Table 3 Ordered logistic regression analyses for predicting liver fibrosis stages in the development cohort

Variable	Coefficient (95% confidence interval)	Standard error	Wald	*P*-value
Univariate analysis				
Log$_e$ (platelet count ($\times 10^9$/l))	−2.859 (−3.858 to −1.860)	0.510	31.461	<0.001
Log$_e$ (age (days))	1.812 (1.119–2.506)	0.354	26.213	<0.001
Log$_e$ (TB (mg/dl))	1.517 (0.891–2.142)	0.319	22.565	<0.001
Log$_e$ (albumin (g/dl))	−7.950 (−11.270 to −4.631)	1.694	22.038	<0.001
Log$_e$ (PT-INR)	7.126 (4.125–10.127)	1.531	21.662	<0.001
Log$_e$ (ChE (IU/l))	−2.841 (−4.078 to −1.604)	0.631	20.272	<0.001
Log$_e$ (DB (mg/dl))	1.269 (0.706–1.832)	0.287	19.534	<0.001
Log$_e$ (GGT (IU/l))	−0.926 (−1.398 to −0.454)	0.241	14.772	<0.001
Log$_e$ (AST (IU/l))	0.924 (0.235–1.612)	0.351	6.920	0.009
Log$_e$ (ALT (IU/l))	0.278 (−0.312–0.868)	0.301	0.852	0.36
Multivariate analysis				
Log$_e$ (TB (mg/dl))	1.185 (0.574–1.796)	0.312	14.452	<0.001
Log$_e$ (platelet count ($\times 10^9$/l))	−1.882 (−3.052 to −0.712)	0.597	9.935	0.002
Log$_e$ (age (days))	1.093 (0.232–1.955)	0.439	6.190	0.01

ALT, alanine aminotransferase; AST, aspartate aminotransferase; ChE, cholinesterase; DB, direct bilirubin; GGT, γ-glutamyltransferase; PT-INR, prothrombin time-international normalized ratio; TB, total bilirubin.

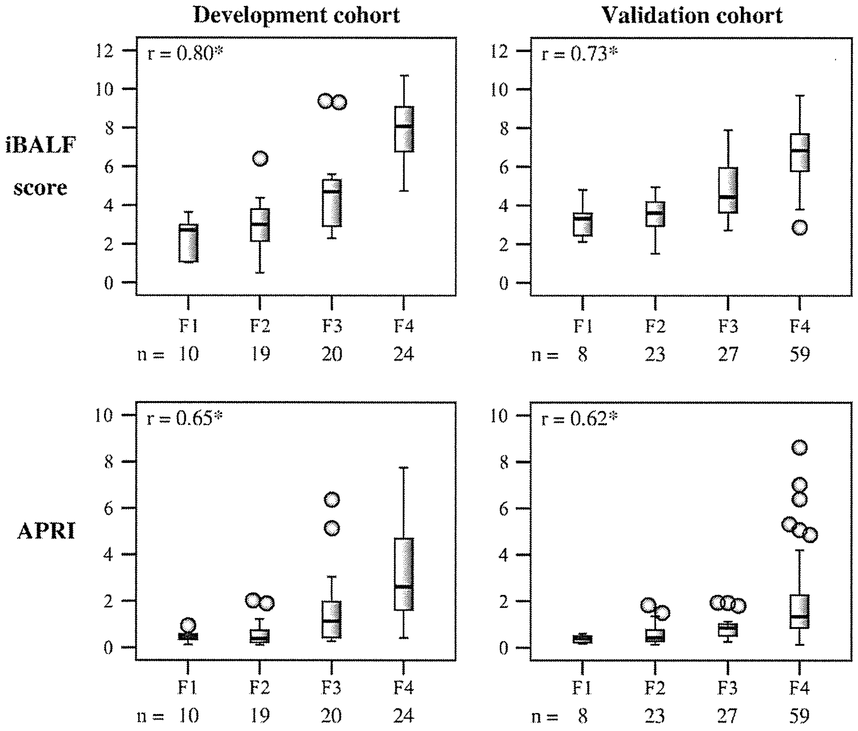

Figure 2 Values of the infant biliary atresia liver fibrosis (iBALF) score and aspartate aminotransferase-to-platelet ratio index (APRI) according to the histological fibrosis stages. Boxplots show the median values with the interquartile ranges, and error bars indicate the smallest and the largest values within 1.5 box-lengths of the upper and the lower quartiles. Circles represent the individual points for outliers. Correlations between the markers and the fibrosis stages were evaluated using the Spearman correlation coefficient (*r*); *$P<0.001$.

For infants with BA at presentation, two types of surgical procedure could be chosen—bile drainage surgery or liver transplantation. There were two reports regarding effects on outcomes after liver transplantation comparing early failure of hepatoportoenterostomy, which was defined as the need for liver transplantation within the first year of life, and primary liver transplantation. Alexopoulos *et al.*[10] described that early failure of hepatoportoenterostomy adversely affected patient and graft survival rates. Neto *et al.*[11] reported that early failure of hepatoportoenterostomy had no effect on patient and graft survival, that late failure of hepatoportoenterostomy had a protective effect compared with primary liver transplantation, and that previous hepatoportoenterostomy increased biliary complications and bowel perforations after liver transplantation. Thus, it is important to know which patients can benefit from bile drainage surgery at presentation. In this study, we attempted to reveal the association between the iBALF score at the initial surgery and prognosis using the BALF score at

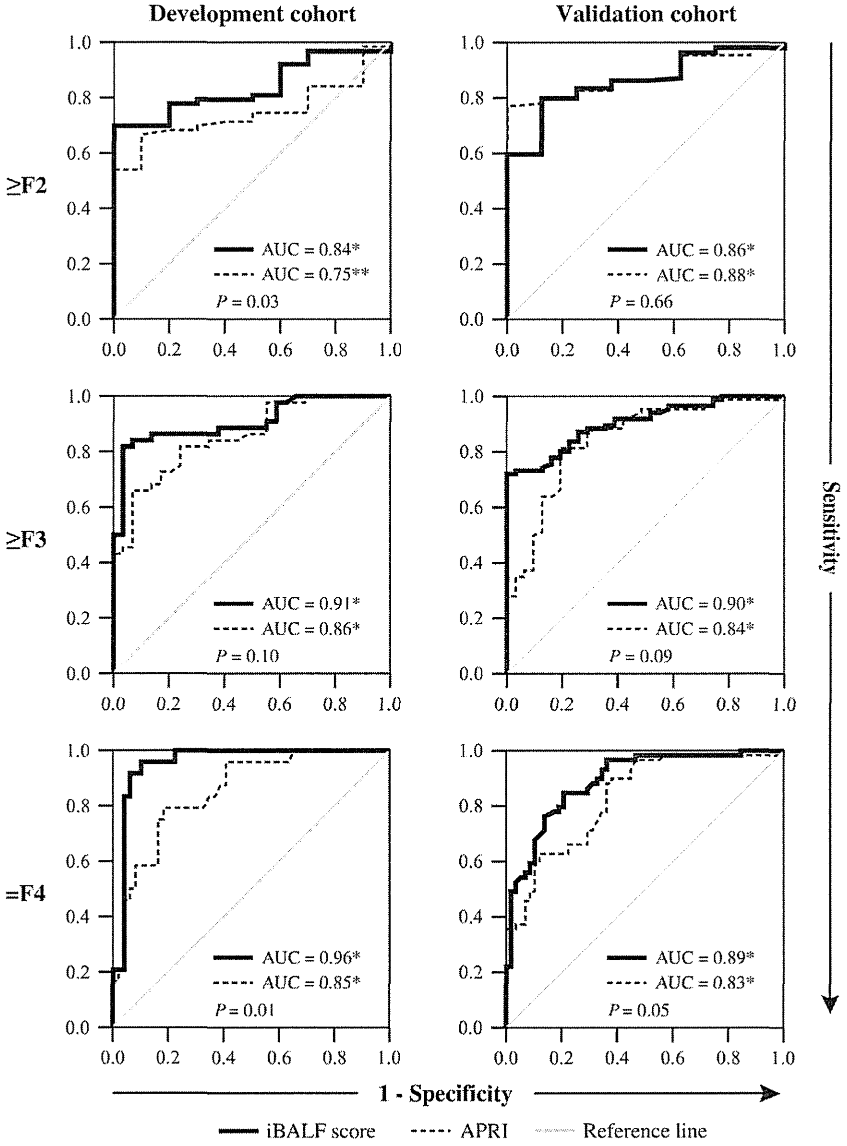

Figure 3 Receiver-operating characteristic curves of two fibrosis markers for diagnosing each fibrosis stage. Evaluated noninvasive markers included the infant biliary atresia liver fibrosis (iBALF) score (thick lines) and the aspartate aminotransferase-to-platelet ratio index (APRI, dashed lines). Gray lines indicate the reference lines. The diagnostic power of each marker was assessed by calculating the area under the curve (AUC); $^*P<0.001$, $^{**}P=0.01$. The P values in the panels represent the differences between AUCs of the iBALF score and the APRI using the DeLong test.

1 year of age. The results (Figure 4) suggest that BA patients with an iBALF score >6 at presentation might require liver transplantation rather than bile drainage surgery. However, the number of these severely affected patients was small in both cohorts. Except for these severely affected patients, the iBALF score at the initial surgery did not seem to be associated with native liver survival at 1 year of age. There was no correlation between the iBALF score at the initial surgery and the BALF score at 1 year of age among the patients with native liver survival, suggesting that liver fibrosis at the initial surgery had a limited effect on liver fibrosis progression or remission. We previously reported similar data on the actual fibrosis stages in 15 patients aged ≥ 2 years who underwent serial histological examinations at the time of initial surgery and after surgery and who were included in the development cohort of the current study: seven of these 15 patients showed remission of fibrosis, five showed the same fibrosis stage, and three showed progression of fibrosis.[4] We believe that effective postsurgical antifibrotic therapy for BA patients is needed and that noninvasive fibrosis monitoring would be highly valuable in clinical practice and study.

In addition to our previous report, several other studies have proposed noninvasive markers to assess liver fibrosis in BA patients. The APRI, which was originally developed to predict cirrhosis in hepatitis C patients,[8] has been widely investigated in BA patients. Kim *et al.*[12] described that the correlation coefficient between the APRI and Metavir fibrosis score from 35 patients at the time of hepatoportoenterostomy was

0.77 ($P < 0.001$) and that the AUCs of the APRI for ≥ F3 and F4 fibrosis were 0.92 and 0.91, respectively. By contrast, Lind *et al.*[13] reported that the APRI was not significantly different according to the fibrosis stage in 31 patients at the time of hepatoportoenterostomy. In 23 patients after successful hepatoportoenterostomy (median, 4.2 years; range, 1.6–18.9 years after surgery), Lampela *et al.*[14] described a significant correlation between the APRI and Metavir fibrosis score ($r = 0.63$, $P < 0.001$) and a good diagnostic accuracy of the APRI for ≥ F3 with 93% sensitivity and 67% specificity. Another noninvasive fibrosis marker, transient elastography (Fibroscan), was more recently investigated to assess liver

Table 4 Cutoff values and diagnostic accuracies of the infant biliary atresia liver fibrosis (iBALF) score for predicting histological fibrosis stages

	n (%)	Cutoff	Sensitivity	Specificity	Accuracy
Development cohort (n = 73)					
≥ F2	63 (86.3%)	3.00	77.8%	80.0%	78.1%
≥ F3	44 (60.3%)	3.99	86.4%	86.2%	86.3%
= F4	24 (32.9%)	5.75	91.7%	93.9%	93.2%
Validation cohort (n = 117)					
≥ F2	109 (93.2%)	3.56	83.5%	75.0%	82.9%
≥ F3	86 (73.5%)	4.34	80.2%	80.6%	80.3%
= F4	59 (50.4%)	5.12	84.7%	79.3%	82.0%

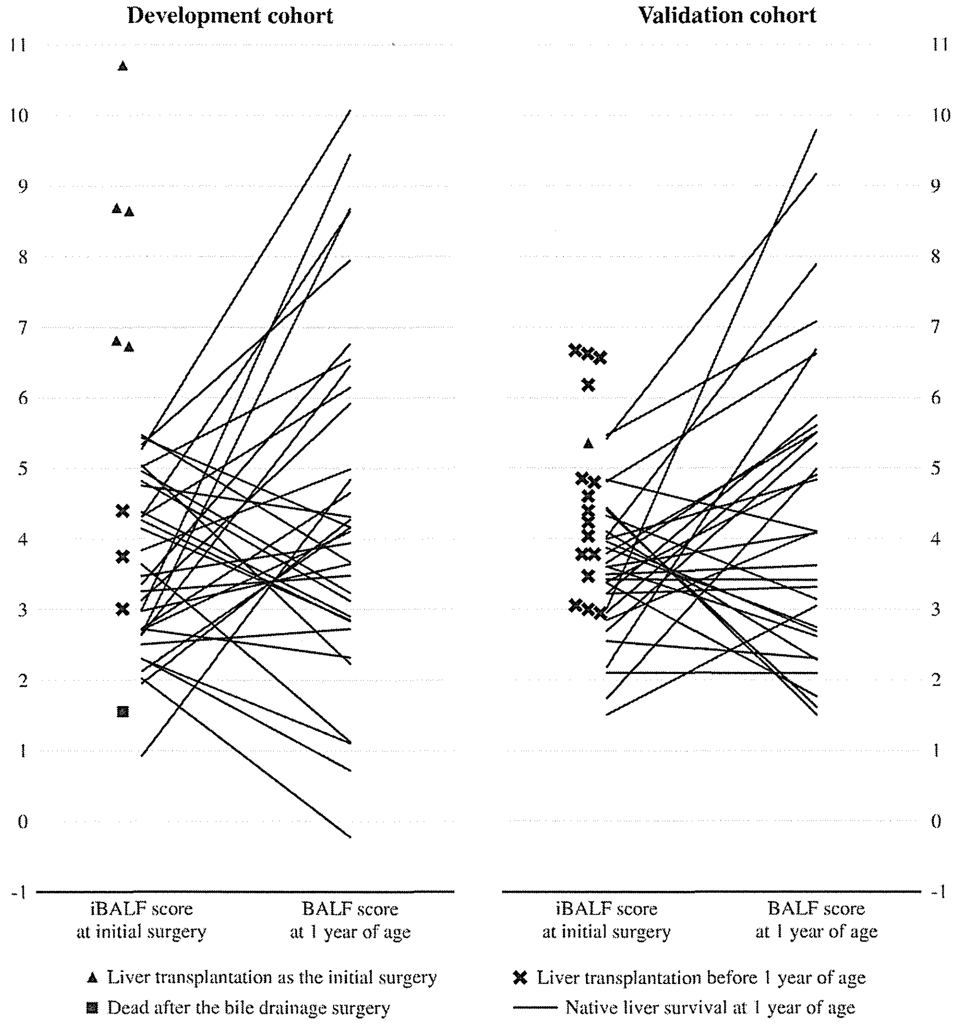

Figure 4 Relationships between the infant biliary atresia liver fibrosis (iBALF) score at the initial surgery and prognosis. Triangles indicate the patients receiving liver transplantation as the initial surgery. Crosses represent the patients requiring liver transplantation after bile drainage surgery before 1 year of age. The square indicates the patient who died after bile drainage surgery. The patients who survived with their native liver at 1 year of age are expressed by lines between the iBALF score at the bile drainage surgery and the biliary atresia liver fibrosis (BALF) score at 1 year of age.

stiffness using the ultrasound technique; Shin *et al.*[15] described that liver stiffness measurements obtained via transient elastography significantly correlated with Metavir fibrosis stages ($r = 0.63$, $P < 0.001$) and had good diagnostic powers for predicting severe fibrosis (≥F3; AUC = 0.86) and cirrhosis (F4; AUC = 0.96) in 47 BA patients aged <1 year at the time of hepatoportoenterostomy with liver biopsy or liver transplantation. Moreover, the APRI and transient elastography had already been investigated for associations with esophageal varices, an important consequence of liver fibrosis and portal hypertension, in postsurgical BA patients.[14,16–18] The current study suggests the advantages of the iBALF score over the APRI: stronger correlation with the fibrosis stages and more favorable diagnostic power than the APRI. Unlike the elastography methods, the iBALF score has good accessibilities, such as no need for a special device and simple equation components that allow retrospective calculation.

Although the current study indicated that the iBALF was a good noninvasive fibrosis marker even in the validation cohort, it has several limitations. First, patients were selected from three institutions, two of which were assigned to the development cohort and one to the validation cohort, resulting in significant differences in patient characteristics and blood test results between the cohorts. BA patients aged <1 year can be divided into three situations: patients before surgery, patients with a good postsurgical course, and patients requiring liver transplantation after bile drainage surgery. Although we intended that the iBALF-scoring system could apply in all situations, needle biopsy examinations for postsurgical patients with good bile drainage were performed at only one of the three participating institutions, thus the sample size was too small. To reflect the data from patients with a good postsurgical course in the iBALF score composition, we assigned the small number of these patients to the development cohort rather than randomly assigning them to the development cohort or the validation cohort. Thus, the relationships between liver fibrosis stage and the iBALF score of patients with a good postsurgical course could not be validated. In addition, there was a probable difference in the timing of liver transplantation between the institutions. Because of serious deceased donor organ shortages in Japan,[19] the timing of liver transplantation using liver allografts from living donors probably reflected the transplantation policy of each institution, resulting in significantly different ranges of the iBALF score in F4 patients between the cohorts and wide overlap in the ranges of the F3 and F4 groups in the validation cohort. The second limitation was general problems in prior studies of noninvasive fibrosis markers using the biopsy examinations as a reference standard: namely, biopsy sampling errors,[20] and observer variability.[21] Subcapsular wedge biopsy examination, which was used in most subjects in the current study, would tend to overestimate liver fibrosis. Thus, the fibrosis stages evaluated based on liver biopsy examinations might have false-positive and false-negative results.

In this study, we developed the iBALF score as a noninvasive surrogate fibrosis marker for BA patients aged <1 year, in addition to the previously developed BALF-scoring system for BA patients aged ≥1 year. Although some concerns remain, the iBALF score was validated to strongly correlate with liver fibrosis stage and to have good diagnostic powers for predicting liver fibrosis. The iBALF and BALF scores may be useful in future clinical studies as surrogate fibrosis markers.

CONFLICT OF INTEREST

Guarantor of the article: Tatsuo Kuroda, MD, PhD.
Specific author contributions: Hirofumi Tomita designed the study, collected and interpreted the data, performed the statistical analysis, and drafted the manuscript; Yasushi Fuchimoto designed the study, collected the data, and critically reviewed the manuscript. A. Fujino and T. Kuroda designed the study, interpreted the data, and critically reviewed the manuscript. K. Hoshino, M. Sakamoto, M. Kasahara, Y. Kanamori, and M. Nakano designed the study and critically reviewed the manuscript. Y. Masugi, A. Nakazawa, and S. Akatsuka participated in the histological evaluations and critically reviewed the manuscript. Y. Yamada and F. Yoshida collected the data and critically reviewed the manuscript. All authors have seen and approved the final version of the manuscript.
Financial support: This study was supported by a grant from The Ministry of Health, Labour and Welfare of Japan (H26-Nanchi-Ippan-082, Health Labour Sciences Research Grants for Research on intractable diseases).
Potential competing interests: None.

Study Highlights

WHAT IS CURRENT KNOWLEDGE

- ✓ Although liver fibrosis is a prominent feature of biliary atresia (BA) patients, noninvasive liver fibrosis markers in BA patients have been limited.
- ✓ We previously developed a BA liver fibrosis (BALF) score as the first specific liver fibrosis marker for BA patients aged ≥1 year.

WHAT IS NEW HERE

- ✓ We developed a novel noninvasive fibrosis marker for BA patients aged <1 year—the infant BALF (iBALF) score.
- ✓ The iBALF score was validated to be a good noninvasive marker of native liver fibrosis for BA patients during infancy.
- ✓ The iBALF and BALF scores can monitor liver fibrosis in a similar manner before and after 1 year of age, respectively.
- ✓ The BA patients with an iBALF score >6 at presentation had poor outcome on native liver survival at 1 year of age.

1. Hartley JL, Davenport M, Kelly DA. Biliary atresia. *Lancet* 2009; **374**: 1704–1713.
2. Haafiz AB. Liver fibrosis in biliary atresia. *Expert Rev Gastroenterol Hepatol* 2010; **4**: 335–343.
3. Lykavieris P, Chardot C, Sokhn M *et al.* Outcome in adulthood of biliary atresia: a study of 63 patients who survived for over 20 years with their native liver. *Hepatology* 2005; **41**: 366–371.
4. Tomita H, Masugi Y, Hoshino K *et al.* Long-term native liver fibrosis in biliary atresia: development of a novel scoring system using histology and standard liver tests. *J Hepatol* 2014; **60**: 1242–1248.
5. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996; **24**: 289–293.
6. Ichida F, Tsuji T, Omata M *et al.* New Inuyama classification; new criteria for histological assessment of chronic hepatitis. *Int Hepatol Commun* 1996; **6**: 112–119.
7. Nio M, Ohi R. Biliary atresia. *Semin Pediatr Surg* 2000; **9**: 177–186.

8. Wai CT, Greenson JK, Fontana RJ *et al.* A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003; **38**: 518–526.
9. Tanaka T, Yamashita A, Ichihara K. Reference intervals of clinical tests in children determined by a latent reference value extraction method. *J Jpn Pediatr Soc* 2008; **112**: 1117–1132.
10. Alexopoulos SP, Merrill M, Kin C *et al.* The impact of hepatic portoenterostomy on liver transplantation for the treatment of biliary atresia: early failure adversely affects outcome. *Pediatr Transplant* 2012; **16**: 373–378.
11. Neto JS, Feier FH, Bierrenbach AL *et al.* Impact of Kasai portoenterostomy on liver transplantation outcomes: A retrospective cohort study of 347 children with biliary atresia. *Liver Transpl* 2015; **21**: 922–927.
12. Kim SY, Seok JY, Han SJ *et al.* Assessment of liver fibrosis and cirrhosis by aspartate aminotransferase-to-platelet ratio index in children with biliary atresia. *J Pediatr Gastroenterol Nutr* 2010; **51**: 198–202.
13. Lind RC, Verkade HJ, Porte RJ *et al.* Aspartate transaminase-to-platelet ratio index is not correlated with severity of fibrosis or survival in children with biliary atresia. *J Pediatr Gastroenterol Nutr* 2012; **54**: 698.
14. Lampela H, Kosola S, Heikkila P *et al.* Native liver histology after successful portoenterostomy in biliary atresia. *J Clin Gastroenterol* 2014; **48**: 721–728.
15. Shin NY, Kim MJ, Lee MJ *et al.* Transient elastography and sonography for prediction of liver fibrosis in infants with biliary atresia. *J Ultrasound Med* 2014; **33**: 853–864.
16. Chang HK, Park YJ, Koh H *et al.* Hepatic fibrosis scan for liver stiffness score measurement: a useful preendoscopic screening test for the detection of varices in postoperative patients with biliary atresia. *J Pediatr Gastroenterol Nutr* 2009; **49**: 323–328.
17. Chongsrisawat V, Vejapipat P, Siripon N *et al.* Transient elastography for predicting esophageal/gastric varices in children with biliary atresia. *BMC Gastroenterol* 2011; **11**: 41.
18. Colecchia A, Di Biase AR, Scaioli E *et al.* Non-invasive methods can predict oesophageal varices in patients with biliary atresia after a Kasai procedure. *Dig Liver Dis* 2011; **43**: 659–663.
19. Tanabe M, Kawachi S, Obara H *et al.* Current progress in ABO-incompatible liver transplantation. *Eur J Clin Invest* 2010; **40**: 943–949.
20. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003; **38**: 1449–1457.
21. Bedossa P. The French METAVIR Cooperative Study Group. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. *Hepatology* 1994; **20**: 15–20.

Molecular Genetic Dissection and Neonatal/Infantile Intrahepatic Cholestasis Using Targeted Next-Generation Sequencing

Q8 Takao Togawa, MD[1], Tokio Sugiura, MD, PhD[1], Koichi Ito, MD, PhD[1], Takeshi Endo, MD, PhD[1], Kohei Aoyama, MD[1], Kei Ohashi, MD[1], Yutaka Negishi, MD[1], Toyoichiro Kudo, MD, PhD[2], Reiko Ito, MD, PhD[2], Atsuo Kikuchi, MD, PhD[3], Natsuko Arai-Ichinoi, MD, PhD[3], Shigeo Kure, MD, PhD[3], and Shinji Saitoh, MD, PhD[1]

Objectives To ascertain a molecular genetic diagnosis for subjects with neonatal/infantile intrahepatic cholestasis (NIIC) by the use of next-generation sequencing (NGS) and to perform a genotype-phenotype correlation.

Study design We recruited Japanese subjects with NIIC who had no definitive molecular genetic diagnosis. We developed a diagnostic custom panel of 18 genes, and the amplicon library was sequenced via NGS. We then compared clinical data between the molecular genetically confirmed subjects with NIIC.

Results We analyzed 109 patients with NIIC ("genetic cholestasis," 31 subjects; "unknown with complications" such as prematurity, 46 subjects; "unknown without complications," 32 subjects), and a molecular genetic diagnosis was made for 28 subjects (26%). The rate of positive molecular genetic diagnosis in each category was 22 of 31 (71%) for the "genetic cholestasis" group, 2 of 46 (4.3%) for the "unknown with complications" group, and 4 of 32 (12.5%) for the "unknown without complications" group. The grouping of the molecular diagnoses in the group with genetic cholestasis was as follows: 12 with Alagille syndrome, 5 with neonatal Dubin-Johnson syndrome, 5 with neonatal intrahepatic cholestasis caused by citrin deficiency, and 6 with progressive familial intrahepatic cholestasis or benign recurrent intrahepatic cholestasis with low gamma-glutamyl transpeptidase levels. Several clinical datasets, including age of onset, direct bilirubin, and aminotransferases, were significantly different between the disorders confirmed using molecular genetic diagnosis.

Conclusion Targeted NGS can be used for molecular genetic diagnosis in subjects with NIIC. Clinical diagnosis should be accordingly redefined in the view of molecular genetic findings. *(J Pediatr 2016;■:■-■).*

The etiologic diversity of neonatal/infantile cholestasis has been described previously.[1,2] The most commonly identifiable etiologies are biliary atresia (25%-35%), genetic intrahepatic cholestasis (25%), and metabolic diseases (20%).[2,3] Recent advances in the understanding of the molecular basis of cholestatic syndromes have enabled the classification of these syndromes and have offered an opportunity for the development of diagnostic methods that take into account the genetic makeup of neonatal/infantile intrahepatic cholestasis (NIIC).[1-3] Alagille syndrome (ALGS), progressive familial intrahepatic cholestasis (PFIC), neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD), and other conditions were distinguished from idiopathic neonatal hepatitis between the early 1970s and the 2000s.[1,2] In addition, in the past 2 decades, a wide range of disease-causing genes underlying the pathogenesis of NIIC has been identified.[1,4] The high genetic heterogeneity and variability of NIIC make it difficult to survey for pathogenic variants in clinical practice, and molecular genetic diagnosis challenging.

Next-generation sequencing (NGS) technologies have revolutionized genomic and clinical genetic research.[5,6] NGS offers comprehensive sequencing of multiple known causative or associated genes in highly heterogeneous diseases.[5] Here, we constructed an NIIC gene panel that included 18 candidate genes. In this study we aimed to use this panel and NGS to ascertain the molecular genetic diagnosis of subjects with NIIC. We then compared clinical and laboratory findings for patients in whom a molecular genetic diagnosis was made to obtain a comprehensive understanding of NIIC.

ALGS	Alagille syndrome
ALT	Alanine aminotransferase
AR	Autosomal-recessive inheritance
AST	Aspartate aminotransferase
BRIC	Benign recurrent intrahepatic cholestasis
CNV	Copy number variation
D.Bil	Direct bilirubin
DJS	Dubin-Johnson syndrome
GGT	Gamma glutamyl transpeptidase
HGVB	Human Genetic Variation Browser
Ion PGM	Ion Torrent Personal Genome Machine
IR 4.0	Ion Reporter 4.0
MAF	Minor allele frequency
MLPA	Multiplex ligation-dependent probe amplification
NGS	Next-generation sequencing
NICCD	Neonatal intrahepatic cholestasis caused by citrin deficiency
NIIC	Neonatal/infantile intrahepatic cholestasis
PFIC	Progressive familial intrahepatic cholestasis
T.Bil	Total bilirubin

From the [1]Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; [2]Department of Hepatology, National Medical Center for Children and Mothers, National Center for Child Health and Development, Tokyo, Japan; and [3]Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan

Supported by Grant-in-Aid for Young Scientists (26870489 [to T.S.]). The authors declare no conflicts of interest. Q1

http://dx.doi.org/10.1016/j.jpeds.2016.01.006

Methods

Written informed consent was obtained from the parents. Experimental protocols were approved by the Ethical Committee for the Study of Human Gene Analysis at Nagoya City University Graduate School of Medical Sciences (approval number 150).

Serum-conjugated hyperbilirubinemia is the most common marker of cholestasis.[7] Here, cholestasis was defined as follows: (1) a serum direct bilirubin (D.Bil) level of >1.0 mg/dL, if the total bilirubin (T.Bil) was <5.0 mg/dL; or (2) a serum D.Bil level of more than 20% of the T.Bil level if the T.Bil level was >5.0 mg/dL.[3,7] From April 2013 to August 2015, we recruited Japanese subjects with NIIC as follows. The entry criteria included: (1) cholestasis; (2) onset <12 months of age; (3) date of birth between January 2010 and December 2014; and (4) no definitive molecular diagnosis previously. The exclusion criteria were extrahepatic cholestasis, such as biliary structural abnormality, or chromosomal abnormality.

Clinical Diagnosis of NIICs and Collection of the Laboratory Findings

Clinical diagnosis of all the subjects was confirmed by reviewing the available clinical and laboratory records at the time of registration. We classified the recruited subjects into 3 subcategories: (1) "genetic cholestasis," ie, subjects who were clinically diagnosed with known genetic cholestasis syndromes such as ALGS, PFIC, NICCD, and Dubin-Johnson syndrome (DJS); (2) "unknown with complications," ie, subjects with no definitive clinically identified etiology, although they showed potential cholestasis-causative complications such as prematurity, infections, and metabolic or hormonal system abnormalities; and (3) "unknown without complications," ie, subjects with no definitive etiology or no potentially cholestasis-causative complications.

ALGS with autosomal-dominant inheritance was diagnosed according to the classical definition of the presence of 3 of 5 major clinical criteria.[8] Diagnosis of neonatal DJS with autosomal-recessive inheritance (AR) was determined by the presence of a brown/black liver and a normal value of aspartate aminotransferase/alanine aminotransferase (AST/ALT).[9-11] Diagnosis of NICCD with AR was based on clinical suspicion, supported by laboratory evidence including serum amino acid profiles or liver histology.[12] The clinical diagnosis of PFIC/benign recurrent intrahepatic cholestasis (BRIC) with AR was based on the clinical history of infantile age, serologic low/normal gamma glutamyl transpeptidase (GGT) level, and histologic features.[13-15] For laboratory findings at presentation, we adopted the measurements available at the same period as the greatest D.Bil value obtained.

Targeted Genes, Amplicon Library Preparation, and NGS

An amplicon library of the target exons and flanking sequences was prepared with the use of an Ion AmpliSeq Custom Panel (Life Technologies, Carlsbad, California). This custom NIIC panel (panel ID: IAD34922) contained the *JAG1, NOTCH2, ABCC2, SLC25A13, ATP8B1, ABCB11, ABCB4, TJP2, HSD3B7, AKR1D1, CYP7B1, VPS33B, BAAT, EPHX1, SLC10A1, ABCB1, SLC4A2*, and *SLCO1A2* genes. The first 14 genes listed are known intrahepatic cholestasis disease-causing genes,[1,4] and the last 4 are potential candidate genes encoding proteins that play a role in bile acid transport. The genes on the panel were selected on the basis of the prevalence of the variants in the Japanese population, and thus customized for Japanese patients with NIIC. The number of exons, amplicons, and total targeted bases were 348, 546, and 52 795 bases, respectively. This NIIC panel allowed theoretical coverage of 98.5% of the targeted sequences (**Table I**; available at www.jpeds.com). Genomic DNA was extracted from peripheral blood using the QIAamp Blood Midi Kit (QIAGEN, Hilden, Germany). The library was constructed by use of the Ion AmpliSeq Library Kit 2.0 (Life Technologies, Carlsbad, California). Emulsion polymerase chain reaction was carried out using the Ion OneTouch 2 system (Life Technologies). NGS was performed with the Ion Torrent Personal Genome Machine (Ion PGM) system (Life Technologies).

Sequence Data Analysis Using Bioinformatics and Validation Analysis

Sequence data analysis pipelines were established with use of the workflow in CLC Genomics Workbench 7.0 (CLC bio, Aarhus, Denmark) and Ion Reporter 4.0 (IR 4.0; Life Technologies). IR 4.0 could potentially call a copy number variation (CNV). The Human Gene Mutation Database professional in January 2015 (http://www.hgmd.org/, BioBase), Sorting Intolerant From Tolerant, Polymorphism Phenotyping, and Human Splicing Finder ver 3.0 were used with computational predictive programs.[16-18] Minor allele frequency (MAF) was referred to the Japanese data set of the Human Genetic Variation Browser (HGVB, URL: http://www.genome.med.kyoto-u.ac.jp/SnpDB) that was released in November 2013 and to the Exome Aggregation Consortium, Cambridge, Massachusetts (http://exac.broadinstitute.org; accessed August 2015). The American College of Medical Genetics and Genomics interpretation guidelines were followed in assessing the pathogenicity of the detected variants.[19] The nomenclature of identified variants was assigned according to the guidelines of the Human Genome Variation Society version 2 (http://www.hgvs.org/mutnomen/). Chromosomal coordinates were assigned according to the GRCh37/hg19 assembly. All candidate variants were validated by conventional Sanger sequencing, and CNVs were confirmed by multiplex ligation-dependent probe amplification (MLPA) analysis.[20]

Preliminary Validation of the System on 5 Positive Control Cases

To evaluate the performance of our NGS-based molecular diagnosis system, we tested samples from 5 patients with NIIC genotyped by conventional Sanger sequencing. These samples carried 9 disease-causing variants in total: 6 single-nucleotide variants, 1 small insertion, 1 small deletion, and 1 gross deletion (516 bp) in *JAG1*, *SLC25A13*, or *ABCB11*. The basic sequence data showed an average depth of coverage in the target regions of 600-fold and that 97.6% of the target regions had a coverage of more than 100-fold. We successfully identified all 9 variants, confirming the efficacy of our diagnostic system.

Comparison of the Clinical and Laboratory Findings among Subjects with the Molecular Diagnosis

The values of clinical and laboratory findings among subjects with a molecular genetic diagnosis were analyzed statistically.

Statistical Analyses

Statistical analyses were performed with GraphPad Prism 6.05 for windows (GraphPad Software, Inc, La Jolla, California). Kruskal-Wallis nonparametric one-way ANOVA followed by Dunn multiple comparison test were performed, and a *P* value of <.05 was considered significant.

Results

One hundred nine subjects fulfilled our criteria and were enrolled in this study (**Figure 1**; available at www.jpeds.com). Of these 109 subjects, 31 (28%) were clinically diagnosed with genetic cholestasis (**Table II**). For all subjects, the median age of onset of symptoms associated with cholestasis was 1 month, and the median D.Bil value was 4.1 mg/dL. The median age of the subjects when we performed the NGS screening was 5 months (2-15 months, IQR).

Summary of Ion PGM Sequencing

The median number of total sequenced bases per subject, of mapped reads, and of mean read length were 61.9 mega bases (42.7-76.9, IQR), 414 kilo reads (294-525, IQR), and 148 bases (139-152, IQR), respectively. The average depth of coverage in the target region was 672-fold (447-849, IQR); 97.0% of the target regions had more than 100-fold coverage.

Molecular Genetic Diagnosis

Our sequencing and bioinformatic analysis indicated that 28 (26%) subjects received a molecular genetic diagnosis of pathogenic or likely pathogenic variants of *JAG1*, *NOTCH2*, *ABCC2*, *SLC25A13*, *ATP8B1*, or *ABCB11*. The molecular genetic diagnosis in 22 (79%) of the 28 subjects was consistent with the clinical diagnosis, as these subjects had been categorized as "genetic cholestasis." For the remaining 6 subjects who had been clinically categorized as "unknown with complications" (2 subjects) or "unknown without complications" (4 subjects), the molecular genetic diagnosis was as follows: 1 with ALGS, 3 with neonatal DJS, 1 with NICCD, and 1 with PFIC2/BRIC2. **Table III** (available at www.jpeds.com) shows the characteristics of the variants in the 28 subjects with a molecular genetic diagnosis. We

Table II. Baseline demographic and clinical characteristics*

Clinical diagnosis	No.	Sex (male)	Age (mo) at: Onset	Age (mo) at: NGS†	GA (w)	BW (g)	T.Bil (mg/dL)	D.Bil (mg/dL)	AST (IU/L)	ALT (IU/L)	GGT (IU/L)
Genetic cholestasis	31	13									
ALGS	13	5	1 (0-2)	4 (2-15)	38 (37-39)	2584 (2306-2790)	10.7 (6.0-14.2)	7.0 (4.4-9.0)	231 (112-299)	159 (70-293)	290 (106-1023)
Neonatal DJS	2	2	0 (0-0)	1.5 (1-2)	39 (39-39)	3135 (2774-3496)	16.8 (10.2-23.4)	9.8 (6.2-13.4)	32 (19-44)	19 (12-25)	148 (97-199)
NICCD	4	3	4 (4-6)	6 (4-12)	40 (39-41)	2768 (2546-3148)	5.2 (2.5-10)	2.3 (1.4-3.3)	124 (106-194)	58 (29-97)	192 (104-273)
PFIC/BRIC	12	3	2 (1-2)	10 (4-20)	38 (37-39)	2863 (2585-3255)	10.8 (3.4-15.1)	6.4 (2.2-11.4)	109 (45-233)	62 (30-153)	48 (16-67)
Unknown, complication (+)	46	25									
Perinatal abnormalities	31	17	1 (0-3)	5 (2-8)	34 (28-37)	1760 (774-2360)	6.2 (5.1-7.7)	4.1 (2.9-5.4)	125 (50-206)	54 (21-122)	110 (59-207)
Infections	7	3	2 (1-2)	3 (1-5)	39 (38-40)	3252 (2995-3380)	5.2 (3.5-10.8)	2.6 (2.4-7.1)	149 (42-836)	130 (13-517)	84 (59-204)
Metabolic acidosis	3	2	1 (1-9)	2 (1-11)	39 (37-40)	2675 (1992-2824)	7.3 (6.8-9.5)	4.5 (2.1-4.8)	2164 (143-4190)	874 (56-1673)	222 (151-317)
Abnormal hormone values	3	2	1 (0-2)	17 (2-33)	39 (36-40)	2850 (2190-3120)	10.8 (2.3-11.4)	4.2 (1.0-8.7)	137 (56-263)	43 (26-185)	100 (89-237)
Hemochromatosis	1	1	0	13	37	2122	12.7	4.6	166	31	26
Milk allergy	1	0	0	1	39	2808	14.2	3.1	50	12	1247
Unknown, complication (−)	32	16	1 (0-3)	10 (3-25)	39 (38-40)	2920 (2734-3159)	6.8 (5.3-10.9)	3.7 (2.3-7.2)	75 (48-140)	49 (30-100)	81 (45-119)
Total	109	54	1 (0-3)	5 (2-15)	38 (37-39)	2720 (2253-3045)	7.0 (5.2-10.9)	4.1 (2.6-7.0)	113 (53-213)	59 (30-151)	90 (58-215)

BW, birth weight; *GA*, gestational age; *No.*, number.

*Data are median values (IQR).

†The median age of the subjects when we performed the NGS screening.

confirmed all the variants shown in **Table III** by using conventional Sanger sequencing or MLPA analysis.

Eleven of the 13 clinically diagnosed ALGS patients had genetic mutations in either *JAG1* or *NOTCH2*; we identified 10 heterozygous pathogenic or likely pathogenic variants in *JAG1* (NCU01-10) and 1 likely pathogenic variant in *NOTCH2* (NCU11). We also identified 1 subject (NCU23) with a *JAG1* missense mutation in the "unknown with complications" group. This subject was a preterm infant with infantile cholestasis and ALGS-like facial features but no other symptoms. Thus, 12 subjects in total were defined as having ALGS by molecular genetic diagnosis. Regarding variant type in *JAG1*, we identified 1 subject (NCU10) with a deletion encompassing an entire exon, which was called by IR 4.0 as CNV 1 in the region of chr20: 10 619 976-10 654 275. We confirmed the deletion in *JAG1* using the SALSA MLPA probemix P184-C2 *JAG1* (MRC-Holland bv; Amsterdam, the Netherlands; data not shown).

Two subjects (NCU12, 13) who had been clinically diagnosed with neonatal DJS due to a black liver in macroscopic findings had compound heterozygous pathogenic or likely pathogenic variants in *ABCC2*. Surprisingly, 3 subjects (NCU25-27) who had been categorized under the group "unknown without complications" had pathogenic variants in a compound heterozygous state. These 3 subjects showed a normal value of serum AST/ALT, and their D.Bil decreased spontaneously; however, they did not have a black liver, as determined by macroscopic or microscopic study following experimental laparotomy. Thus, 5 subjects were confirmed with neonatal DJS by molecular genetic diagnosis.

All 4 subjects (NCU14-17) who had been diagnosed clinically with NICCD were confirmed by molecular genetic diagnosis as having compound heterozygous pathogenic variants in *SLC25A13*. We also identified 1 subject (NCU28) with compound heterozygous pathogenic variants in *SLC25A13* from the "unknown without complications" group. Subject NCU28 presented clinically with prolonged jaundice and failure to thrive but with a normal newborn screening test and a normal serum amino acid profile.

Five of the 12 subjects that were diagnosed clinically with PFIC/BRIC were shown by molecular genetic diagnosis to have compound heterozygous pathogenic or likely pathogenic variants; 2 (NCU18, 19) in *ATP8B1* and 3 (NCU20-22) in *ABCB11*, which molecularly confirmed these subjects to have PFIC1/BRIC1 and PFIC2/BRIC2, respectively. One subject (NCU24), who had been categorized as "unknown with complications," was molecularly diagnosed with compound heterozygous likely pathogenic variants in *ABCB11*.

No molecular genetic diagnosis was confirmed in *ABCB4*, *TJP2*, *HSD3B7*, *AKR1D1*, *CYP7B1*, *VPS33B*, *BAAT*, *EPHX1*, *SLC10A1*, *ABCB1*, *SLCO1A2*, or *SLC4A2*. However, we found 12 subjects with a pathogenic or a likely pathogenic variant on a single allele in an AR disease gene: 5 subjects in *ABCC2*, 1 in *SLC25A13*, 5 in *ABCB11*, and 1 in *ABCB4*. Therefore, these subjects did not receive a definitive molecular genetic diagnosis.

Comparison of Clinical and Laboratory Findings between the 28 Subjects with Those of the Molecular Genetic Diagnosis

Statistical significance was detected for age of onset, gestational age, birth weight, D.Bil, AST, ALT, and GGT between subjects of ALGS, neonatal DJS, NICCD, and PFIC/BRIC (**Figure 2**). The age of onset of neonatal DJS was significantly earlier than that of NICCD (0 months vs 4 months), and the D.Bil level also was significantly greater than that of NICCD (11 mg/dL vs 2.6 mg/dL) (**Figure 2**, A and B). Birth weight in ALGS was significantly smaller than that in neonatal DJS (2538 g vs 2936 g), and, on the other hand, AST/ALT levels in neonatal DJS were significantly lower than those in ALGS (25/16 IU/L vs 265/196 IU/L) (**Figure 2**, C-E). As might be predicted, the GGT level in PFIC/BRIC was significantly lower than that in ALGS (17 IU/L vs 513 IU/L) (**Figure 2**, F). Unexpectedly, the gestational age in PFIC/BRIC was significantly earlier than that in NICCD (37.5 weeks vs 40 weeks). No significant differences were found, however, in the T.Bil, total bile acids, or total cholesterol levels between these groups.

Discussion

In this study, we analyzed 109 patients with NIIC, and a molecular genetic diagnosis was made for 26% of them (28/109). We categorized the subjects into 3 groups: "genetic cholestasis" including ALGS, neonatal DJS, NICCD, and PFIC/BRIC (31 subjects); "unknown with complications" (32 subjects), and "unknown without complications" (46 subjects). The positive molecular genetic diagnosis rate in the "genetic cholestasis" group was 71% (22/31). It is noteworthy that the results of all the molecular diagnoses were consistent with those of clinical diagnosis in this group. Thus, precise clinical evaluation would be warranted to predict underlying genetic diagnosis. Nevertheless, a patient with clinical features of ALGS was reported to have compound heterozygous variants of *ATP8B1* without pathogenic variants of *JAG1* or *NOTCH2*,[21] indicating potential genetic heterogeneity in NIIC. Forty-six subjects were clinically categorized as "unknown with complications," and a molecular genetic diagnosis was made in 2 subjects (4.3%). The low positive rate could be explained by the existence of other complications that may have caused the NIIC, such as prematurity, infection, and metabolic or hormonal system abnormality.[3,22] Thirty-two subjects were categorized clinically under "unknown without complications" and a molecular genetic diagnosis was made in 4 subjects (12.5%), which supported the aforementioned finding. Taken together, 6 subjects (7.7%) of unknown etiology were successfully categorized by the molecular genetic diagnosis; 1 with ALGS, 3 with neonatal DJS, 1 with NICCD, and 1 with PFIC2/BRIC2. Molecular genetic diagnosis is undoubtedly important to confirm the clinical diagnosis and to allow the provision of appropriate genetic counseling.[8,14,20,23] Therefore, targeted NGS as reported in this study would be a powerful and practical

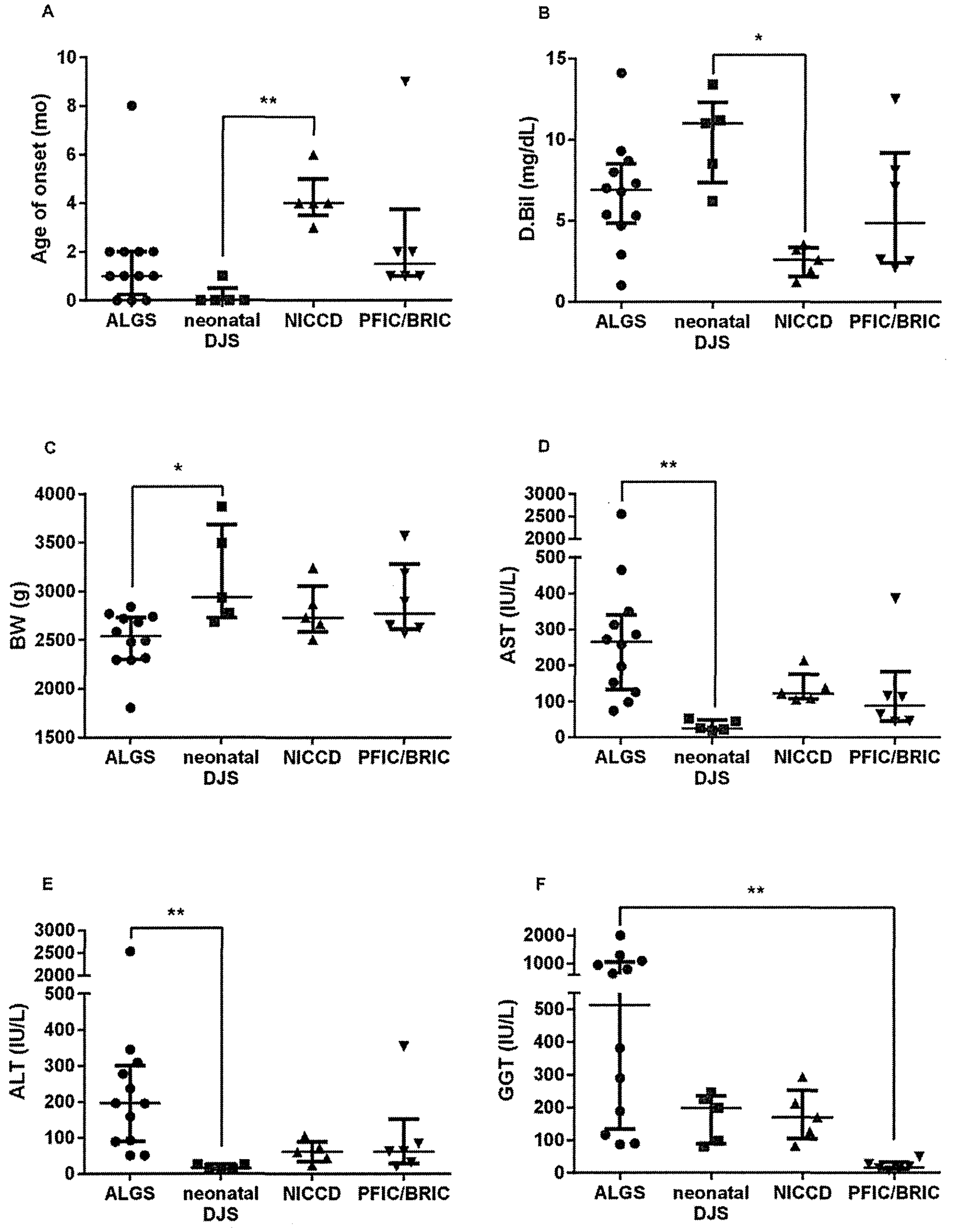

Figure 2. A, The age of onset; **B,** serum D.Bil; **C,** birth weight (BW); **D,** AST; **E,** ALT; and **F,** GGT were compared for the indicated groups of subjects. Kruskal-Wallis nonparametric one-way ANOVA followed by Dunn multiple comparison test were performed. $^{*}P < .05$; $^{**}P < .01$; error bars, IQR.

solution for the diagnosis of NIIC, especially for patients with atypical clinical presentation.

In our assay, we detected 1 whole exon deletion, 6 small heterozygous deletions, and 1 small heterozygous insertion using 2 different variant calling algorithms other than single-nucleotide variants (**Table III**). Notably, we identified whole exon deletion of *JAG1* with IR 4.0, in which the CNV calling is based on sequenced read depth. In a previous report of molecular genetic diagnosis for patients with NIIC, Liu et al[23,24] developed a hybridization-based resequencing chip, the *Jaundice Chip*, which includes 5 disease-causing genes; *SERPINA1*, *JAG1*, *ATP8B1*, *ABCB11*, and *ABCB4*. Single-nucleotide variants were detected with high accuracy by this chip. However, the authors stated that heterozygous deletions or insertions might not be detected by this chip because the strand with normal sequences would produce a normal readout.

According to the Human Gene Mutation Database professional, more than 450 *JAG1* disease-causing variants have been registered, and approximately 50% of these have been classified as gross deletions or small indels. Thus, the approach to a CNV calling with a potentially predictable bioinformatic tool should be performed in targeted NGS for patients with NIIC. Recently, Herbst et al[25] performed NGS of 6 children with infantile cholestasis for analysis of 93 genes with partially overlapping phenotypes associated with inherited cholestatic diseases. Using the Illumina MiSeq system, they made a molecular genetic diagnosis in 4 of the 6 patients and identified 6 novel variants in *PKHD1*, *ABCB11*, and *NPC1*, thus supporting the alternative NGS approach. In cases in which no variant or only a single allelic variant in a candidate AR gene are identified with targeted NGS, a further genetic investigation, such as whole-genome sequencing or an analysis of highly conserved promoter sequences in the noncoding region, might be applied.

Previous reports suggest that a heterozygous variant in *ATP8B1*, *ABCB11*, or *ABCB4* may play a role in the susceptibility to infantile intrahepatic cholestasis.[26-28] We found 5 subjects with a single heterozygous variant in *ABCB11*, including 4 missense mutations and 1 small insertion at the essential splice site, with no other candidate pathogenic variants. All of them showed a low GGT value in serum, and 2 subjects presented complete remission of cholestasis, suggesting a clinical diagnosis of PFIC/BRIC. Therefore, it is plausible that another mutation remained unidentified in this gene, or that a single heterozygous alteration might predispose to cholestasis. This issue remains to be solved in future studies.

We identified 5 neonatal DJS (including 2 siblings) subjects by molecular genetic analysis. In particular, 2 of these 5 subjects were identified as early as 2 months of age. DJS is diagnosed rarely in the infantile period, and only 5 neonatal DJS cases have undergone molecular genetic analysis.[9-11] Therefore, our two 2-month-old cases are the earliest cases to be diagnosed by molecular genetic analysis.[9-11] Six pathogenic or likely pathogenic variants in *ABCC2*, p.R100Ter, c.1815+2T>A, c.1967+2T>C, p.R768W, c.2439+2T>C, and p.K961R, were identified in our neonatal DJS (**Table III**). The p.R768W occurred with a high frequency, with 3 of 10 affected alleles in our genetically confirmed subjects and the MAF was registered as 0.252% in HGVB. We therefore suspected the presence of high-frequency pathogenic variants in *ABCC2* of neonatal DJS cases in the Japanese population such as in *SLC25A13* of NICCD.[12,29] The total MAF of these 6 variants in Japanese and East Asian or non-East Asian populations was calculated by HGVB and Exome Aggregation Consortium as 0.935%, 0.104%, and 0.00888%, respectively (**Table III**). The cumulative carrier rates of the 6 variants were predicted as approximately 1/54, 1/483, and 1/5634 on the basis of the Hardy-Weinberg principle. Thus, the estimated incidence of the affected patients with the 6 variants was calculated as approximately 1/12 000, 1/930 000, and 1/130 000 000. We speculated that cases of neonatal DJS would be more frequent in the Japanese population than in East Asian and non-East Asian populations, or that the constituent of causative pathogenic variants of neonatal DJS would be diverse between the ethnic groups.

We performed statistical analysis of the values of clinical findings for 28 subjects with a molecular genetic diagnosis. In our study, the characteristic clinical/laboratory findings of neonatal DJS showed the greatest serum value of D.Bil, and the earliest age of onset (11 mg/dL and 0 months, respectively). In ALGS, the median age of onset, AST/ALT, and GGT were 1 month, 265/196 IU/L, and 513 IU/L, respectively. Also, the GGT value in PFIC/BRIC clearly showed a low level (17 IU/L, median). These findings might help to make a differential diagnosis among conjugated hyperbilirubinemia in early infancy.

Our NIIC panel could be used in future application in clinical settings to detect disease-causing pathogenic variants as early as possible, to decrease the need of invasive procedures, to give disease-specific treatment, and to provide genetic counseling for the families. All of our patients with neonatal DJS underwent exploratory laparotomy or laparoscopy to exclude biliary atresia within 2 months of age, and 3 of them did not display a black liver. Early introduction of NGS-based molecular diagnosis would have made a proper diagnosis before the surgical intervention in these patients. Now, we have several alternatives to NGS-based molecular genetic diagnosis. Whole-exome sequencing can identify causative genetic alterations; however, whole-exome sequencing is still expensive and interpretation of results is time consuming. Small panel-based approaches are cheap, fast, and easy to interpret.[30] They also can identify CNVs, not detected by Sanger sequencing, as shown in this study. Therefore, small panels, including ours, are a practical approach for patients with NIIC, especially in early infancy, to avoid unnecessary surgical interventions.

This study presents some limitations. First, we could not access a large number of controls samples to measure the accuracy and analytical sensitivity of our assay. The accuracy and sensitivity of Ion PGM based sequencing has been reported to be compatible with other methods,[30] but we only tested our panel using 5 positive control patients with 9 disease-causing variants in total. Therefore, some mutations may not be detectable with our method. Second, our diagnostic gene panel was customized for Japanese patients with NIIC. Some cholestatic genes that are important for patients in western countries are not included, such as cystic fibrosis and alpha-antitrypsin deficiency, etc and NICCD is common in Japanese population.[29,31] An expanded panel encompassing more relevant genes could be used for the further investigations including diverse ethnic population.

In conclusion, we successfully achieved molecular genetic diagnosis in 28 (26%) of 109 Japanese patients with NIIC using targeted NGS and bioinformatics. Targeted NGS is a powerful tool that is feasible for the diagnosis of NIIC and should be introduced for early diagnosis of NIIC of unknown etiology as well as for the confirmation of clinically diagnosed genetic cholestasis syndromes. ■

We thank Drs Kenji Goto and Kohachiro Sugiyama (both deceased) for their help.

Submitted for publication Oct 10, 2015; last revision received Dec 21, 2015; accepted Jan 5, 2016.

Reprint requests: Dr Tokio Sugiura, MD, PhD, Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Kawasumi-1, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan. E-mail: tokio@med.nagoya-cu.ac.jp

References

1. Balistreri WF, Bezerra JA, Jansen P, Karpen SJ, Shneider BL, Suchy FJ. Intrahepatic cholestasis: summary of an American Association for the Study of Liver Diseases single-topic conference. Hepatology 2005;42:222-35.
2. Balistreri WF, Bezerra JA. Whatever happened to "neonatal hepatitis"? Clin Liver Dis 2006;10:27-53.
3. Feldman AG, Sokol RJ. Neonatal Cholestasis. Neoreviews 2013;14. Q7
4. Sambrotta M, Strautnieks S, Papouli E, Rushton P, Clark BE, Parry DA, et al. Mutations in TJP2 cause progressive cholestatic liver disease. Nat Genet 2014;46:326-8.
5. Koshimizu E, Miyatake S, Okamoto N, Nakashima M, Tsurusaki Y, Miyake N, et al. Performance comparison of bench-top next generation sequencers using microdroplet PCR-based enrichment for targeted sequencing in patients with autism spectrum disorder. PLoS One 2013;8:e74167.
6. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30:434-9.
7. Moyer V, Freese DK, Whitington PF, Olson AD, Brewer F, Colletti RB, et al. Guideline for the evaluation of cholestatic jaundice in infants: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. J Pediatr Gastroenterol Nutr 2004;39:115-28.
8. Turnpenny PD, Ellard S. Alagille syndrome: pathogenesis, diagnosis and management. Eur J Hum Genet 2012;20:251-7.
9. Lee JH, Chen HL, Chen HL, Ni YH, Hsu HY, Chang MH. Neonatal Dubin-Johnson syndrome: long-term follow-up and MRP2 mutations study. Pediatr Res 2006;59:584-9.
10. Pacifico L, Carducci C, Poggiogalle E, Caravona F, Antonozzi I, Chiesa C, et al. Mutational analysis of ABCC2 gene in two siblings with neonatal-onset Dubin Johnson syndrome. Clin Genet 2010;78:598-600.
11. Okada H, Kusaka T, Fuke N, Kunikata J, Kondo S, Iwase T, et al. Neonatal Dubin-Johnson syndrome: Novel compound heterozygous mutation in the ABCC2 gene. Pediatr Int 2014;56:e62-4.
12. Ohura T, Kobayashi K, Tazawa Y, Abukawa D, Sakamoto O, Tsuchiya S, et al. Clinical pictures of 75 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD). J Inherit Metab Dis 2007;30:139-44.
13. Pawlikowska L, Strautnieks S, Jankowska I, Czubkowski P, Emerick K, Antoniou A, et al. Differences in presentation and progression between severe FIC1 and BSEP deficiencies. J Hepatol 2010;53:170-8.
14. van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, Bull LN, et al. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 2004;127:379-84.
15. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E. Progressive familial intrahepatic cholestasis. Orphanet J Rare Dis 2009;4:1.
16. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812-4.
17. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894-900.
18. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.
19. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405-24.
20. Jurkiewicz D, Gliwicz D, Ciara E, Gerfen J, Pelc M, Piekutowska-Abramczuk D, et al. Spectrum of JAG1 gene mutations in Polish patients with Alagille syndrome. J Appl Genet 2014;55:329-36.
21. Grochowski CM, Rajagopalan R, Falsey AM, Loomes KM, Piccoli DA, Krantz ID, et al. Exome sequencing reveals compound heterozygous mutations in ATP8B1 in a JAG1/NOTCH2 mutation-negative patient with clinically diagnosed Alagille syndrome. Am J Med Genet A 2015;167A:891-3.
22. Jancelewicz T, Barmherzig R, Chung CT, Ling SC, Kamath BM, Ng VL, et al. A screening algorithm for the efficient exclusion of biliary atresia in infants with cholestatic jaundice. J Pediatr Surg 2015;50:363-70.
23. Matte U, Mourya R, Miethke A, Liu C, Kauffmann G, Moyer K, et al. Analysis of gene mutations in children with cholestasis of undefined etiology. J Pediatr Gastroenterol Nutr 2010;51:488-93.
24. Liu C, Aronow BJ, Jegga AG, Wang N, Miethke A, Mourya R, et al. Novel resequencing chip customized to diagnose mutations in patients with inherited syndromes of intrahepatic cholestasis. Gastroenterology 2007;132:119-26.
25. Herbst SM, Schirmer S, Posovszky C, Jochum F, Rodl T, Schroeder JA, et al. Taking the next step forward - Diagnosing inherited infantile cholestatic disorders with next generation sequencing. Mol Cell Probes 2015;29:291-8.
26. Jacquemin E, Malan V, Rio M, Davit-Spraul A, Cohen J, Landrieu P, et al. Heterozygous FIC1 deficiency: a new genetic predisposition to transient neonatal cholestasis. J Pediatr Gastroenterol Nutr 2010;50:447-9.
27. Hermeziu B, Sanlaville D, Girard M, Leonard C, Lyonnet S, Jacquemin E. Heterozygous bile salt export pump deficiency: a possible genetic predisposition to transient neonatal cholestasis. J Pediatr Gastroenterol Nutr 2006;42:114-6.
28. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E. The spectrum of liver diseases related to ABCB4 gene mutations: pathophysiology and clinical aspects. Semin Liver Dis 2010;30:134-46.
29. Tabata A, Sheng JS, Ushikai M, Song YZ, Gao HZ, Lu YB, et al. Identification of 13 novel mutations including a retrotransposal insertion in SLC25A13 gene and frequency of 30 mutations found in patients with citrin deficiency. J Hum Genet 2008;53:534-45.
30. Li X, Buckton AJ, Wilkinson SL, John S, Walsh R, Novotny T, et al. Towards clinical molecular diagnosis of inherited cardiac conditions: a comparison of bench-top genome DNA sequencers. PLoS One 2013;8:e67744.
31. Yamashiro Y, Shimizu T, Oguchi S, Shioya T, Nagata S, Ohtsuka Y. The estimated incidence of cystic fibrosis in Japan. J Pediatr Gastroenterol Nutr 1997;24:544-7.

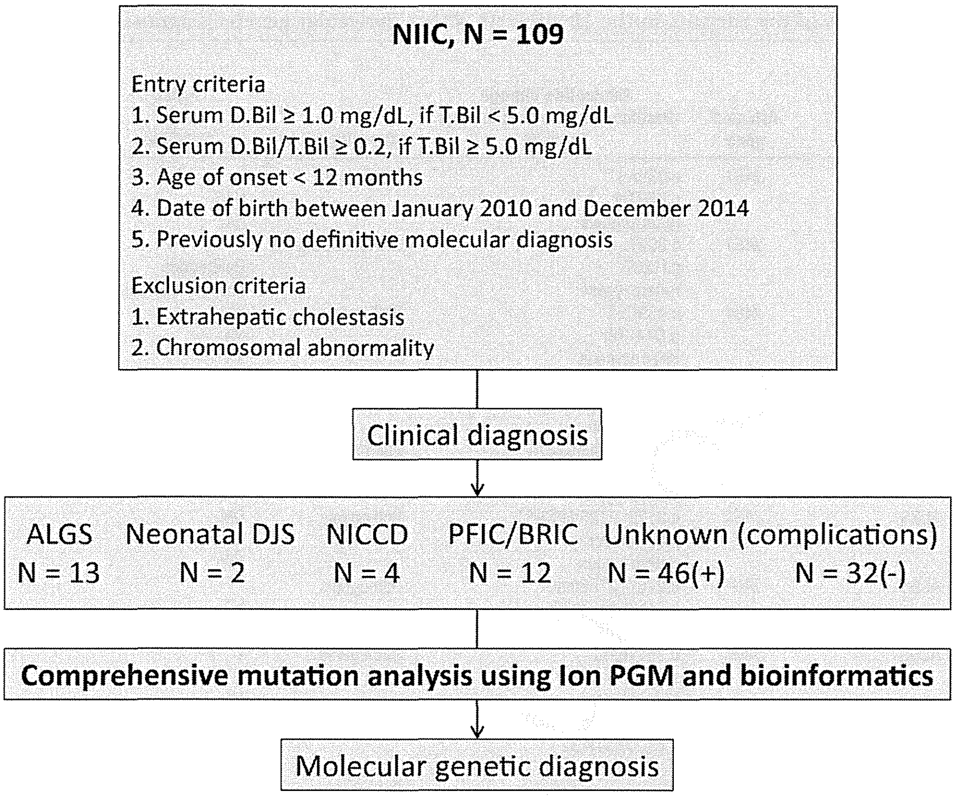
NIIC, N = 109
Entry criteria
1. Serum D.Bil ≥ 1.0 mg/dL, if T.Bil < 5.0 mg/dL
2. Serum D.Bil/T.Bil ≥ 0.2, if T.Bil ≥ 5.0 mg/dL
3. Age of onset < 12 months
4. Date of birth between January 2010 and December 2014
5. Previously no definitive molecular diagnosis
Exclusion criteria
1. Extrahepatic cholestasis
2. Chromosomal abnormality
Clinical diagnosis
ALGS N = 13
Neonatal DJS N = 2
NICCD N = 4
PFIC/BRIC N = 12
Unknown (complications) N = 46(+) N = 32(-)
Comprehensive mutation analysis using Ion PGM and bioinformatics
Molecular genetic diagnosis

Figure 1. Flow diagram of the study and study subjects.

Table I. The panel of 18 targeted genes

Gene name	Disorder/function	No. exons	No. amplicons	Total no. targeted bases	% of coverage of target sequence
JAG1	ALGS	26	35	3657	99.9
NOTCH2	ALGS	35	63	7602	97.9
ABCC2	DJS	32	44	4638	99.7
SLC25A13	NICCD	19	28	2116	99.1
ATP8B1	PFIC1/BRIC1	27	45	3756	98.1
ABCB11	PFIC2/BRIC2	27	43	3966	99.0
ABCB4	PFIC3	28	44	4068	97.6
TJP2	PFIC4, FHCA	27	38	4158	95.6
HSD3B7	BAS	7	9	1170	100
AKR1D1	BAS	10	12	999	99.6
CYP7B1	BAS	6	17	1521	92.6
VPS33B	ARCS1	23	23	1854	100
BAAT	FHCA	3	10	1257	100
EPHX1	FHCA	8	15	1368	100
SLC10A1	NTCP	5	10	1050	98.7
ABCB1	MDR1	27	46	3843	99.2
SLC4A2	AE2	24	38	3759	98.6
SLCO1A2	OATP-1	14	26	2013	96.9
	Total	348	546	52 795	98.5

AE2, anion exchange protein 2; *ARCS1*, arthrogryposis, renal dysfunction and cholestasis syndrome-1; *BAS*, bile acid synthetic defect; *FHCA*, familial hypercholanemia; *MDR1*, multidrug resistance protein 1; *No.*, number; *NTCP*, Na(+)/taurocholate transport protein; *OATP-1*, organic anion transporting polypeptide 1.

Table III. Characteristics of the variants in the 28 subjects with a molecular genetic diagnosis

							MAF	
Subjects	Sex	Clinical diagnosis	Affected gene	Nucleotide change Predicted amino acid change Zygosity	Classification*	HGMD SIFT PolyPhen	HGVB ExAC (ESA) ExAC (Non-ESA)	Parental origin
NCU01	F	ALGS	*JAG1*	c.238A>T p.K80Ter Heterozygous	Pathogenic	- NA NA	- - -	De novo
NCU02	M	ALGS	*JAG1*	c.359T>C p.I120T Heterozygous	Likely Pathogenic	- Deleterious possibly damaging	- - -	ND
NCU03	F	ALGS	*JAG1*	c.439C>T p.Q147Ter Heterozygous	Pathogenic	DM NA NA	- - -	Maternal
NCU04	M	ALGS	*JAG1*	c.551G>A p.Arg184His Heterozygous	Pathogenic	DM Deleterious probably damaging	- - -	ND
NCU05	F	ALGS	*JAG1*	c.785_789delGTGAT p.Cys262fsTer1 Heterozygous	Pathogenic	DM NA NA	- - -	De novo
NCU06	F	ALGS	*JAG1*	c.2122_2125delCAGT p.Q708fsTer34 Heterozygous	Pathogenic	DM NA NA	- - -	ND
NCU07	M	ALGS	*JAG1*	c.2767_2768insG p.Asp923fsTer29 Heterozygous	Pathogenic	- NA NA	- - -	De novo
NCU08	F	ALGS	*JAG1*	c.2927delC p.T976fsTer8 Heterozygous	Pathogenic	- NA NA	- - -	De novo
NCU09	F	ALGS	*JAG1*	c.3007_3010delGAGC p.E1003fsTer32 Heterozygous	Pathogenic	- NA NA	- - -	De novo
NCU10	F	ALGS	*JAG1*	Whole exons deletion (Chromosome 20: 10 619 976-10 654 275)	Pathogenic	- NA NA	- - -	De novo
NCU11	M	ALGS	*NOTCH2*	c.5758G>C p.A1920P Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - -	Maternal
NCU12	M	Neonatal DJS	*ABCC2*	c.1815+2T>A Splice-site disruption Heterozygous	Pathogenic	DM NA NA	- 1.16e-04 -	Maternal
			ABCC2	c.2302C>T p.R768W Heterozygous	Pathogenic	DM Deleterious probably damaging	2.52e-03 3.47e-04 5.33e-05	Paternal
NCU13	M	Neonatal DJS	*ABCC2*	c.2302C>T p.R768W Heterozygous	Pathogenic	DM Deleterious probably damaging	2.52e-03 3.47e-04 5.33e-05	Paternal
			ABCC2	c.2882A>G Splice-site disruption (p.K961R†) Heterozygous	Likely pathogenic	- Tolerated benign	2.74e-03 3.50e-04 -	Maternal
NCU14	M	NICCD	*SLC25A13*	c.615+1G>C Splice-site disruption Heterozygous	Pathogenic	DM NA NA	- - -	Maternal
			SLC25A13	c.1592G>A p.G531Asp Heterozygous	Pathogenic	DM Deleterious probably damaging	1.67e-03 - -	Paternal
NCU15	M	NICCD	*SLC25A13*	c.674C>A p.S225Ter Heterozygous	Pathogenic	DM NA NA	1.67e-03 1.16e-04 -	Maternal
			SLC25A13	c.852_855delTATG p.R284fsTer3 Heterozygous	Pathogenic	DM NA NA	2.06e-03 3.71e-03 -	Paternal
NCU16	M	NICCD	*SLC25A13*	c.852_855delTATG p.R284fsTer3 Heterozygous	Pathogenic	DM NA NA	2.06e-03 3.71e-03 -	ND
			SLC25A13	c.1078C>T p.R360Ter Heterozygous	Pathogenic	DM NA NA	- 1.16e-04 3.55e-05	ND

(*continued*)

Table III. Continued

							MAF	
Subjects	Sex	Clinical diagnosis	Affected gene	Nucleotide change Predicted amino acid change Zygosity	Classification*	HGMD SIFT PolyPhen	HGVB ExAC (ESA) ExAC (Non-ESA)	Parental origin
NCU17	F	NICCD	*SLC25A13*	c.852_855delTATG p.R284fsTer3 Heterozygous	Pathogenic	DM NA NA	2.06e-03 3.71e-03 -	Paternal
			SLC25A13	c.1177+1G>A Splice site disruption Heterozygous	Pathogenic	DM NA NA	3.18e-03 1.04e-03 -	Maternal
NCU18	F	PFIC/BRIC	*ATP8B1*	c.916T>C p.C306R Heterozygous	Pathogenic	DM Deleterious probably damaging	- - -	Maternal
			ATP8B1	c.2854C>T p.R952Ter Heterozygous	Pathogenic	DM NA NA	- - 8.87e-06	Paternal
NCU19	F	PFIC/BRIC	*ATP8B1*	c.922G>A p.G308S Heterozygous	Likely Pathogenic	- Deleterious probably damaging	- - -	Paternal
			ATP8B1	c.3579_3589delACGGCAGCAGG p.R1193fsTer39 Heterozygous	Pathogenic	- NA NA	- - -	Maternal
NCU20	F	PFIC/BRIC	*ABCB11*	c.386G>A p.C129Y Heterozygous	Likely Pathogenic	- Tolerated possibly damaging	- - -	Maternal
			ABCB11	c.1460G>A p.R487H Heterozygous	Pathogenic	DM Tolerated possibly damaging	- 1.22e-04 1.19e-04	Paternal
NCU21	F	PFIC/BRIC	*ABCB11*	c.386G>A p.C129Y Heterozygous	Likely pathogenic	- Tolerated possibly damaging	- - -	ND
			ABCB11	c.3839T>A p.I1280N Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - -	ND
NCU22	M	PFIC/BRIC	*ABCB11*	c.3121T>C p.Y1041H Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - -	Maternal
			ABCB11	c.3904G>T p.E1302Ter Heterozygous	Pathogenic	DM NA NA	- - -	Paternal
NCU23	M	Unknown with complications	*JAG1*	c.1262G>A p.C421Y Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - -	Maternal
NCU24	F	Unknown with complications	*ABCB11*	c.1709C>T p.A570V Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - -	Paternal
			ABCB11	c.3802C>T p.R1268W Heterozygous	Likely pathogenic	- Deleterious probably damaging	- - 8.92e-06	Maternal
NCU25‡	F	Unknown without complications	*ABCC2*	c.298C>T p.R100Ter Heterozygous	Pathogenic	DM NA NA	- 1.16e-04 2.66e-05	Paternal
			ABCC2	c.2439+2T>C Splice site disruption Heterozygous	Pathogenic	DM NA NA	2.72e-03 1.16e-04 8.89e-06	Maternal
NCU26‡	M	Unknown without complications	*ABCC2*	c.298C>T p.R100Ter Heterozygous	Pathogenic	DM NA NA	- 1.16e-04 2.66e-05	Paternal
			ABCC2	c.2439+2T>C Splice-site disruption Heterozygous	Pathogenic	DM NA NA	2.72e-03 1.16e-04 8.89e-06	Maternal

(*continued*)

Table III. Continued

							MAF	
Subjects	Sex	Clinical diagnosis	Affected gene	Nucleotide change Predicted amino acid change Zygosity	Classification*	HGMD SIFT PolyPhen	HGVB ExAC (ESA) ExAC (Non-ESA)	Parental origin
NCU27	F	Unknown without complications	*ABCC2*	c.1967+2T>C Splice-site disruption Heterozygous	Pathogenic	DM NA NA	1.37e-03 - -	ND
			ABCC2	c.2302C>T p.R768W Heterozygous	Pathogenic	DM Deleterious probably damaging	2.52e-03 3.47e-04 5.33e-05	ND
NCU28	F	Unknown without complications	*SLC25A13*	c.15G>A Splice-site disruption Heterozygous	Pathogenic	DM NA NA	- - -	Maternal
			SLC25A13	c.1177+1G>A Splice-site disruption Heterozygous	Pathogenic	DM NA NA	3.18e-03 1.04e-03 -	Paternal

DM, disease-causing mutation; *ExAC (ESA)*, Exome Aggregation Consortium (East Asian); *ExAC (Non-ESA)*, Exome Aggregation Consortium (Non-East Asian); *F*, female; *HGMD*, Human Gene Mutation Database; *M*, male; *NA*, not applicable; *ND*, not determined; *PolyPhen*, Polymorphism Phenotyping; *SIFT*, Sorting Intolerant From Tolerant; -, no registration data.

Accession number: *JAG1*, NM_000214.2; *NOTCH2*, NM_024408.3; *ABCC2*, NM_000392.3; *SLC25A13*, NM_014251.2; *ATP8B1*, NM_005603.4; *ABCB11*, NM_003742.2.

*According to the American College of Medical Genetics and Genomics interpretation guidelines.[19]

†The mutation in NCU-13 was predicted as a "Broken WT Donor Site" by Human Splicing Finder.

‡Sibling.

nature
COMMUNICATIONS

ARTICLE

Received 22 Nov 2014 | Accepted 7 Jul 2015 | Published 21 Aug 2015 DOI: 10.1038/ncomms9018 OPEN

Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals

Masao Nagasaki[1,2,3,*], Jun Yasuda[1,2,*], Fumiki Katsuoka[1,2,*], Naoki Nariai[1,†], Kaname Kojima[1,2], Yosuke Kawai[1,2], Yumi Yamaguchi-Kabata[1,2], Junji Yokozawa[1,2], Inaho Danjoh[1,2], Sakae Saito[1,2], Yukuto Sato[1,2], Takahiro Mimori[1], Kaoru Tsuda[1], Rumiko Saito[1], Xiaoqing Pan[1,†], Satoshi Nishikawa[1], Shin Ito[1], Yoko Kuroki[1,†], Osamu Tanabe[1,2], Nobuo Fuse[1,2], Shinichi Kuriyama[1,2,4], Hideyasu Kiyomoto[1,2], Atsushi Hozawa[1,2], Naoko Minegishi[1,2], James Douglas Engel[5], Kengo Kinoshita[1,3,6], Shigeo Kure[1,2], Nobuo Yaegashi[1,2], ToMMo Japanese Reference Panel Project[#] & Masayuki Yamamoto[1,2]

The Tohoku Medical Megabank Organization reports the whole-genome sequences of 1,070 healthy Japanese individuals and construction of a Japanese population reference panel (1KJPN). Here we identify through this high-coverage sequencing (32.4 × on average), 21.2 million, including 12 million novel, single-nucleotide variants (SNVs) at an estimated false discovery rate of <1.0%. This detailed analysis detected signatures for purifying selection on regulatory elements as well as coding regions. We also catalogue structural variants, including 3.4 million insertions and deletions, and 25,923 genic copy-number variants. The 1KJPN was effective for imputing genotypes of the Japanese population genome wide. These data demonstrate the value of high-coverage sequencing for constructing population-specific variant panels, which covers 99.0% SNVs of minor allele frequency ≥0.1%, and its value for identifying causal rare variants of complex human disease phenotypes in genetic association studies.

[1] Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai 980-8573, Japan. [2] Graduate School of Medicine, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan. [3] Graduate School of Information Sciences, Tohoku University, 6-3-09, Aramaki Aza-Aoba, Aoba-ku, Sendai 980-8579, Japan. [4] International Research Institute of Disaster Science, Tohoku University, 468-1, Aramaki Aza-Aoba, Aoba-ku, Sendai 980-0845, Japan. [5] Department of Cell and Developmental Biology, University of Michigan Medical School, 109 Zina Pitcher Place, Ann Arbor, Michigan 48109-2200, USA. [6] Institute of Development, Aging and Cancer, Tohoku University, 4-1, Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan. [*] These authors contributed equally to this work. [†] Present addresses: Institute for Genomic Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA (N.N. & Y.T.); National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104-0045, Japan (X.P.); National Center for Child Health and Development, National Medical Center for Children and Mothers Research Institute, 2-10-1, Okura, Setagaya-ku, Tokyo 157-8535, Japan (Y.K.). [#] A full list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to M.N. (email: nagasaki@megabank.tohoku.ac.jp) or to M.Y. (email: masiyamamoto@med.tohoku.ac.jp).

Tohoku Medical Megabank Organization (ToMMo) established a biobank that combines medical and genome information for the community medical system in the Tohoku region, located in the northeast part of Japan. We initiated the prospective genome cohort study in the region to identify genetic and environmental factors in diseases, and to enable personalized medicine based on an individual's genomic information. In the experimental design, we performed whole-genome sequencing (WGS) of 1,070 samples by PCR-free sequencing with more than 30 × coverage genome wide. This enabled us to identify very rare as well as novel single-nucleotide variants (SNVs), which was impossible to find by single-nucleotide polymorphism (SNP) microarrays or are difficult to find by low coverage sequencing on the same sample size. Notably, all sequencing and bioinformatics analyses were conducted using the same protocols in a single institute, allowing stringent control over systematic errors that might arise from using different equipment, protocols or bioinformatics pipelines.

Since the International Human Genome Project was completed[1], a great deal of effort has been devoted to discovering, cataloguing and haplotyping common nucleotide sequence variants in populations by targeted sequencing and SNP arrays, such as in the International HapMap Project[2]. The knowledge of these variants enabled genome-wide association studies (GWASs), through which many SNPs associated with human traits and diseases have been discovered[3,4] under the assumption of common-disease/common-variants hypothesis[5]. However, these identified SNPs can only explain a small fraction of genotype–phenotype relationships underlying the problem of 'missing heritability[6]'. Recent GWAS conducted on a large number of case and control samples revealed that lower-frequency variants contribute to a substantial fraction of the heritability of common diseases, such as type 2 diabetes[7] and cancers[8]. The 1000 Genomes Project (1KGP) has catalogued human genetic variations at minor allele frequency (MAF) $>1\%$ through WGS from multiple ethnic groups[9]. However, low frequency ($0.5\%<\mathrm{MAF}\leq5\%$), rare ($0.1\%<\mathrm{MAF}\leq0.5\%$) or very-rare ($\mathrm{MAF}\leq0.1\%$) variants have not been thoroughly catalogued, in comparison with common variants ($\mathrm{MAF}>5\%$) because the existence of lower-frequency variants is population-specific[9]. Targeted high-coverage sequencing has been conducted to detect very-rare variants in the coding sequences of European Americans and African Americans[10,11], while, in contrast, the majority of rare and very-rare variants in intergenic regions have not yet been discovered. Since the ENCODE project showed that a substantial fraction of noncoding regions (80.4%) are biochemically active[12], these data suggest that these regions may also be associated with phenotypes or diseases.

Structural variants (SVs), including copy-number variants (CNVs), have also been catalogued in several studies[13,14]. In addition, a number of studies have revealed associations of SVs with occurrence of disease, including autism[15], schizophrenia[16] and Crohn's disease[17]. CNVs may also explain traits of populations, such as dietary habits of agricultural societies and hunter–gatherers[18], or drug responses[19]. For identifying SVs, a middle-coverage sequencing strategy (~13 ×) with 250 parent–offspring families by the Genome of the Netherlands succeeded in extending the catalogue of deletions from 20 to 100 bp compared with the 1KGP data set[20]. As high-coverage WGS is becoming less expensive, it is now feasible to detect lower-frequency SVs as well as SNVs[21–23] in specific target populations[24].

From the identified SNVs, we construct a reference panel of 1,070 Japanese individuals (1KJPN), including some very-rare SNVs. The 1KJPN cohort provides unique insights into the landscape of functional variations, especially in noncoding regions. We demonstrate here that 1KJPN is useful for genotype imputation for the Japanese population. In this analysis, a functional variant associated with Moyamoya disease (MMD) was identified through the imputed genotypes based on 1KJPN.

Results

Data processing and variant discovery. From the collected DNA samples in the ToMMo biobank, we selected 1,344 candidates for constructing 1KJPN, considering the traceability of participants' information, and the quality and abundance of DNA samples for SNP array genotyping and WGS analysis. All participants provided written informed consent, and all DNA samples and personal information were analysed anonymously. All the DNA samples were genotyped with Illumina HumanOmni2.5-8 BeadChip (Omni2.5). Among the genotyped samples, 1,070 samples were then selected by filtering out close relatives and outliers (Supplementary Fig. 1).

PCR bias is one of the major sources of sequencing error[25]. Hence, the selected 1,070 samples were sequenced by Illumina HiSeq 2500 using the latest PCR-free protocol (162 bp paired-end reads and 550 bp insert size, improving the accuracy for detecting SVs[26]; Fig. 1a and Methods). An in-house density check protocol was employed before sequencing[27], and subsequent data quality-control (QC) was performed with originally developed software[28] named SUGAR to maximize the quality and throughput of each experiment. In total, 100.4 trillion bases of DNA sequence reads were generated (Table 1). The sequence reads were aligned to the human reference genome (GRCh37/hg19) with decoy sequences (hs37d5), and then variants were called by several computational algorithms (see Methods). This strategy led to the discovery of 29.6 million SNVs (the high-sensitive SNVs), 1.97 million short deletions (72.6% novel), 1.38 million short insertions (<100 bp; 75.0% novel), 47,343 large deletions and 9,354 large insertions (equal to or longer than 100 bp) in autosomes (Table 1).

To obtain reliable SNV calls, we applied multiple filtering steps (Supplementary Fig. 2), including the depth of coverage of reads (Fig. 1a), software-derived biases, departure from Hardy–Weinberg equilibrium (HWE) and complexity of genomic regions around variants (Methods and Supplementary Table 1). The performance of genotype calls in the high-confidence SNVs was improved after filtering (Fig. 1b). Consequently, we obtained a 21.2-million SNV call set, which hereafter referred to as the high-confidence SNVs (56.6% are novel; Table 1 and Fig. 1c). The high novel SNV ratio is consistent with a previous observation that rare variants tend to be population-specific[20]. The false discovery rate (FDR) of the high-confidence SNVs was confirmed using several different experimental technologies (see Methods); the FDR for SNVs, deletions and insertions were 0% (0 out of 174; confidence interval (CI) 0.0–1.10%), 0% (0 out of 32; CI 0.0–5.78%) and 3.85% (1 out of 22; CI 0.49–19.34%), respectively (Supplementary Table 2). We further conducted validation experiments for novel SNVs using a custom-designed Illumina SNP array. Combined with the genotyping results obtained with Omni2.5, the overall FDR was 0.8% with CI 0.63–0.97% (Supplementary Table 3 and Methods). It is important to note that the estimates of FDRs were not strongly affected by MAF, indicating that the discoveries of novel variants in this study are fairly robust with respect to the allele frequency.

Estimation of variant discovery rate. We estimated the rate of variant discovery with the sample size of 1,070. Because the distribution of allele frequency in a population is affected by underlying demographic history[29,30], we inferred the demographic model of 1KJPN population from the site frequency spectrum (SFS) constructed from the intergenic

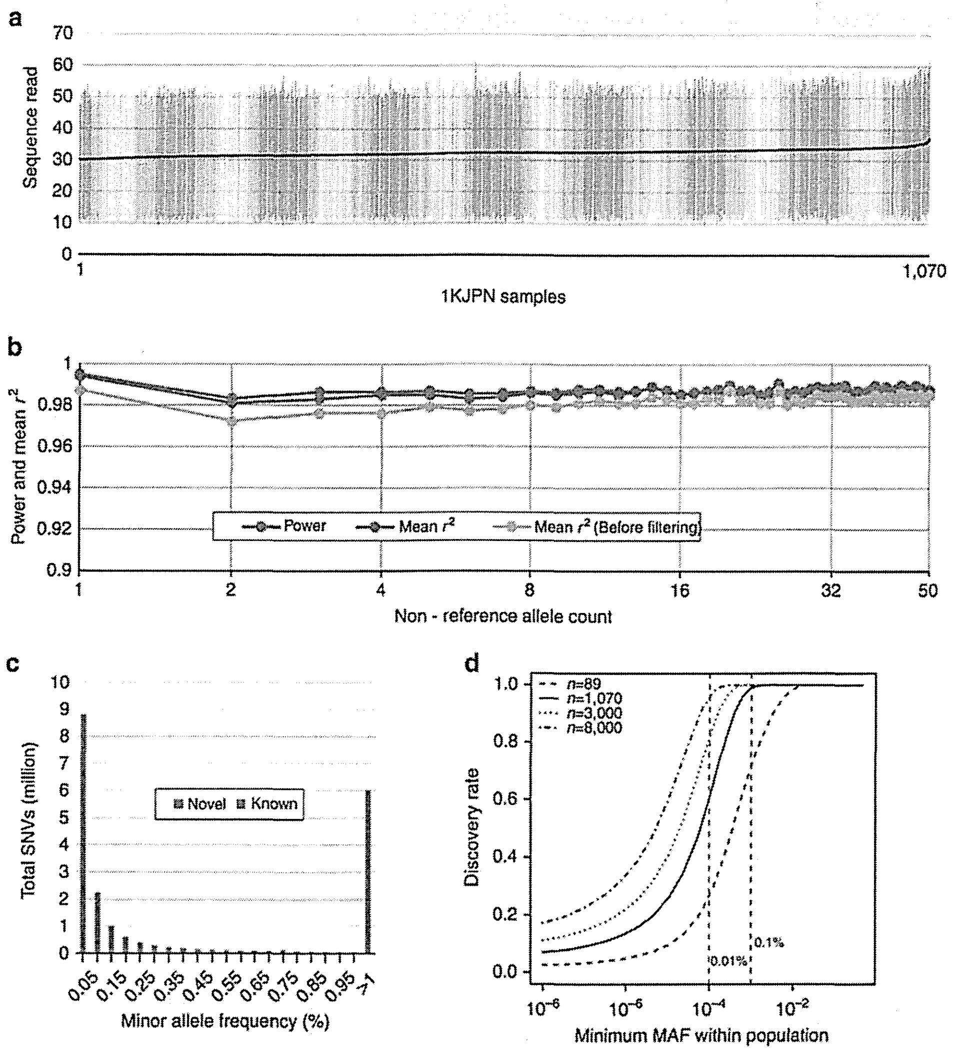

Figure 1 | SNVs in 1KJPN. (**a**) Statistics on read depth in 1KJPN. The vertical bars indicate the minimum and maximum depth of the number of sequence reads on each individual after filtering. They were sorted according to the average sequenced read depth (the black line). (**b**) The plot shows the power to detect SNVs (blue) of the confidence SNVs and the mean r^2 values before (yellow) and after (orange) filtering with SNP array data for the same sample on non-reference allele counts ranging from 1 to 50. The r^2 between genotypes from the SNVs in 1KJPN and the SNP array data is given by the squared Pearson correlation. (**c**) The numbers of novel and known SNVs in each MAF bin. The novel SNV frequency begins to dominate for lower MAFs. (**d**) The rate of variant discovery by minimum MAF in the 1KJPN population. The rates of variant discovery in our sequencing strategy were plotted against minimum MAF in the 1KJPN population by different sampling size. The distribution of population MAF was estimated on the basis of the demographic model shown in Supplementary Fig. 3.

regions (Supplementary Fig. 3a). As expected from excess in rare variants, the population of 1KJPN has experienced recent population expansion (Supplementary Fig. 3b), which is consistent with previous studies[10,30]. This demographic model is used for the calculation of the variant discovery rate (Fig. 1d and Supplementary Methods). According to Fig. 1d, 99.0 % of SNVs with MAF 0.1% or larger were expected to be captured by the present sampling strategy.

Functional impact of very-rare variants. The high-coverage PCR-free protocol of WGS has generated higher-power SNP discovery, especially for rare and very-rare SNVs. On comparison of the SFSs of intergenic region of 1KJPN and 1KGP SNVs, the data demonstrate a higher proportion of very-rare SNVs in the 1KJPN data set than in 1KGP phase 1 data set (Fig. 2a). In addition, although the number of SNVs in 1KGP was generally larger than that in 1KJPN, the number of very-rare variants detected in intergenic region was higher in 1KJPN than in 1KGP (Fig. 2b). These observations imply that there was less bias in discovering the very-rare variants of 1KJPN even in the intergenic region, although a possibility of faster expansion rate in 1KJPN cannot be excluded.

Because deleterious mutations are removed from populations faster than neutral mutations, SNVs observed at lower frequency in a population are indicative of purifying (negative) selection, and their selection strength differs among the various functional genomic categories. Along with this idea, the SFS has been analysed to evaluate relative influence of (mostly) negative selection on SNVs of each functional category in large sequencing projects[9,10,20]. Because we conducted the WGS without PCR amplification or exome capture, it is expected that there are less bias in variant detection between coding and noncoding regions. Therefore, the SFS of functional categories can be directly

Table 1 | Summary of WGS of Japanese individuals and variant detection in autosomes.

Info		
Total samples		1,070
Total raw bases		100.4 trillion bases
Mean sequenced depth		32.4 ×
SNVs	High-sensitive SNVs	High-confidence SNVs
Total	29,588,649	21,221,195
Number of known variants*	12,308,520	9,219,783
Number of novel variants*	17,280,129	12,001,412
Novelty rate	58.40%	56.55%
Average number per sample	3,886,081	2,716,853
Average individual heterozygosity	2,252,841	1,532,773
	Length	
Deletions	1 bp ≤ length < 100 bp	100 bp ≤ length
Number of sites overall	1,969,302	47,343
Number of novel variants†	1,429,636	—
Novelty rate	72.60%	—
Number of inframe/frameshift	3,112/4,454	—
Average number per sample	190,857	2,654
	Length	
Insertions	1 bp ≤ length < 100 bp	100 bp ≤ length
Number of sites overall	1,384,230	9,354
Number of novel variants†	1,037,839	9,354
Novelty rate	74.98%	—
Number of inframe/frameshift	1,577/2,506	—
Average number per sample	159,359	45

SNV, single-nucleotide variant; WGS, whole-genome sequencing.
All data listed here are limited to the autosomal genome.
*Comparison based on dbSNP build 138.
†The decision of novel sites is described in Methods.

compared with intergenic region. We classified the high-confidence SNVs into predicted functional categories and evaluated the effect of purifying selection as a fraction of very-rare variants (FVRV; Fig. 2c–f). The FVRV of intergenic regions was 40.1%, which was the lowest among all categories (Fig. 2c), supporting the notion that most of the sites in intergenic regions are evolutionarily neutral. In contrast, FVRVs of noncoding regions other than intergenic were significantly higher than the FVRV of intergenic regions—introns (41.6%), synonymous (43.7%), 3′-untranslated region (UTR, 43.9%) and 5′-UTR (45.0%)—implying that a substantial fraction of noncoding regions are functional and under weak purifying selection. Similar tendencies were observed for insertions and deletions (Supplementary Fig. 4a). We conducted the same analysis with 1KGP phase I data set. In contrast to 1KJPN, the FVRVs of 5′-UTR, 3′-UTR and intron from 1KGP data set were lower than the FVRV of synonymous SNVs (Supplementary Fig. 4b). This might not be a signature of weak purifying selection on UTR and intron regions. This is rather due to the low power of SNP discovery of 1KGP in these regions where majority of them have been sequenced with low coverage in the project.

Mutations that disrupt protein and/or transcript structure are highly detrimental. This analysis reconfirmed that the FVRV of nonsynonymous transcribed SNPs (52.5%) was distinctly higher than synonymous (43.7%) and intronic (41.6%) variants (Fig. 2c). Nonetheless, the FVRV of loss of function mutations (61.4%) was much higher than nonsynonymous SNVs (Fig. 2d). In addition, we detected heterogeneity in the FVRV of nonsynonymous SNVs in terms of functional consequences predicted by PolyPhen-2, as previously reported[10]. The FVRV of SNVs that were predicted to be 'probably damaging' was the highest (61.8%), followed by the fraction that was 'possibly damaging' (56.8%) and finally 'benign' (48.2%). We can also infer the impact of purifying selection on disease-causing mutations, those categorized as 'disease mutations' in the Human Gene Mutation Database (HGMD)[31] in terms of FVRV. The FVRV of disease mutations was 48.4%, which is very close to benign SNVs.

Although the intergenic region exhibits the lowest FVRV, the ENCODE project revealed that a large proportion of intergenic regions may be associated with biochemical activity[32]. Thus, we inferred the influence of natural selection on intergenic regions using the predicted chromatin state[33] from the chromatin immunoprecipitation-Seq data produced by the ENCODE Consortium[32]. Among seven categories of predicted chromatin states, the SNVs observed on genomic segments bearing some functionally predicted activity exhibited higher FVRVs than repressed or low-activity regions (Fig. 2e). The difference in FVRV among chromatin states was small, but significant. This indicates weak selection on specific intergenic regions, such as promoters and enhancers, for gene regulation. Furthermore, we observed that the FVRV of microRNAs (miRNAs), but not lincRNAs, was higher, not only than for intergenic regions but also for functionally predicted ENCODE regions.

Notably, the degree of deleterious SNVs predicted by scaled C score in Combined Annotation Dependent Depletion[34], which incorporates several annotations such as conservation metrics and regulatory information, was highly correlated with the FVRV of the 1KJPN variants (Fig. 2f). These observations clearly