessitating the use of the concatenation step for sample preparation.

In our workflow validation using ten positive controls, the mutation detection rate was lower with PGM than MiSeq. False

Table 5. Rare SNVs identified with amino acid changes and computational predictions of pathogenicity.

Gene	Accession No.	Nucleotide : amino acid change	MutationTaster	Polyphen2[1] (Hum Div)	SIFT[2]	GERP[3]	HGMD[4]	genotype (allele) cases	genotype (allele) controls	OR (95% CI)	*p* value	Patient
BRAF	NM_004333	c.976A>G:p.I326V	polymorphism	0	0.71	−5.32	none	1/28 (1/56)	1/212 (1/424)	7.82 (0.46–128.60)	0.22	A682
CACNA1C	NM_001129837	c.4706C>T:p.P1569L	polymorphism	0.001	**0.04**	2.39	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A681
CHD7	NM_017780	c.7880G>A:p.R2627Q	polymorphism	**0.997**	**0.01**	**5.56**	CHARGE syndrome (R2627X)	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A634
CHD7	NM_017780	c.7652C>A:p.T2551N	polymorphism	0.01	0.31	**5.63**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A447
CNTNAP2	NM_014141	c.1276C>A:p.L426I	polymorphism	**0.977**	**0.03**	**5.7**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A479
CNTNAP2	NM_014141	c.1448G>A:p.R483Q	polymorphism	**0.991**	0.4	**5.07**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A621
DMD	NM_004007	c.3479A>G:p.N1160S	polymorphism	**0.973**	0.22	1.36	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
DMD	NM_004007	c.2473A>G:p.M825V	polymorphism	0.026	0.36	3.87	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
MID1	NM_001193278	c.555G>A:p.M185I	polymorphism	**0.839**	**0.01**	**5.64**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A669
NIPBL	NM_015384	c.1553C>T:p.T518I	polymorphism	0.275	**0**	**5.88**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A681
NRXN1	NM_001135659	c.455G>A:p.G152D	**disease causing**	0	1	4.97	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A711
NSD1	NM_022455	c.2087T>C:p.V696A	polymorphism	0.189	**0.02**	3.94	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A464
PNKP	NM_007254	c.56C>T:p.A19V	polymorphism	0.026	0.27	4.55	none	3/28 (3/56)	1/212 (1/424)	**25.32 (2.54–252.76)**	**0.005**	A627, A651, A674
RAI1	NM_030665	c.1148C>T:p.P383L	polymorphism	1	**0**	**5.55**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A663
RAI1	NM_030665	c.4238T>C:p.M1413T	polymorphism	0.011	**0**	2.93	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A634
RELN	NM_005045	c.8915A>C:p.K2972T	**disease causing**	**0.996**	**0.03**	**5.89**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A653
SCN1A	NM_001165963	c.4612G>A:p.V1538I	polymorphism	**0.89**	0.09	**5.76**	Dravet syndrome	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A695
SHANK3	NM_033517	c.3169C>T:p.L1057F	polymorphism	0.232	0.27	3.19	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A668
SLC6A4	NM_001045	c.878C>T:p.S293F	**disease causing**	0	0.25	3	Serotonin transporter deficency	1/28 (1/56)	4/212 (4/424)	1.93 (0.21–17.87)	0.47	A674
TSC2	NM_000548	c.2032G>A:p.A678T	polymorphism	0.016	0.23	−0.706	none	1/28 (1/56)	1/212 (1/424)	7.82 (0.48–128.60)	0.22	A647
VPS13B	NM_015243	c.820T>G:p.F274V	polymorphism	0.314	**0**	**5.45**	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A663
VPS13B	NM_017890	c.11960C>G:p.P3987R	polymorphism	0.437	0.06	3.85	none	1/28 (1/56)	0/212 (0/424)	n.d.	0.12	A619

[1]PolyPhen2 scores close to 1 are likely to be pathogenic (highlighted in bold). HumDiv-trained Polyphen-2 assumes even mildly deleterious alleles as damaging to evaluate rare alleles potentially involved in complex phenotypes.
[2]SIFT scores less than 0.05 are likely to be pathogenic (highlighted in bold).
[3]GERP scores above 5 are highly conserved (highlighted in bold).
[4]The Human Gene Mutation Database (HGMD) was searched to identify SNVs registered as disease causing mutations. Carrier frequencies of each SNV were statistically compared between ASD patients (cases) and in-house normal 212 controls (controls). Results are presented as odds ratio's (OR) and p values. Pathogenic findings are shown in bold. CI, confidence interval; wt, wild type allele; mut, mutant allele; n.d., not determined.
doi:10.1371/journal.pone.0074167.t005

negatives are largely due to the weakness in indel detection, implying not only excess false positive, but also increased false negative indel calls with PGM. Another typical false negative mutation identified with PGM was detected at amplicon ends. This may happen more readily with PGM as the read length is not as long as expected. On the other side, the higher coverage of the MiSeq data is expected due to the longer read lengths as well as paired end reads. With regards to SNV detection, both PGM and MiSeq showed high mutation detection rates (6/7 mutations, 85.7% in PGM vs. 7/7 mutations, 100% in MiSeq). Target resequencing of 28 patients with ASD identified 21 candidate SNVs in PGM versus 22 in MiSeq, again showing similar SNV variant detection abilities. Although there is a higher false-positive SNV call rate with PGM compared to MiSeq due to the same factors observed in positive control studies, At present it would be reasonable to apply PGM for SNV detection. Recent rapid updates of the device, chemistry and mapping/mutation detection software in PGM may potentially reduce these drawbacks in the near future.

ASD is a genetically heterogeneous disease, with a complex genetic architecture [4]. In particular, rare SNVs with a multigenic contribution are expected to play a specific role in the molecular pathogenesis of ASD. We have shown that our workflow works rapidly and inexpensively to address this issue by demonstrating our successful identification of novel candidate SNVs in ASD. Notably, A19V in *PNKP* was identified significantly more in patients with ASD than controls. PNKP (polynucleotide kinase 3′-phosphatase) is a bi-functional enzyme that possesses both DNA 3′-phosphatase and DNA 5′-kinase activities, and associates with the single strand break repair machinery. Single strand break could be hazardous to the cell if left unrepaired, especially in central nervous system since frequently single strand breaks could happen [28]. *PNKP* is mutated in microcephaly, early-onset, intractable seizures and developmental delay (MCSZ), in autosomal recessive manner. Patients with MCSZ sometimes show variable behavioral problems, mainly hyperactivity [29]. Considering enzymatic activity of PNKP and its stability as reported [30], clinical symptoms of individuals with the heterozygous variant may not be as severe as MCSZ, however it could not be denied that possible decrease in enzyme activity or protein level of PNKP comparing to wild type might affect the normal development of central nervous system. It was implied that *PNKP* might be a candidate for ASD-related gene by copy number analysis previously [31]. We showed for the first time a candidate variant associated with ASD. Further study with larger samples is necessary to confirm its pathogenicity. It is also noted that there were some genes such as *CHD7, CNTNAP2, DMD*, and *RAI1*, in which two patients had private rare variants. It is speculated that the private variants of those might accumulate in ASD populations.

In conclusion, we present the comparison of two bench-top sequencers, PGM and MiSeq, through the newly developed workflow for the investigation of ASD. Analyzing larger sample sets may lead to unraveling of the missing heritability of ASD.

Supporting Information

Figure S1 Comparison of overall sequencing quality between PGM and MiSeq. (A) Box plots of base-call quality scores across all bases obtained using PGM with a 316 chip (left panel) or MiSeq (right panel). Green and red areas indicate quality scores above 28 and below 20, respectively. Yellow boxes show upper and lower quartiles with whiskers indicating 10% and 90% quartiles. Red horizontal lines indicate the median value. Blue curves represent the mean quality scores. Quality scores are given based on the calculation of Phred-scaled quality values using q = -10log10(P), with P being the estimated error probability for that base-call. (B) Quality score distribution over all sequence reads obtained using PGM with a 316 chip (left panel in red) or MiSeq (right panel in blue). Combined data from four samples are displayed. Mean quality scores across all base-calls from a particular sequence, calculated as the Phred score, are shown on the X axis, and the number of reads with the specified mean sequence quality on the Y axis. (C) Distribution of read length from all sequence reads obtained using PGM with a 316 chip (left panel in red) or MiSeq (right panel in blue). Read lengths are shown on the X axis, and the number of reads with the specified read lengths on the Y axis. (D) Mapping quality from all sequence reads obtained using PGM with a 316 chip (red bars) or MiSeq (blue bars). The number of reads with a mapping quality of either <40 or ≥ 40 in each device (left panel). The percentage of reads with mapping quality ≥ 40 in each device (right panel). MQ, mapping quality.
(TIF)

Figure S2 Comparison between PGM and MiSeq of mutations and sequence reads from positive control samples. (A) The c.342_344delinsAGGAGTT mutation detected in Sample 1. (B) The c.243_244insC mutation detected in Sample 7. In both panels, data was obtained from either PGM (upper) or MiSeq (lower). Both the c.342_344delinsAGGAGTT mutation and the c.243_244insC mutation were not detected in PGM with neither PGM-TMAP-Variant Caller algorithm nor PGM-Novolign-GATK algorithm. Forward and reverse read strands are shown in pink and blue, respectively. Red and blue arrows indicate insertion and deletion positions, respectively, which were confirmed by Sanger sequencing. The horizontal bar indicates the deletion call, and symbols within the read strands (I·) indicate insertion calls detected by either PGM or MiSeq. In (A) and (B), the true inserted sequence depicted by "I·" commonly detected by PGM and MiSeq is AACTCC and C, respectively. The DNA sequence surrounding a mutation is shown below the IGV graphics. WT, wild type; Pt, patient.
(TIF)

Table S1 RainDance ASDSeq™ Core Research Screening Panel.
(PDF)

Table S2 Summary of SNP/indel detection with PGM and MiSeq.
(PDF)

Table S3 Summary of target resequencing and prioritization.
(PDF)

Table S4 Clinical features of patients with novel SNVs.
(PDF)

Table S5 Multiple mutations detected in patients with ASD.
(PDF)

Table S6 Comparison of PGM and MiSeq analysis cost and expected yield.
(PDF)

Acknowledgments

We thank all the participants for their cooperation in this research. We also thank Ms. Y. Yamashita, Ms. K. Takabe and Mr. T. Miyama from the

Department of Human Genetics, Yokohama City University Graduate School of Medicine, and Dr. D. Yamaguchi from BITS, for their technical assistance.

Author Contributions

Conceived and designed the experiments: EK SM N. Matsumoto. Performed the experiments: EK SM. Analyzed the data: EK SM MN YT N. Miyake HS. Contributed reagents/materials/analysis tools: NO. Wrote the paper: EK SM N. Matsumoto.

References

1. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70: 863–885.
2. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators CDC (2012) Prevalence of autism spectrum disorders-Autism and Developmental Disabilities Monitoring Network. MMWR Surveill Summ 61: 1–19.
3. American Psychiatric Association (2000) Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition-Text Revision (DSMIV-TR); Association AP, editor. Washington D.C.
4. Devlin B, Scherer SW (2012) Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev 22: 229–237.
5. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, et al. (2011) Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 68: 1095–1102.
6. Stein JL, Parikshak NN, Geschwind DH (2013) Rare inherited variation in autism: beginning to see the forest and a few trees. Neuron 77: 209–211.
7. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, et al. (2012) Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism 3: 9.
8. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, et al. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 25: 63–77.
9. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, et al. (2012) Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Hum Mol Genet 21: 4781–4792.
10. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485: 237–241.
11. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485: 246–250.
12. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485: 242–245.
13. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74: 285–299.
14. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, et al. (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77: 235–242.
15. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164.
16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498.
17. Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, et al. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods 7: 250–251.
18. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8: 186–194.
19. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18: 1851–1858.
20. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434–439.
21. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13: 341.
22. Thorvaldsdottir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform.
23. Prasad HC, Zhu CB, McCauley JL, Samuvel DJ, Ramamoorthy S, et al. (2005) Human serotonin transporter variants display altered sensitivity to protein kinase G and p38 mitogen-activated protein kinase. Proc Natl Acad Sci U S A 102: 11545–11550.
24. Depienne C, Trouillard O, Saint-Martin C, Gourfinkel-An I, Bouteiller D, et al. (2009) Spectrum of SCN1A gene mutations associated with Dravet syndrome: analysis of 333 patients. J Med Genet 46: 183–191.
25. Jongmans MC, Admiraal RJ, van der Donk KP, Vissers LE, Baas AF, et al. (2006) CHARGE syndrome: the phenotypic spectrum of mutations in the CHD7 gene. J Med Genet 43: 306–314.
26. Meader S, Hillier LW, Locke D, Ponting CP, Lunter G (2010) Genome assembly quality: assessment and improvement using the neutral indel model. Genome Res 20: 675–684.
27. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012: 251364.
28. Reynolds JJ, Stewart GS (2013) A single strand that links multiple neuropathologies in human disease. Brain 136: 14–27.
29. Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, et al. (2010) Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. Nat Genet 42: 245–249.
30. Reynolds JJ, Walker AK, Gilmore EC, Walsh CA, Caldecott KW (2012) Impact of PNKP mutations associated with microcephaly, seizures and developmental delay on enzyme activity and DNA strand break repair. Nucleic Acids Res 40: 6608–6619.
31. Gregory SG, Connelly JJ, Towers AJ, Johnson J, Biscocho D, et al. (2009) Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med 7: 62.

Mutations in *B3GALT6*, which Encodes a Glycosaminoglycan Linker Region Enzyme, Cause a Spectrum of Skeletal and Connective Tissue Disorders

Masahiro Nakajima,[1,21] Shuji Mizumoto,[2,21] Noriko Miyake,[3,21] Ryo Kogawa,[2] Aritoshi Iida,[1] Hironori Ito,[4] Hiroshi Kitoh,[5] Aya Hirayama,[6] Hiroshi Mitsubuchi,[7] Osamu Miyazaki,[8] Rika Kosaki,[9] Reiko Horikawa,[10] Angeline Lai,[11] Roberto Mendoza-Londono,[12] Lucie Dupuis,[12] David Chitayat,[12] Andrew Howard,[13] Gabriela F. Leal,[14] Denise Cavalcanti,[15] Yoshinori Tsurusaki,[3] Hirotomo Saitsu,[3] Shigehiko Watanabe,[16] Ekkehart Lausch,[17] Sheila Unger,[18] Luisa Bonafé,[19] Hirofumi Ohashi,[16] Andrea Superti-Furga,[19] Naomichi Matsumoto,[3] Kazuyuki Sugahara,[2] Gen Nishimura,[20] and Shiro Ikegawa[1,*]

Proteoglycans (PGs) are a major component of the extracellular matrix in many tissues and function as structural and regulatory molecules. PGs are composed of core proteins and glycosaminoglycan (GAG) side chains. The biosynthesis of GAGs starts with the linker region that consists of four sugar residues and is followed by repeating disaccharide units. By exome sequencing, we found that *B3GALT6* encoding an enzyme involved in the biosynthesis of the GAG linker region is responsible for a severe skeletal dysplasia, spondyloepimetaphyseal dysplasia with joint laxity type 1 (SEMD-JL1). *B3GALT6* loss-of-function mutations were found in individuals with SEMD-JL1 from seven families. In a subsequent candidate gene study based on the phenotypic similarity, we found that *B3GALT6* is also responsible for a connective tissue disease, Ehlers-Danlos syndrome (progeroid form). Recessive loss-of-function mutations in *B3GALT6* result in a spectrum of disorders affecting a broad range of skeletal and connective tissues characterized by lax skin, muscle hypotonia, joint dislocation, and spinal deformity. The pleiotropic phenotypes of the disorders indicate that *B3GALT6* plays a critical role in a wide range of biological processes in various tissues, including skin, bone, cartilage, tendon, and ligament.

Skeletal dysplasias represent a vast collection of genetic disorders of the skeleton, currently divided into 40 groups.[1] Spondyloepimetaphyseal dysplasia (SEMD) is one group (group 13) of skeletal dysplasia that contains more than a dozen distinctive diseases. SEMD with joint laxity (SEMD-JL) is a subgroup of SEMD that consists of type 1 (SEMD-JL1 [MIM 271640]) and type 2 (SEMD-JL2 [MIM 603546]). SEMD-JL1 or SEMD-JL Beighton type is an autosomal-recessive disorder that shows mild craniofacial dysmorphism (prominent eye, blue sclera, long upper lip, small mandible with cleft palate) and spatulate finger with short nail.[2] The large joints of individuals with SEMD-JL1 are variably affected with hip dislocation, elbow contracture secondary to radial head dislocation, and clubfeet. Joint laxity is particularly prominent in the hands. Skeletal changes of SEMD-JL1 are characterized by moderate platyspondyly with anterior projection of the vertebral bodies, hypoplastic ilia, and mild metaphyseal flaring.[3] Kyphoscoliosis progresses with age, leading to a short trunk, whereas platyspondyly become less conspicuous and the vertebral bodies appear squared in shape with age. Recently, dominant kinesin family member 22 (*KIF22* [MIM 603213]) mutations have been found in SEMD-JL2;[4,5] however, the genetic basis of SEMD-JL1 remains unknown.

To identify the SEMD-JL1-causing mutation, we performed whole-exome sequencing experiments. We recruited seven individuals with SEMD-JL1 from five unrelated Japanese families (F1–F5) and a Singapore/Japanese family (F6) (Table 1). One family (F1) had a pair of affected sibs (P1 and P2) from nonconsanguineous parents. Genomic DNA was extracted by standard procedures

[1]Laboratory for Bone and Joint Diseases, Center for Integrative Medical Sciences, RIKEN, Tokyo 108-8639, Japan; [2]Laboratory of Proteoglycan Signaling and Therapeutics, Frontier Research Center for Post-Genomic Science and Technology, Graduate School of Life Science, Hokkaido University, Sapporo 001-0021, Japan; [3]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan; [4]Department of Orthopaedic Surgery, Central Hospital, Aichi Prefectural Colony, Kasugai 480-0392, Japan; [5]Department of Orthopaedic Surgery, Nagoya University School of Medicine, Nagoya 466-8550, Japan; [6]Department of Pediatrics, Akita Prefectural Center on Development and Disability, Akita 010-1407, Japan; [7]Department of Neonatology, Kumamoto University Hospital, Kumamoto 860-8556, Japan; [8]Department of Radiology, National Center for Child Health and Development, Tokyo 157-8535, Japan; [9]Division of Medical Genetics, National Center for Child Health and Development, Tokyo 157-8535, Japan; [10]Division of Endocrinology and Metabolism, National Center for Child Health and Development, Tokyo 157-8535, Japan; [11]Department of Paediatric Medicine, KK Women's and Children's Hospital, Singapore 229899, Singapore; [12]Department of Paediatrics, The Hospital for Sick Children and University of Toronto, Toronto, ON M5G 1X8, Canada; [13]Department of Surgery, The Hospital for Sick Children and University of Toronto, Toronto, ON M5G 1X8, Canada; [14]The Professor Fernando Figueira Integral Medicine Institute (IMIP), Recife, PE 50070-550, Brazil; [15]Skeletal Dysplasia Group, Department of Medical Genetics, Faculty of Medical Sciences, State University of Campinas (UNICAMP), Campinas, SP 13083-970, Brazil; [16]Division of Medical Genetics, Saitama Children's Medical Center, Saitama 339-8551, Japan; [17]Division of Paediatric Genetics, Centre for Pediatrics and Adolescent Medicine, University of Freiburg, Freiburg 79106, Germany; [18]Medical Genetics Service, University of Lausanne, CHUV, Lausanne 1011, Switzerland; [19]Department of Pediatrics, University of Lausanne, CHUV, Lausanne 1011, Switzerland; [20]Department of Pediatric Imaging, Tokyo Metropolitan Children's Medical Center, Fuchu 183-8561, Japan
[21]These authors contributed equally to this work
*Correspondence: sikegawa@ims.u-tokyo.ac.jp
http://dx.doi.org/10.1016/j.ajhg.2013.04.003.

Table 1. Clinical and Radiographic Findings of the Individuals with *B3GALT6* Mutations

Subject ID	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Family ID	F1	F1	F2	F3	F4	F5	F6	F7	F8	F9	F9	F10
Clinical diagnosis	SEMD-JL1	SEMD-JL1	SEMD-JL1	SEMD-JL1	SEMD-JL1	SEMD-JL1	SEMD-JL1	SEMD-JL1	EDS-PF	EDS-PF	EDS-PF	EDS-PF
General Information												
Ethnicity	Japanese	Japanese	Japanese	Japanese	Japanese	Japanese	Japanese/ Singaporean	Vietnamese	Italian	Italian/ Canadian	Italian/ Canadian	Brazilian
Gender	M	M	F	M	F	F	M	M	M	F	F	F
Age	34 years	31 years	12 years, 7 months	6 years	5 years, 1 month	12 years	2 years, 9 months	34 years	8 months	7 years	1 month	5 years, 1 month
Gestational age	39 weeks, 2 days	full term	37 weeks	40 weeks, 1 day	39 weeks, 5 days	full term	39 weeks	full term	ND	36 weeks	37 weeks	39 weeks
Birth length (cm)	ND	ND	36	ND	43.1	42	43	(average)	ND	44	44	44
Birth weight (g)	ND	2,200	2,124	2,832	2,535	2,222	2,485	3,500	ND	2,097	2,790	3,300
Clinical Features												
Height (cm) (SD)[a]	127.7 (−7.4)	130 (−7.0)	88.8 (−10.7)	94 (−4.0)	90 (−4.0)	118.4 (−5.1)	78.2 (−4.0)	118 (−9.1)	66 (−1.6)	90 (−6.8)	45 (−3.7)	81 (−5.9)
Weight (kg) (SD)[a]	40.3 (−2.2)	36.9 (−2.5)	13.2 (−3.7)	15.4 (−1.5)	14.4 (−1.3)	23.2 (−2.0)	10.6 (−1.9)	28 (−3.3)	5.65 (−3.0)	13.9 (−2.2)	2.65 (−2.8)	8.5 (−8.4)
Craniofacial												
Flat face with prominent forehead	ND	ND	+	+	+	+	+	−	+	+	+	+
Prominent eyes, proptosis	ND	ND	+	−	−	+	+	−	+	+	+	+
Blue sclerae	ND	ND	+	+	+	−	+	−	+	+	+	−
Long upper lip	ND	ND	−	+	+	−	+	+	+	+	+	−
Micrognathia	ND	ND	+	+	+	+	−	+	−	−	−	−
Cleft palate	ND	ND	−	−	−	−	−	−	−	−	−	+
Musculoskeletal												
Kyphoscoliosis[b]	+ (7 months)	+ (1.2 years)	+ (8 months)	+ (infancy)	+ (2 years)	+ (3 months)	+ (8 months)	+ (1 year)	+ (6 months)	++ (prenatal)	++ (prenatal)	++ (2 years)
Spatulate finger	−	ND	+	+	+	+	−	−	+	+	+	−
Finger laxity	ND	ND	+	+	−	−	+	−	++	+	+	+
Large joint laxity	ND	ND	+	+	−	−	+	−	++	++	++	+
Restricted elbow movement	+	ND	+	+	+	−	−	+	+	+	+	+
Hand contracture	−	−	−	−	−	+	−	−	−	+	+	−

(*Continued on next page*)

Table 1. ***Continued***

Subject ID	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Hip dislocation	–	–	–	+	–	+	–	–	–	+	+	+
Clubfeet	–	–	+	–	–	–	+	–	–	+	+	–
Muscular hypotonia	–	–	+	–	–	–	–	–	++	++	++	++
Skin and Hair												
Doughy skin	ND	ND	+	–	–	–	+	–	++	+	+	+
Hyperextensibility	ND	ND	+	–	–	–	+	–	++	+	+	–
Cutis laxa	ND	ND	–	–	–	–	–	–	+	+	–	+
Sparse hair	ND	ND	–	–	–	–	–	–	+	+	+	–
Others			MR, DD				camptodactyly			DD		pectus excavatum
Radiological Features												
Platyspondyly	+[c]	+[c]	+[c]	+	+	+	+	+	+	+	+	+
Anterior beak of vertebral body[b]	+	+	– (4 years)	– (5 years)	+	+	+	–	+	+	+	+
Short ilia	+	+	+	+	+	+	+	+	+	+	+	+
Prominent lesser trochanter	+	+	+	–	+	+	+	+	+	+	+	+
Metaphyseal flaring	+	+	+	+	+	+	+	+	+	–	+	+
Epiphyseal dysplasia of femoral head	–	–	–	+	–	+	–	–	–	–	+	+
Elbow malalignment	ND	ND	+	+	+	+	+	+	+	+	+	+
Advanced carpal ossification[b]	– (9 years)	ND	– (12 years)	+	+	+	+	ND	+	– (7 years)	–	– (5 years)
Carpal fusion	ND	ND	+	–	–	–	–	–	–	–	–	–
Metacarpal shortening	ND	ND	+	+	+	+	+	+	–	–	+	–
Overtubulation	–	–	–	–	–	–	–	–	+	+	+	+

Abbreviations are as follows: SEMD-JL1, spondyloepimetaphyseal dysplasia with joint laxity type 1; EDS-PF, Ehlers-Danlos syndrome, progeroid form; ND, no data; MR, mitral regurgitation; DD, developmental delay.
[a]At last presentation.
[b]Age at medical attention provided in parentheses.
[c]Absent at age 20 years in P1 and P2 and at age 12 years in P3.

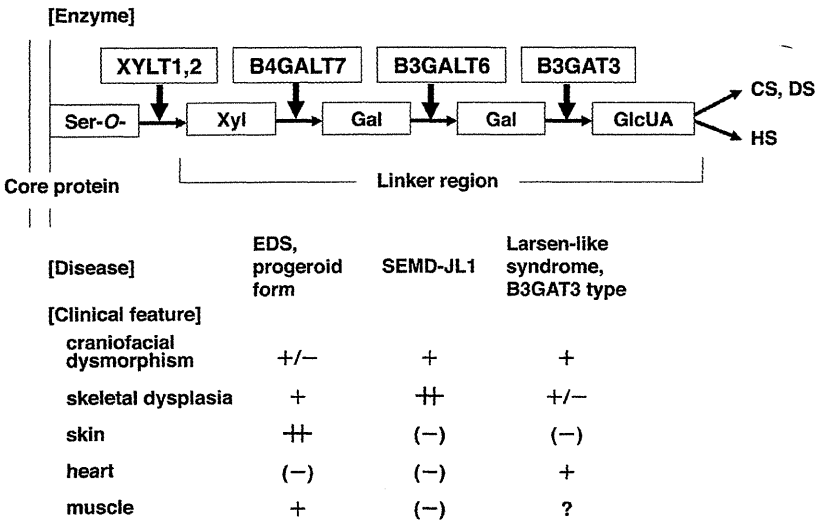


Figure 1. Enzymes Involved in Biosynthesis of the Glycosaminoglycan Linker Region and Summary Features of Diseases Caused by Their Defects Based on a Conventional Concept for the Diseases

The biosyntheses of GAGs start with the formation of a common tetrasaccharide linker sequence covalently attached to the core protein. The linker region synthesis involves a single linear pathway composed of four successive steps catalyzed by distinctive enzymes. Abbreviations are as follows: XYLT, β-xylosyltransferase; B4GALT7, xylosylprotein β1,4-galactosyltransferase, polypeptide 7 (β1,4-galactosyltransferase-I); B3GALT6, UDP-Gal, βGal β1,3-galactosyltransferase polypeptide 6 (β1,3-galactosyltransferase-II); B3GAT3, β-1,3-glucuronyltransferase 3 (glucuronosyltransferase I); Ser-*O*, the serine residue of the GAG attachment site on the proteoglycan core protein; Xyl, xylose; Gal, galactose; GlcUA, D-glucuronic acid; CS, chondroitin sulfate; DS, dermatan sulfate; HS, heparan sulfate; EDS, Ehlers-Danlos syndrome; SEMD-JL1, spondyloepimetaphyseal dysplasia with joint laxity type 1.

from peripheral blood, saliva, or Epstein-Barr virus-immortalized lymphocyte of the individuals with SEMD-JL1 and/or their parents after informed consent. The study was approved by the ethical committee of RIKEN and participating institutions. We captured the exomes of the seven subjects as previously described.[6,7] In brief, we sheared genomic DNA (3 μg) by Covaris S2 system (Covaris) and processed with a SureSelect All Exon V4 kit (Agilent Technologies). We sequenced DNAs captured by the kit with HiSeq 2000 (Illumina) with 101 base pair-end reads. We performed the image analysis and base calling by HiSeq Control Software/Real Time Analysis and CASAVA1.8.2 (Illumina) and mapped the sequences to human genome hg19 by Novoalign. We processed the aligned reads by Picard to remove PCR duplicate. The mean depth of coverage for reads was 132.8×, and, on average, 91.0% of targeted bases had sufficient coverage (20× coverage) and quality for variant calling (Table S1 available online). The variants were called by Genome Analysis Toolkit 1.5-21 (GATK) with the best practice variant detection with the GAKT v.3 and annotated by ANNOVAR (2012 February 23).

Based on the hypothesis that SEMD-JL1 is inherited in an autosomal-recessive fashion, we filtered variants with the script created by BITS (Tokyo, Japan) according to following conditions: (1) variants registered in ESP5400, (2) variants found in our in-house controls (n = 274), (3) synonymous changes, (4) rare variants registered in dbSNP build 135 (MAF < 0.01), and (5) variants associated with segmental duplication. After combining variants selected by the homozygous mutation model and the compound heterozygous mutation model, we selected genes shared by individuals from three or more families. The analysis of the next-generation sequencing identified possible compound heterozygous variants in *B3GALT6* in individuals from three families (Table S2). In addition, two other subjects had possible causal heterozygous variants of *B3GALT6*.

B3GALT6 (RefSeq accession number NM_080605.3) is a single-exon gene on chromosome 1p36.33. It encodes UDP-Gal:βGal β1,3-galactosyltransferase polypeptide 6 (or galactosyltransferase-II: GalT-II), an enzyme involved in the biosynthesis of the glycosaminoglycan (GAG) linker region.[8] The biosyntheses of dermatan sulfate (DS), chondroitin sulfate (CS), and heparin/heparan sulfate (HS) GAGs start with the formation of a tetrasaccharide linker sequence, glucuronic acid-β1-3-galactose-β1-3-galactose-β1-4-xylose-β1 (GlcUA-Gal-Gal-Xyl), which is covalently attached to the core protein. The linker region synthesis involves a single linear pathway composed of four successive steps catalyzed by distinctive enzymes (Figure 1). The first step is the addition of xylose to the hydroxy group of specific serine residues on the core protein by xylosyltransferases from UDP-Xyl, followed by two distinct galactosyltransferases (GalT-I and II) and a glucuronosyltransferase from UDP-Gal and UDP-GlcUA, respectively. The next hexosamine addition is critical because it determines which GAG (i.e., CS, DS, or HS) is assembled on the linker region. GalT-II encoded by *B3GALT6* functions in the third step of the linker formation (Figure 1).

To confirm the results obtained by the next-generation sequencing, we examined the seven subjects used for the next-generation sequencing and an additional subject from a Vietnamese family (F7) by direct sequence of the PCR products from genomic DNAs using 3730xl DNA Analyzer (Applied Biosystems). The Sanger sequencing confirmed all *B3GALT6* mutations found by the next-generation sequencing and identified additional *B3GALT6* mutations. The results indicated that *B3GALT6* mutations were found in all subjects (Tables 2 and S1). All but P4 from F3 were compound heterozygotes of missense mutations. In P4, only a heterozygous c.1A>G (p.Met1?) mutation was found, although we searched for a *B3GALT6* mutation in the entire coding region, 5′ and 3′ UTRs, and flanking

Table 2. *B3GALT6* Mutations in Spondyloepimetaphyseal Dysplasia with Joint Laxity Type 1 and Ehlers-Danlos Syndrome, Progeroid Form

Family	Clinical Diagnosis	Nucleotide Change	Amino Acid Change
F1	SEMD-JL1	c.1A>G	p.Met1?
		c.694C>T	p.Arg232Cys
F2	SEMD-JL1	c.1A>G	p.Met1?
		c.466G>A	p.Asp156Asn
F3[a]	SEMD-JL1	c.1A>G	p.Met1?
F4	SEMD-JL1	c.1A>G	p.Met1?
		c.694C>T	p.Arg232Cys
F5	SEMD-JL1	c.694C>T	p.Arg232Cys
		c.899G>C	p.Cys300Ser
F6	SEMD-JL1	c.1A>G	p.Met1?
		c.193A>G	p.Ser65Gly
F7	SEMD-JL1	c.200C>T	p.Pro67Leu
		c.694C>T	p.Arg232Cys
F8	EDS-PF	c.353delA	p.Asp118Alafs*160
		c.925T>A	p.Ser309Thr
F9	EDS-PF	c.588delG	p.Arg197Alafs*81
		c.925T>A	p.Ser309Thr
F10	EDS-PF	c.16C>T	p.Arg6Trp
		c.415_423del	p.Met139Ala141del

The nucleotide changes are shown with respect to *B3GALT6* mRNA sequence. The corresponding predicted amino acid changes are numbered from the initiating methionine residue.
[a]Only a heterozygous mutation was found.

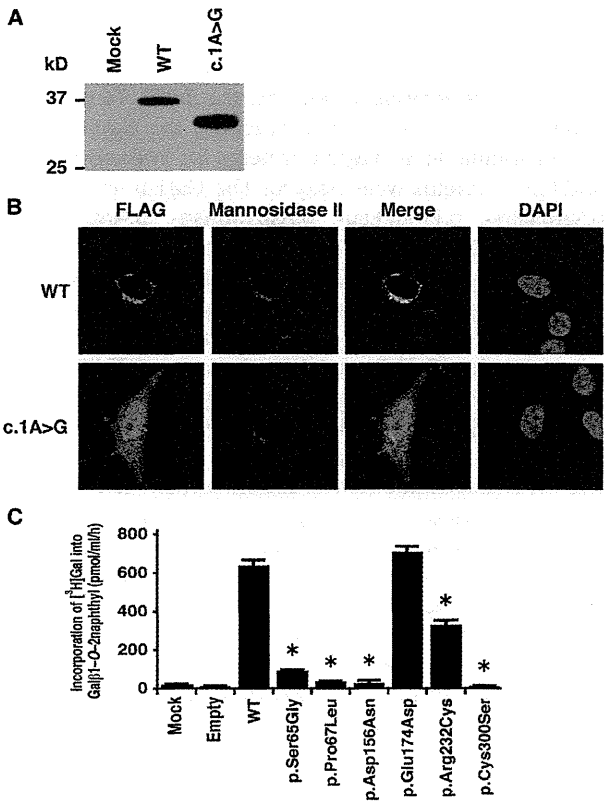


Figure 2. Analyses of B3GALT6 Missense Mutant Proteins Identified in Individuals with SEMD-JL1 In Vitro
(A) Immunoblot analysis of lysates from HeLa cells expressing transfected wild-type (WT) and mutant (c.1A>G) *B3GALT6*. The mutant *B3GALT6* yields a shortened protein. The difference of the molecular sizes between WT and mutant proteins is approximately ~4 kD.
(B) Subcellular localization of B3GALT6. HeLa cells were transfected with WT and mutant (c.1A>G) *B3GALT6*. Cells were stained with anti-FLAG (green), anti-α-mannosidase II (red), and 4′,6-diamidino-2-phenylindole (DAPI; blue). WT was expressed in the Golgi, but the mutant was found in cytoplasm and nucleus.
(C) Decreased enzyme activities of the missense mutant proteins (p.Ser65Gly, p.Pro67Lys, p.Asp156Asn, p.Arg232Cys, and p.Cys300Ser). p.Glu174Asp is a common polymorphism in the public database. The GalT-II activity is measured by incorporation of [^{3}H]Gal into Galβ1-*O*-2naphthyl (*p*mol/ml/hr) and represents the averages of three independent experiments performed in triplicate. Empty and mock indicate the GalT-II activity obtained with the conditioned medium transfected with or without an empty vector. *$p < 0.0001$ versus WT (one-way analysis of variance with Dunnett's adjustment).

regions of *B3GALT6*. Most of the mutations are predicted to be disease causing by in silico analysis. The c.1A>G (p.Met1?) mutation was found in individuals from five of the seven families.

Although mutations affecting initiation codons have been reported to be pathogenic in several diseases,[9] the effects of initiation codon mutations on the encoded protein are variable among the genes. We therefore investigated the effect of the c.1A>G (p.Met1?) mutation on the protein by using C-terminally FLAG-tagged B3GALT6 with and without the mutation expressed in HeLa cells (RIKEN Cell Bank). We detected the mutant B3GALT6 protein with a molecular weight ~4 kD lower compared with the wild-type (WT) protein (Figure 2A). These results suggest that translation initiation at the second ATG of the coding sequence, at position c.124, would become the initiation codon because of the mutation, probably resulting in an N-terminal deletion of 41 amino acids (p.Met1_Ala41del), in the same open reading frame that contains the transmembrane domain. We then examined the subcellular localization of the mutant B3GALT6 protein by immunocytochemistry. The immunofluorescence for WT-B3GALT6 was observed in a perinuclear region overlapping with that for α-mannosidase II, a marker of the Golgi as previously reported.[8] In contrast, the immunofluorescence for the mutant B3GALT6 protein was observed in the nucleus and cytoplasm (Figure 2B). Therefore, the mutant protein can be considered to be functionally null because of the mislocalization.

To investigate the causality of other *B3GALT6* missense mutations, we also examined the subcellular localization of the mutant B3GALT6 proteins by immunocytochemistry. c.193A>G (p.Ser65Gly), c.200C>T (p.Pro67Leu),