XFG.[34] In addition, the titer of *H. pylori* antibody in the aqueous humor might reflect the severity of glaucomatous damage in POAG patients. We hypothesized that some types of chronic infection and/or inflammation can lead to the development of glaucoma especially POAG.

In conclusion, we have identified *TLR4* SNPs as genetic susceptibility alleles for POAG, NTG, and XFG in the Japanese population. Our findings would support the idea that changes in the regulation of TLR signaling in human glaucoma may be associated with innate and adaptive immune responses. Further investigations on different ethnic populations, and on the structure and function of the *TLR4* protein, would be helpful in understanding the pathogenesis of POAG, NTG, and XFG.

ALL AUTHORS HAVE COMPLETED AND SUBMITTED THE ICMJE FORM FOR DISCLOSURE OF POTENTIAL CONFLICTS OF interest and none were reported. This study was supported in part by a Grant-In-Aid for Scientific Research from the Ministry of Education, Science, and Culture of the Japanese Government (NF; C-22591928), by grant from the Ministry of Health, Labor and Welfare of Japan to N.F., and by a grant from the Japan-China Medical Association, Japan, to D.S. Involved in conception and design of the study (Y.T., D.S., Y.M., N.F.); data collection (Y.T., D.S., Ai.S., To.F., N.Y., Ta.F., H.A., H.I., X.Z., At.S., Y.O., K.N.), analysis (Y.T., D.S., N.F.), and interpretation of data (T.Y., T.N., N.F.); and preparation, review, and approval of the manuscript (K.N., T.N., N.F.). The purpose and procedures of the experiment were explained to all patients, and informed consent was obtained. The procedures used conformed to the tenets of the Declaration of Helsinki, and this study was approved by the Tohoku University Institutional Review Board prospectively. Informed consent for participation in this research was obtained from all patients.

Y. Takano and D. Shi contributed equally to this work.

The authors thank Duco Hamasaki, Professor Emeritus, Bascom Palmer Eye Institute, University of Miami, Miami, Florida, for his critical comments and valuable assistance.

REFERENCES

1. Quigley H. Number of people with glaucoma worldwide. *Br J Ophthalmol* 1996;80(5):389–393.
2. Werner EB. Normal-tension glaucoma. In: Ritch R, Shields MB, Krupin T, eds. The Glaucomas, St. Louis: Mosby; 1996:769–797.
3. Shiose Y, Kitazawa Y, Tsukahara S, et al. Epidemiology of glaucoma in Japan—a nationwide glaucoma survey. *Jpn J Ophthalmol* 1991;35(2):133–155.
4. Iwase A, Suzuki Y, Araie M, et al. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology* 2004;111(9):1641–1648.
5. Shields MB. Molecular genetics and pharmacogenomics of the glaucomas. In: Shields MB, ed. Shields Textbook of Glaucoma, 6th ed. Baltimore (MD): Lippincott Williams & Wilkins; 2011:139–148.
6. Stone EM, Fingert JH, Alward WL, et al. Identification of a gene that causes primary open angle glaucoma. *Science* 1997;275(5300):668–670.
7. Rezaie T, Child A, Hitchings R, et al. Adult-onset primary open-angle glaucoma caused by mutations in optineurin. *Science* 2002;295(5557):1077–1079.
8. Monemi S, Spaeth G, DaSilva A, et al. Identification of a novel adult-onset primary open-angle glaucoma (POAG) gene on 5q22.1. *Hum Mol Genet* 2005;14(6):725–733.
9. Wiggs JL, Auguste J, Allingham RR, et al. Lack of association of mutations in optineurin with disease in patients with adult-onset primary open-angle glaucoma. *Arch Ophthalmol* 2003;121(8):1181–1183.
10. Hauser MA, Allingham RR, Linkroum K, et al. Distribution of WDR36 DNA sequence variants in patients with primary open-angle glaucoma. *Invest Ophthalmol Vis Sci* 2006;47(6):2542–2546.
11. Schlotzer-Schrehardt U, Naumann GO. Ocular and systemic pseudoexfoliation syndrome. *Am J Ophthalmol* 2006;141(5):921–937.
12. Thorleifsson G, Magnusson KP, Sulem P, et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* 2007;317(5843):1397–1400.
13. Hayashi H, Gotoh N, Ueda Y, Nakanishi H, Yoshimura N. Lysyl oxidase-like 1 polymorphisms and exfoliation syndrome in the Japanese population. *Am J Ophthalmol* 2008;145(3):582–585.
14. Fuse N, Miyazawa A, Nakazawa T, Mengkegale M, Otomo T, Nishida K. Evaluation of LOXL1 polymorphisms in eyes with exfoliation glaucoma in Japanese. *Mol Vis* 2008;14:1338–1343.
15. Ritch R. Exfoliation syndrome-the most common identifiable cause of open-angle glaucoma. *J Glaucoma* 1994;3(2):176–177.
16. Wax MB. Is there a role for the immune system in glaucomatous optic neuropathy? *Curr Opin Ophthalmol* 2000;11(2):145–150.
17. Wax MB, Tezel G, Saito I, et al. Anti-Ro/SS-A positivity and heat shock protein antibodies in patients with normal-pressure glaucoma. *Am J Ophthalmol* 1998;125(2):145–157.
18. Tezel G, Seigel GM, Wax MB. Autoantibodies to small heat shock proteins in glaucoma. *Invest Ophthalmol Vis Sci* 1998;39(12):2277–2287.
19. Park JS, Svetkauskaite D, He Q, et al. Involvement of toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J Biol Chem* 2004;279(9):7370–7377.
20. Akira S, Takeda K, Kaisho T. Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat Immunol* 2001;2(8):675–680.
21. Shibuya E, Meguro A, Ota M, et al. Association of Toll-like receptor 4 gene polymorphisms with normal tension glaucoma. *Invest Ophthalmol Vis Sci* 2008;49(10):4453–4457.
22. Suh W, Kim S, Ki CS, Kee C. Toll-like receptor 4 gene polymorphisms do not associate with normal tension glaucoma in a Korean population. *Mol Vis* 2011;17:2343–2348.
23. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133–139.

24. Chaudhuri N, Dower SK, Whyte MK, Sabroe I. Toll-like receptors and chronic lung disease. *Clin Sci (Lond)* 2005; 109(2):125–133.
25. Zhang X, Shan P, Jiang G, Cohn L, Lee PJ. Toll-like receptor 4 deficiency causes pulmonary emphysema. *J Clin Invest* 2006;116(11):3050–3059.
26. Tezel G, Hernandez R, Wax MB. Immunostaining of heat shock proteins in the retina and optic nerve head of normal and glaucomatous eyes. *Arch Ophthalmol* 2000; 118(4):511–518.
27. Luo C, Yang X, Kain AD, Powell DW, Kuehn MH, Tezel G. Glaucomatous tissue stress and the regulation of immune response through glial Toll-like receptor signaling. *Invest Ophthalmol Vis Sci* 2010;51(11):5697–5707.
28. Radstake TR, Franke B, Hanssen S, et al. The Toll-like receptor 4 Asp299Gly functional variant is associated with decreased rheumatoid arthritis disease susceptibility but does not influence disease severity and/or outcome. *Arthritis Rheum* 2004;50(3):999–1001.
29. Kuuliala K, Orpana A, Leirisalo-Repo M, et al. Polymorphism at position +896 of the toll-like receptor 4 gene interferes with rapid response to treatment in rheumatoid arthritis. *Ann Rheum Dis* 2006;65(9):1241–1243.
30. Werner M, Topp R, Wimmer K, et al. TLR4 gene variants modify endotoxin effects on asthma. *J Allergy Clin Immunol* 2003;112(2):323–330.
31. Sackesen C, Karaaslan C, Keskin O, et al. The effect of polymorphisms at the CD14 promoter and the TLR4 gene on asthma phenotypes in Turkish children with asthma. *Allergy* 2005;60(12):1485–1492.
32. Danesh J, Collins R, Peto R. Chronic infections and coronary heart disease: is there a link? *Lancet* 1997;350(9075): 430–436.
33. Oshima T, Ozono R, Yano Y, et al. Association of Helicobacter pylori infection with systemic inflammation and endothelial dysfunction in healthy male subjects. *J Am Coll Cardiol* 2005;45(8):1219–1222.
34. Kountouras J, Mylopoulos N, Konstas AG, Zavos C, Chatzopoulos D, Boukla A. Increased levels of Helicobacter pylori IgG antibodies in aqueous humor of patients with primary open-angle and exfoliation glaucoma. *Graefes Arch Clin Exp Ophthalmol* 2003;241(11):884–890.

SUPPLEMENTAL TABLE. Primer Sequences for Toll-like Receptor 4 Gene Amplification Used in This Study

	Forward Primer	Reverse Primer	Annealing Temperature(C)
rs10759930	gtacagggtgtttgggagga	catggaccaatgctcttgtg	63
rs1927914	tgatgaggattgaaaatgtgga	acaaaatggtccctcacagc	60
rs1927911	ttaaatactccatatcattggggagac	gagagcattcagaaattagatgg	62
rs12377632	tggtatttggctttctgttcc	aaggtttctggggcaagttt	56
rs2149356	ccttggatcaagtttagccatt	tttccacaaaactcgctcct	60
rs11536889	ccctgtacccttctcactgc	gtttctgaggaggctggatg	62
rs7037117	ttaacccctttcccacctttc	agagtttgggacctgctcaa	60
rs7045953	ttcccatgttccctcatttc	ggggcaaaagagaaactcct	59

JCM
Journals.ASM.org

Use of Illumina Deep Sequencing Technology To Differentiate Hepatitis C Virus Variants

Masashi Ninomiya,[a] Yoshiyuki Ueno,[a] Ryo Funayama,[b] Takeshi Nagashima,[b] Yuichiro Nishida,[b] Yasuteru Kondo,[a] Jun Inoue,[a] Eiji Kakazu,[a] Osamu Kimura,[a] Keiko Nakayama,[b] and Tooru Shimosegawa[a]

Division of Gastroenterology, Tohoku University of Medicine, Sendai, Japan,[a] and Division of Cell Proliferation, Tohoku University of Medicine, Sendai, Japan[b]

Hepatitis C virus (HCV) is a positive-strand enveloped RNA virus that shows diverse viral populations even in one individual. Though Sanger sequencing has been used to determine viral sequences, deep sequencing technologies are much faster and can perform large-scale sequencing. We demonstrate the successful use of Illumina deep sequencing technology and subsequent analyses to determine the genetic variants and amino acid substitutions in both treatment-naïve (patient 1) and treatment-experienced (patient 7) isolates from HCV-infected patients. As a result, almost the full nucleotide sequence of HCV was detectable for patients 1 and 7. The reads were mapped to the HCV reference sequence. The coverage was 99.8% and the average depth was 69.5× for patient 7, with values of 99.4% (coverage) and 51.1× (average depth) for patient 1. In patient 7, amino acid (aa) 70 in the core region showed arginine, with methionine at aa 91, by Sanger sequencing. Major variants showed the same amino acid sequence, but minor variants were detectable in 18% (6/34 sequences) of sequences, with replacement of methionine by leucine at aa 91. In NS3, 8 amino acid positions showed mixed variants (T72T/I, K213K/R, G237G/S, P264P/S/A, S297S/A, A358A/T, S457S/C, and I615I/M) in patient 7. In patient 1, 3 amino acid positions showed mixed variants (L14L/F/V, S61S/A, and I586T/I). In conclusion, deep sequencing technologies are powerful tools for obtaining more profound insight into the dynamics of variants in the HCV quasispecies in human serum.

Hepatitis C virus (HCV) is a positive-strand enveloped RNA virus of approximately 9,600 nucleotide (nt) bases, consisting of a single open reading frame and two untranslated regions, and belongs to the genus *Hepacivirus* within the family *Flaviviridae* (6). The single open reading frame encodes a polyprotein of 3,011 amino acids (aa) that is cleaved by viral and cellular proteases into 10 different proteins. The three structural proteins, which constitute the viral particle, include the core protein and the envelope glycoproteins E1 and E2. Two regions in E2, known as hypervariable regions 1 and 2, are reported to have extreme sequence variability. The seven nonstructural components include the p7 polypeptide, the NS2-3 protease, the NS3 serine protease and RNA helicase, the NS4A polypeptide, the NA4B and NS5A proteins, and the NS5B RNA-dependent RNA polymerase (RdRp) (29). At both ends of the open reading frame lie the 5′- and 3′-untranslated regions (5′-UTR and 3′-UTR). The nucleotide sequence of the 5′-UTR is relatively well conserved among different HCV genotypes. The HCV 5′-UTR contains an internal ribosome entry site (IRES) that directs the cap-independent initiation of virus translation and forms on four characteristic stem-loop structures (17, 18). HCV displays very high genetic variability both in populations and within infected individuals, where it exists as a cluster of closely related but distinct variants, termed "quasispecies," as occurs in many other RNA viruses with a polymerase enzyme lacking proofreading ability (6, 8, 26).

Current standard treatment of chronic HCV infection is based on the combination of pegylated alpha interferon (peg-IFN-α) and ribavirin (RBV). However, patients with a high load of genotype 1b virus ($>1 \times 10^5$ log IU/ml) do not achieve high sustained virological response (SVR) rates (<50%), even when the most effective combination treatment (IFN plus RBV) is administered for 48 weeks (14, 25). Some investigations concerning therapeutic prediction based on virological features revealed that substitutions of arginine for glutamine at amino acid (aa) 70 and/or leucine for methionine at aa 91 in the core region are independent and significant factors associated with SVR or that patients whose viruses have more than 4 amino acid changes in the NS5A interferon sensitivity-determining region (ISDR) (aa 2209 to 2248) have high responses to IFN therapy compared to those for patients with HCV-J (mutant type), whereas patients whose viruses have no amino acid changes (wild type) or 1 to 3 amino acid changes (intermediate type) have low responses (1, 2, 9, 10).

Recently, direct-acting antiviral (DAA) molecules active on HCV, such as NS3/4A protease inhibitors, nucleoside/nucleotide analogue inhibitors of RdRp, nonnucleoside inhibitors of RdRp, and NS5A inhibitors, have been developed. These DAA molecules, either alone or in combination with peg-IFN plus RBV, were recently described as showing large antiviral effects (15, 21). However, the problem that we have to consider next is viral resistance to DAAs due to the selection of viral variants that contain amino acid substitutions altering the drug target and rendering virus less susceptible to the drug's inhibitory activity (35). Additionally, drug-resistant variants already preexist as minor populations within a patient's quasispecies. Drug exposure intensively inhibits replication of the drug-sensitive viral population, and the resistant variants gradually predominate in the HCV population (7). In the future, to determine the most appropriate treatment for HCV

Received 7 September 2011 **Returned for modification** 10 October 2011
Accepted 9 December 2011
Published ahead of print 28 December 2011
Address correspondence to Yoshiyuki Ueno, yueno@med.tohoku.ac.jp.
Supplemental material for this article may be found at http://jcm.asm.org/.

doi:10.1128/JCM.05715-11

patients, analysis of the nucleotide or amino acid sequence of HCV will become important.

Initial attempts to identify the HCV genome sequence relied on Sanger sequencing and the use of PCR primers targeting relatively conserved regions, methods that would likely fail if the virus had more variants (32–34). In recent years, new technologies have been developed that are able to sequence viruses from environmental samples without using specific primers, cloning, and resorting by recombinant DNA techniques and thus can obtain the sequence information for the complete virome in an unbiased way. Metagenomic approaches such as deep sequencing have proven increasingly successful at identifying variants or mutations of the nucleotide sequence (23, 42, 45).

Here we demonstrate the successful use of Illumina deep sequencing technology and subsequent analyses to determine the genetic variants and amino acid substitutions of both treatment-naïve (patient 1) and treatment-experienced (IFN) (patient 7) isolates of HCV without using specific HCV primers.

MATERIALS AND METHODS

Patients. Two patients with chronic hepatitis C virus infection with genotype 1b virus and one healthy control were enrolled in this study. Each serum sample was collected before treatment with peg-IFN-α and RBV and was stored at −20°C until testing. In the laboratory data, the HCV load was 6.8 log IU/ml for patient 1, who was treatment naïve, and 7.0 log IU/ml for patient 7, who was treated with IFN-α2b and RBV for 6 months in 2002, but with no treatment effect (Cobas TaqMan HCV test; Roche Molecular Systems, Pleasanton, CA). More clinical information is described in Table 1.

Sanger sequencing in the core region and NS5A ISDR. Total RNA was extracted from 100-μl serum samples by use of a MagMAX viral RNA isolation kit (Ambion, Austin, TX), and the RNA preparation thus obtained was subjected to cDNA synthesis with reverse transcriptase (SuperScript III RNase H$^-$ reverse transcriptase; Invitrogen) and to PCR amplification using Prime Star HS DNA polymerase (TaKaRa Bio, Shiga, Japan) with nested primers derived from the core region and the NS5A ISDR of the HCV genome. Nested PCR amplification of the core region of the HCV genome was carried out with primers C008 (sense; 5′-AAC CTC AAA GAA AAA CCA AAC G-3′) and C011 (antisense; 5′-CAT GGG GTA CAT YCC GCT YG-3′) in the first round, for 35 cycles (98°C for 10 s, 55°C for 15 s, and 72°C for 1 min, with an additional 7 min in the last cycle), and with primers C009 (sense; 5′-CCA CAG GAC GTY AAG TTC CC-3′) and C010 (antisense; 5′-AGG GTA TCG ATG ACC TTA CC-3′) in the second round, for 25 cycles. Nested primers derived from the NS5A ISDR of the HCV genome were designed to amplify a 188-bp product, using primers C004 (sense; 5′-ATG CCC ATG CCA GGT TCC AG-3′) and C005 (antisense; 5′-AGC TCC GCC AAG GCA GAA GA-3′) in the first round and primers C006 (sense; 5′-ACC GGA TGT GGC AGT GCT CA-3′) and C007 (antisense; 5′-GTA ATC CGG GCG TGC CCA TA-3′) in the second round. The PCR products were sequenced directly on both strands by use of a BigDye Terminator, version 3.1, cycle sequencing kit on an ABI Prism 3100 genetic analyzer (Applied Biosystems, Foster City, CA). Sequence analysis was performed using Genetyx-Mac ver. 12.2.6 (Genetyx Corp., Tokyo, Japan) and ODEN (version 1.1.1) from the DNA Data Bank of Japan (National Institute of Genetics, Mishima, Japan) (19).

Library preparation and Illumina sequencing. Total RNA was extracted from 800 μl of serum by use of a MagMAX viral RNA isolation kit (Ambion) according to the manufacturer's protocol, with the slight modification that carrier RNA was not included. A library was prepared from approximately 200 ng of total RNA by use of an mRNA-seq sample prep kit (Illumina, San Diego, CA). The quality of the library was evaluated with Bioanalyzer (Agilent, Santa Clara, CA). Before deep sequencing, we confirmed the presence of the HCV genome in the libraries by conducting quantitative PCR with StepOnePlus (Applied Biosystems), using SYBR Ex

TABLE 1 Clinical data for patients enrolled in this study[a]

Patient (storage date of sample)	Sex	Age (yr)	Diagnosis	HCV RNA load (log IU/ml)	HCV genotype	Past treatment (period)	Therapeutic effect	Core aa 70	Core aa 91	No. of aa substitutions in NS5A ISDR
1 (October 2008)	Male	43	Chronic hepatitis C	6.8	1b	None	NA	Wild type	Wild type	0
7 (May 2010)	Male	57	Chronic hepatitis C	7.0	1b	IFN-α2b plus RBV (March–September 2002)	Nontherapeutic	Wild type	Mutant	0
9 (January 2011)	Female	64	Control	NA	NA	NA	NA	NA	NA	NA

[a] Abbreviations: aa, amino acid; IFN, interferon; NA, not applicable; RBV, ribavirin. All patients were negative for hepatitis B surface antigen (HBsAg) and hepatitis B surface antibody (HBsAb).

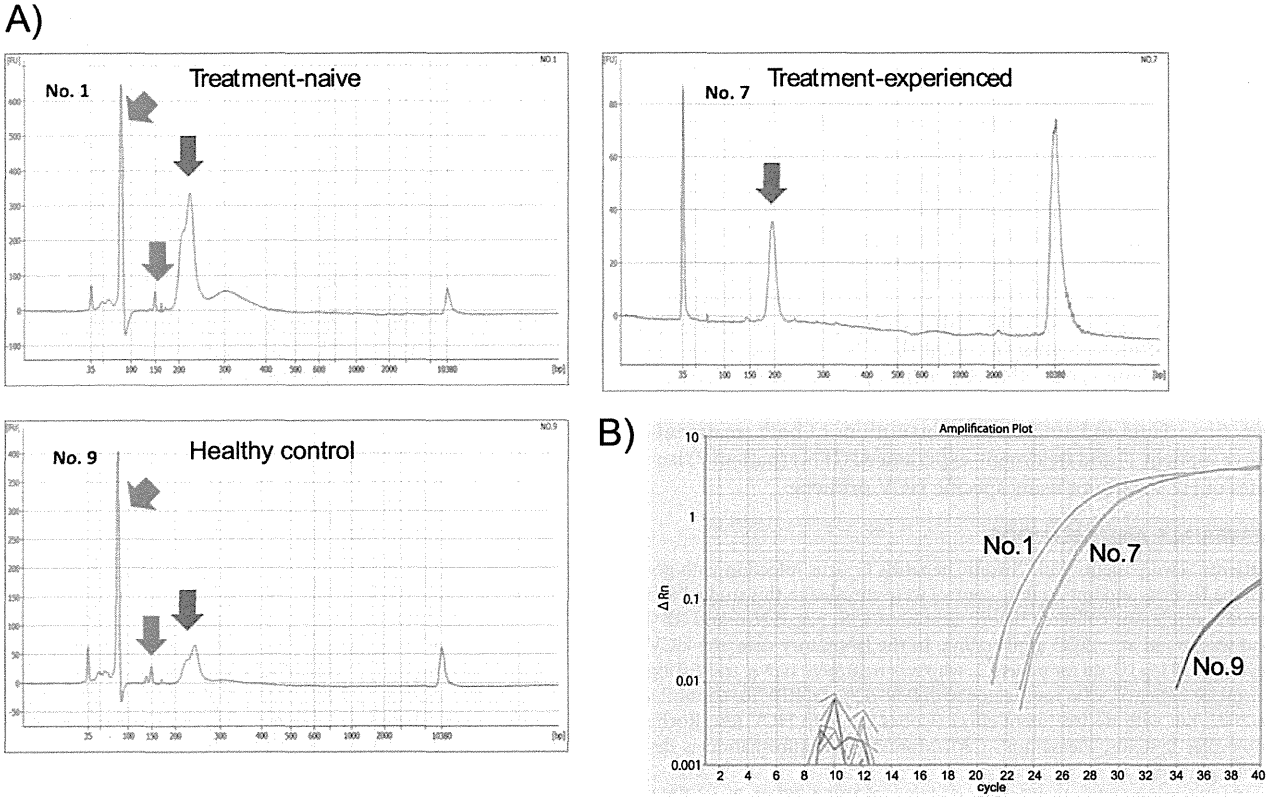

FIG 1 Evaluation of the quality of libraries. (A) The library was well refined for patient 7, but the primer and adaptor dimers were mixed in the libraries of patients 1 and 9, obtained using Bioanalyzer (Agilent). The horizontal axis shows the DNA size, and the vertical axis shows the quantity of DNA. The blue arrows indicate the desired products, and the red arrows indicate the primer or adaptor dimers. The peaks of the wave at 35 bp and 10,380 bp express the marker. (B) Amplification plots for patients 1 (treatment naïve) and 7 (treatment experienced), showing the presence of the HCV genome in the libraries by quantitative PCR with StepOnePlus (Applied Biosystems).

Taq premix (TaKaRa, Shiga, Japan) and the specific primers C112 (sense; 5′-GCW GTS CAR TGG ATG AAC CG-3′) and C113 (antisense; 5′-GCT YTC MGG CAC RTA GTG CG-3′), derived from the 81-bp region encoding HCV NS4B, and then loaded each sample into two or three lanes of a flow cell. Libraries were clonally amplified on the flow cell and sequenced on an Illumina IIx genome analyzer (SCS 2.8 software; Illumina, San Diego, CA), with a 76-mer single end sequence. Image analysis and base calling were performed using RTA 1.8 software.

Analysis. Seventy-six-mer single-end reads were classified by strict bar codes, split into individual reads, and stripped of any remaining primer sequences by using CLC Genomics Workbench (4.6) (http://www.clcbio.co.jp). Sequence reads aligned to the human genome by hg19 (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz), GenBank (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/mrna.fa.gz), RefSeq (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/refMrna.fa.gz), and Ensembl (ftp://ftp.ensembl.org/pub/release62/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.62.cdna.all.fa.gz and ftp://ftp.ensembl.org/pub/release62/fasta/homo_sapiens/ncrna/Homo_sapiens.GRCh37.62.ncrna.fa.gz) were removed in the first mapping analysis of the human genome. Sequence reads not of human origin were aligned with 970 reference HCV sequences registered at the Hepatitis Virus Database server (http://s2as02.genes.nig.ac.jp/index.html) by use of BWA (0.5.9-r16), allowing mismatches within 5 to 10 nucleotide bases (24). The reads could be defined as being of HCV origin by identification with reference to the HCV sequences, allowing mismatches within 10 nucleotide bases. Duplicate reads were completely excluded to avoid sequence bias, using Samtools (0.1.16) (24). Additionally, the variants compared with HCV-J were identified by Samtools. The result of the analysis was displayed using Integrative Genomics Viewer (IGV; 2.0.3) (36).

Ethics statement. Written informed consent was obtained from each individual, and the study for detecting host genomes was approved by the Ethics Committee of the Tohoku University School of Medicine (2010-404).

RESULTS

Evaluation of the quality of the libraries. We conducted deep sequencing analysis for two patients (patient 1 [treatment naïve] and patient 7 [treatment experienced, with IFN]) who had been infected with chronic hepatitis C virus of genotype 1b, as well as one healthy control (patient 9) (Table 1). Since there is only a small quantity of circulating RNAs, including those of viral origin, in serum, it was important to evaluate the quality of the libraries. The library for patient 7 showed good quality using an Agilent bioanalyzer, but the primer and adaptor dimers were mixed in the libraries of patients 1 and 9 (Fig. 1A). Before deep sequencing, we evaluated whether the HCV genome was included in the libraries, and the amplification plots for quantitative PCR showed that the HCV genome could be detected in the libraries from both patients 1 and 7 with the specific primers C112 and C113, derived from the NS4B region (Fig. 1B).

Distribution of free RNA in human serum. To characterize the metagenomics of HCV infection in humans, we analyzed

TABLE 2 Distribution of viral reads in human serum

Sample	No. (%) of reads		
	Patient 1 (treatment naïve)	Patient 7 (treatment experienced)	Patient 9 (healthy control)
Total reads	27,717,487 (100.00)	94,151,356 (100.00)	15,032,130 (100.00)
Adaptor and primer reads	6,502,508 (23.46)	28,605,006 (30.38)	5,713,994 (38.01)
Modified total reads	21,214,979 (76.54)	65,546,350 (69.62)	9,318,136 (61.99)
Reads of human origin	19,761,560 (71.30)	58,446,916 (62.08)	8,660,143 (57.61)
Unknown reads	1,453,419 (5.24)	7,099,434 (7.54)	657,993 (4.38)

the samples by single-end deep sequencing on three lanes for patient 7 and on two lanes for patient 1 and control patient 9, using an Illumina IIx genome analyzer. After trimming the reads to exclude ambiguous nucleotides, primers, or adaptor sequences, 96,079,465 high-quality 76-bp reads were subjected to analysis. From the initial set of reads, a total of 86,868,619 reads were able to be aligned to human genomic DNA.

We then mapped the remaining 9,210,846 reads, including 1,453,419 reads for patient 1, 7,099,434 reads for patient 7, and 657,993 reads for patient 9 (Table 2).

Mapping of the HCV genome sequence. The reads were aligned to 970 HCV genome sequences by using BWA, allowing mismatches within 5 to 10 nucleotide bases. Accordingly, MD5-1 (GenBank accession no. AF165053) was expected to be the closest HCV strain to that in patient 1, and MD2-2 (GenBank accession no. AF165048) was expected to be the closest to that in patient 7. The reads obtained from healthy subject 9 were not aligned to MD5-1 or MD2-2, allowing mismatches within 10 nucleotide bases (see the supplemental material). Whereas some strains, for example, HC-J4, HCV-KT9, and HC-J6, could be mapped to the reads from healthy subject 9, all of the reads were aligned at the 3′-UTR of the U-rich region, and we could not evaluate whether they were of HCV origin. Therefore, we constituted the HCV genome sequences of patients 1 and 7 without the 3′-UTR. In this alignment, the duplicate reads were completely excluded. For patient 1, 6,303 reads were mapped on MD5-1, allowing for 10 mismatched nucleotide bases. The coverage was 99.4%, and the average depth was 51.1× (Fig. 2). For patient 7, 8,583 reads could be identified with MD2-2. The coverage was 99.8%, and the average sequencing depth was 69.5× (Fig. 2).

Notably, the genome sequence could not be obtained for only 8 nt in the 5′-UTR and 52 nt in the core region for patient 1 and for 18 nt in the E2 region for patient 7.

Amino acid substitutions in the core region and the NS5A ISDR. To identify potential mutations at key sites in the genome that mediate the effect of IFN-based therapy, we compared the HCV genome obtained from patient 7 with HCV-J, which is known as the prototypical HCV 1b strain and whose complete genomic sequence has been determined (20). A previous study reported that there were substitutions of aa 70 and/or 91 in the core region and that the number of substitutions within three bases in the region of aa 2209 to 2248 (NS5A ISDR) might be associated with resistance to IFN-based therapy (1, 2, 9, 10). Position 70 in the core region showed arginine, with methionine at aa 91, by Sanger sequencing. In deep sequencing, major variants showed the same amino acid sequence, but minor variants were detectable in 18% (6/34 sequences) of sequences, with replacement of methionine by leucine at aa 91 (Fig. 3A). In the NS5A ISDR, no substitution was indicated for major variants, the same as in direct sequencing by Sanger sequencing, but 16% (14/89 sequences) of sequences showed minor variant replacements of aspartic acid by valine at aa 2220 (Fig. 3B). We validated that more than 10% of the detected variants were effective. For patient 1, core aa 70 (arginine), core aa 91 (leucine), and the number of

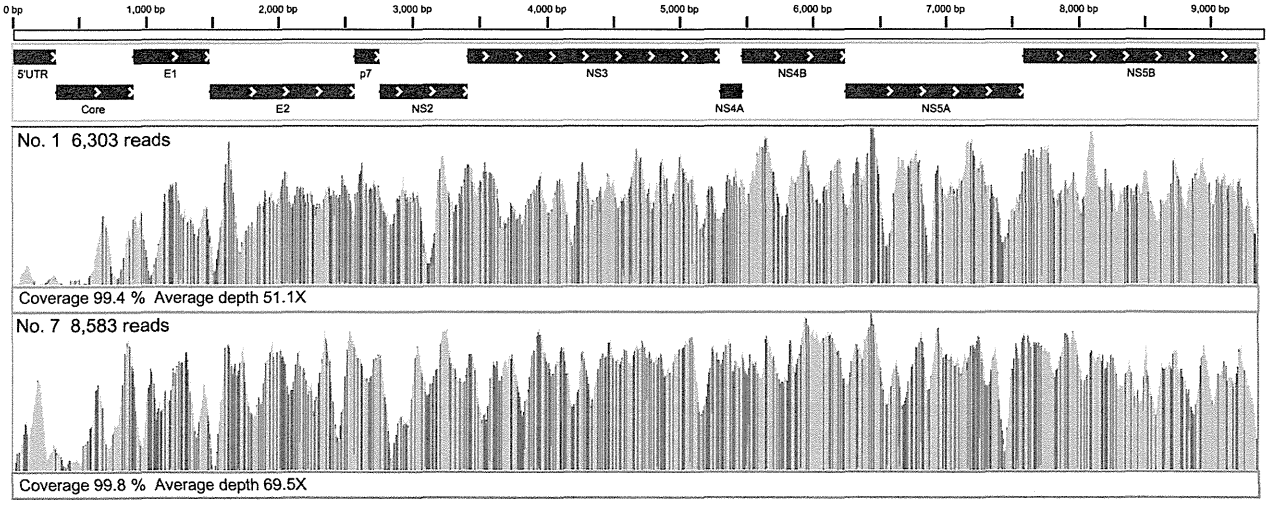

FIG 2 Mapping to the HCV reference genome. For patient 1, 6,303 reads were mapped to MD5-1. The coverage was 99.4%, and the average depth was 51.1×. For patient 7, 8,583 reads were aligned to MD2-2. The coverage was 99.8%, and the average depth was 69.5%.

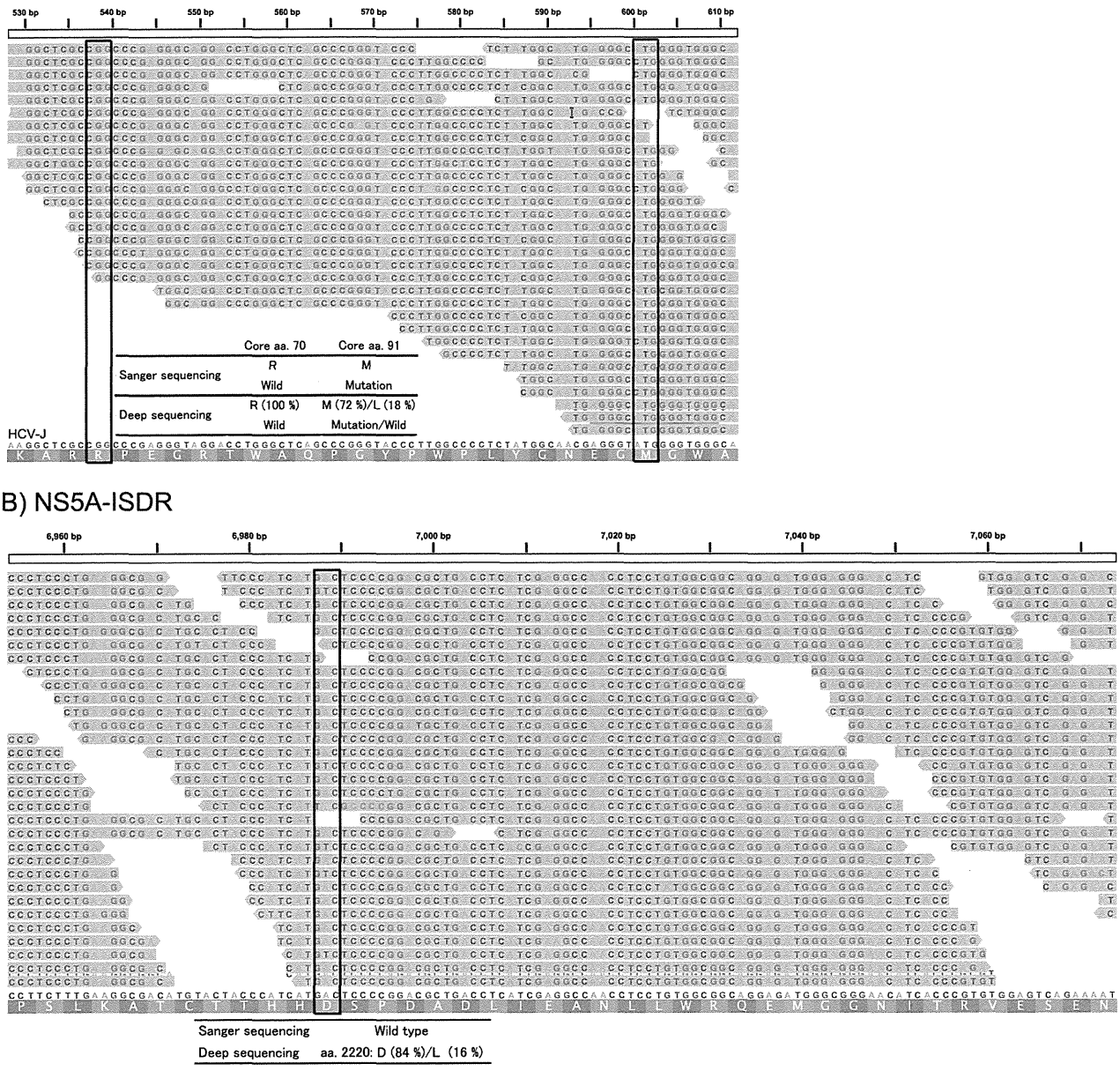

FIG 3 Amino acid substitutions of aa 70 and 91 in the core region (A) and aa 2209 to 2248 of the NS5A ISDR (B). Mutations in these regions were reported to affect the outcome of IFN-based therapies for chronic hepatitis C patients. The lower two lines show the nucleotide sequence and amino acid sequence of HCV-J. Amino acid abbreviations: F, phenylalanine; S, serine; Y, tyrosine; C, cysteine; W, tryptophan; L, leucine; P, proline; H, histidine; Q, glutamine; R, arginine; I, isoleucine; M, methionine; T, threonine; N, asparagine; K, lysine; V, valine; A, alanine; D, aspartic acid; E, glutamic acid; G, glycine.

NS5A ISDR mutations (zero) were the same as those by Sanger sequencing.

Amino acid substitutions in NS3 and NS5B. The recent development of DAA molecules, such as protease inhibitors and polymerase inhibitors, has raised the concern that resistance may weaken the effects of DAA-based therapy (35). It is necessary to obtain the amino acid sequences of NS3 or NS5B variants bearing substitutions which alter the target of the drug. In NS3 in the virus from patient 7, 26 amino acids were changed in comparison with the prototype strain HCV-J. Eight amino acids showed mixed variants, with T72T/I (57 variants [75%]/21 variants [27%]), K213K/R (18 variants [26%]/52 variants [74%]), G237G/S (41 variants [46%]/49 variants [54%]), P264P/S/A (20 variants [42%]/19 variants [40%]/9 variants [19%]), S297S/A (63 variants [81%]/15 variants [19%]), A358A/T (20 variants [21%]/75 variants [79%]), S457S/C (56 variants [64%]/32 variants [36%]), and I615I/M (42 variants [46%]/49 variants [54%]) substitutions (Table 3). For NS5B, the full amino acid sequence was observed, and

TABLE 3 Amino acid substitutions compared with HCV-J in NS3 and NS5B from viruses of patients 1 and 7

Protein and patient	Nucleotide position	Amino acid position	Prototype amino acid	Nucleotide sequence[a]	Amino acid substitution[b]
NS3					
Patient 1	3426	7	S	gCC	A
	3447	14	L	(C/t/g)TT	L/F/V (67 [89]/6 [8]/2 [3])
	3495	30	D	GAg	E
	3513	36	L	gTt	V
	3588	61	S	(T/g)CG	S/A (48 [71]/20 [29])
	3591	62	K	AgG	R
	3618	71	I	gTC	V
	3645	80	Q	CtG	L
	3663	86	P	CaG	Q
	3747	114	V	aTc	I
	3801	132	I	gTC	V
	3855	150	V	GcT	A
	3915	170	I	gT(A/g)	V
	4149	248	I	gTC	V
	4152	249	E	GAc	D
	4191	262	G	aGC	S
	4194	263	G	Gct	A
	4218	271	C	gGC	G
	4302	299	T	tCC	S
	4392	329	I	gTC	V
	4479	358	A	aaC	N
	4551	382	T	tCg	S
	4554	382	G	GcC	A
	4563	386	L	gTC	V
	4659	418	F	TaT	Y
	4758	451	L	gTG	V
	4782	459	A	tCG	S
	4815	470	S	gGg	G
	4935	510	S	aCa	T
	5007	534	S	gGC	G
	5076	557	L	tTC	F
	5163	586	I	A(T/c)A	I/T (4 [12]/30 [88])
	5232	609	V	aTC	I
	5268	621	A	aCA	T
Patient 7	3495	30	D	GAg	E
	3513	36	L	gTt	V
	3531	42	S	aCT	T
	3549	48	V	aTC	I
	3621	72	T	A(C/t)C	T/I (57 [73]/21 [27])
	3663	86	P	CaG	Q
	3672	89	P	tCC	S
	3687	94	M	tTG	L
	3747	114	V	aTc	I
	3801	132	I	gTC	V
	3855	150	V	GcT	A
	3915	170	I	gTA	V
	4044	213	K	A(A/g)g	K/R (18 [26]/52 [74])
	4116	237	G	(G/a)G(C/t)	G/S (41 [46]/49 [54])
	4149	248	I	gTt	V
	4152	249	E	Gac	D
	4194	263	G	Gct	A
	4197	264	P	(C/t/g)CC	P/S/A (20 [42]/19 [40]/9 [19])
	4218	271	C	gGC	G
	4296	297	S	(T/g)CG	S/A (63 [81]/15 [19])
	4302	299	T	tCC	S
	4479	358	A	(G/a)CC	A/T (20 [21]/74 [79])
	4542	379	A	tCA	S
	4551	382	T	tCA	S
	4554	383	G	acC	T
	4563	386	L	aTC	I
	4659	418	F	TaT	Y
	4758	451	L	gTG	V
	4776	457	S	T(C/g)(G/t)	S/C (56 [64]/32 [36])
	4782	459	A	tCg/a	S
	4815	470	S	gcg	A
	5076	557	L	tTC	F
	5172	589	K	AgG	R
	5250	615	I	AT(A/g)	I/M (42 [46]/49 [54])
	5259	618	Y	TtC	F

(Continued on following page)

TABLE 3 (Continued)

Protein and patient	Nucleotide position	Amino acid position	Prototype amino acid	Nucleotide sequence[a]	Amino acid substitution[b]
NS5B					
Patient 1	7659	25	P	gCG	A
	7689	35	S	Aac	N
	7701	39	S	gCC	A
	7725	47	L	CaG	Q
	7827	81	R	Aaa	K
	7839	85	I	gTA	V
	7878	98	K	AgA	R
	7914	110	S	AaC	N
	7932	116	V	aTt	I
	7944	120	R	CaC	H
	7956	124	E	aAG	K
	7989	135	D	aAc	N
	8025	147	V	aTt	I
	8151	189	P	tCC	S
	8205	207	T	gCC	A
	8223	213	C	aaC	N
	8238	218	S	gCA	A
	8289	235	T	gtT	V
	8370	262	V	aTt	I
	8484	300	T	tCT	S
	8532	316	N	tgC	C
	8589	335	A	aaC	N
	8598	338	A	GtC	V
	8784	400	V	GcT	A
	8937	451	C	acT	T
	8976	464	E	cAg	Q
	9153	523	K	Aga	R
	9177	531	K	AgG	R
	9252	556	N	AgC	S
	9276	564	L	gTG	V
	9306	574	L	TgG	W
Patient 7	7599	5	T	tCA	S
	7689	35	S	Aa(c/t)	N
	7701	39	S	gCC	A
	7725	47	L	Caa	Q
	7827	81	R	Aaa	K
	7839	85	I	gTg	V
	7878	98	K	AgA	R
	7914	110	S	AaC	N
	7926	114	R	Aaa	K
	7956	124	E	aAG	K
	8151	189	P	tCC	S
	8205	207	T	gCC	A
	8223	213	C	acC	T
	8238	218	S	gCA	A
	8277	231	N	AgT	S
	8289	235	T	gtT	V
	8298	238	S	gCA	A
	8370	262	V	aTC	I
	8484	300	T	tCT	S
	8532	316	N	tgC	C
	8589	335	A	agC	S
	8784	400	V	GcT	A
	8937	451	C	acT	T
	8940	452	Y	cAC	H
	8946	454	I	gTT	V
	8976	464	E	cAA	Q
	9177	531	K	AgG	R
	9213	543	S	TtC	F
	9231	549	G	aaC	N
	9252	556	N	AgC	S
	9306	574	L	TgG	W

[a] Uppercase letters indicate prototype nucleotides, and lowercase letters indicate mutations.

[b] The numbers and percentages of amino acid bases are displayed in parentheses and brackets, respectively.

31 amino acids were altered in comparison with HCV-J. No mixed variants were seen (Table 3). With patient 1, full amino acid sequences were detected for NS3 and NS5B. Compared with HCV-J, 31 amino acids were altered in NS3, and 3 amino acids showed mixed variants, with L14L/F/V (67 variants [89%]/6 variants [8%]/2 variants [3%]), S61S/A (48 variants [71%]/20 variants [29%]), and I586I/T (4 variants [12%]/30 variants [88%]) substitutions (Table 3). In NS5B, 31 amino acids were converted (Table 3). Note that more than 10% of the minor variants were confirmed as effective.

DISCUSSION

In this study, we attempted to detect the HCV genome directly in human serum without using specific primers and succeeded in determining and certifying nearly the full genome sequence and a high genetic diversity by using deep Illumina sequencing. HCV has already been reported to be a highly variable virus with a quasispecies distribution, large viral populations, and very rapid turnover in individual patients (6, 26). Previous studies using metagenomic sequencing of other viruses from human clinical samples mostly employed pyrosequencing (11, 12, 23, 30, 46). The longer reads from pyrosequencing (250 to 450 bp) facilitate the assembly of individual reads into contigs, which facilitates the classification of the sequence data by homology-based BLAST alignment. In contrast to metagenomic analysis using pyrosequencing, Illumina short-read sequencing enables a greater depth (by an order of magnitude) that is reflected in a very low detection limit. A recent report revealed that viral transcripts could be found at frequencies of <1 in 1,000,000 (28). However, because of short reads, *de novo* assembly without any reference is difficult to conduct, so it is not suitable for discovering an unknown viral genome. However, it seems quite useful for resequencing or detection of variants of known viruses for which abundant nucleotide sequence data have already been reported.

We defined the Illumina 76-mer reads as being of HCV origin by relying on the 970 HCV genome sequences in this study. HCV shows considerable genetic diversity and has been classified into 11 genogroups or 6 groups by the core, E1, or NS5B region, with nucleotide divergence. Only about 75% similarity was shown in the variable region, even for the same group (38, 39, 44). The reads were aligned to each of the 970 HCV sequences, allowing mismatches within 10 nucleotide bases. Under these conditions, we could not map the reads of HCV isolates without the 3′-UTR region for patient 9, who had not been infected with HCV. This is because the 3′-UTR has a U-rich region, and it is impossible to decide the reads of HCV origin with specificity. Therefore, we aligned the reads to the full HCV sequence without the 3′-UTR for patients 1 and 7.

Many variants with different nucleotide bases, known as quasispecies, were detected in both patients 1 and 7. Taking a close look at the mixed variants with amino acid substitutions in NS3, patient 7 showed 8 variants, while there were only 3 variants in patient 1. Recent studies reported that HCV genomic sequences in treatment relapsers displayed significantly more mutations than those in nonresponders. HCV sequence analysis of a 4-year post-antiviral-therapy follow-up revealed that the vast majority of mutations selected during the therapy phase were maintained in the relapsers, while very few new mutations arose during the 4-year posttherapy span (5, 47). Based on the experiments mentioned above, treatment with IFN may lead to the emergence of mutations. Deep sequencing is considered a useful tool for detecting viral variants and determining the mutational rate without cloning.

Although there were only a few detected reads of HCV origin obtained from serum, metagenomic analysis could be conducted with the enormous data sets generated by deep sequencing. Consequently, almost the full genome sequence of HCV was demonstrated by using computational analysis of sequential alignments of individual reads, with average depths of 51.1× and 69.5×. However, the regions for which we could not obtain the sequence were 8 nt in the 5′-UTR and 52 nt in the core region for patient 1 and 18 nt in the E2 region for patient 7. Since the 5′-UTR forms on characteristic stem-loop structures, Sanger sequencing is generally difficult (31). Similarly, even deep Illumina sequencing appears to be difficult. Since the E2 region, known as a hypervariable region, is reported to have extreme sequence variability, it was predicted that there would be too many mismatches with the reference HCV genome and that the reads could not be mapped by this analysis. Analysis of this hypervariable region will require further work.

Comparing the qualities of the libraries from patients 1 and 7, that of patient 7 was well refined; hence, the quality of the library is important for gaining large amounts of expected reads. However, a lot of duplicate reads were found for patient 7, and in fact, to analyze the full sequence of HCV, it is considered sufficient to use two lanes for Illumina sequencing.

Amino acid substitutions of core aa 70 and 91 and within the NS5A ISDR, as well as genetic polymorphisms in the host IL28B gene, encoding IFN-λ-3 on chromosome 19, affect the outcome of interferon-based therapies for chronic hepatitis C patients (1, 2, 9, 10, 16, 40, 43). Even if only the amino acid substitution of a major variant were assumed, an accurate therapeutic effect would be impossible to predict. The proportion of minor variants may change the therapeutic effect, and variants cannot be detected only by direct sequencing using Sanger sequencing. In fact, patient 7 showed a methionine at aa 91 in the core region and no substitution in the NS5A ISDR by direct Sanger sequencing, but deep sequencing indicated minor variants (at aa 91 in the core region [18% of sequences] and in the NS5A ISDR [15% of sequences]).

Recently, DAA molecules have been developed for HCV therapies, and these drugs may lead to the selection of resistant viruses if administered alone. The first-generation NS3/4A inhibitors are telaprevir and boceprevir, and most of the reported clinical data on drug resistance were obtained from patients treated with telaprevir. As an illustration, based on *in vitro* studies, telaprevir resistance related to amino acid substitutions V36A/M/C, T54A/S, R155K/T/Q, A155V/T, and A156T has been reported (4, 22, 37). Substitutions that generated boceprevir resistance included those detected in a patient treated with telaprevir, plus V170A/T and V55A substitutions (13, 41). Of the nonnucleoside inhibitors of RdRp displaying inhibitory activities against the RdRp enzyme at NS5B, few have been reported in the *in vivo* resistance data (27). This is because studies of antiviral efficacy are generally limited to 3 to 5 days. Yet, for example, the S282T substitution has been reported to confer a loss of *in vitro* sensitivity to nucleoside/nucleotide analogue inhibitors (3). In the future, the triple combination of peg-IFN-α, RBV, and a protease inhibitor or several combina-

tions of DAAs will soon become the standard therapy for treatment-naïve and treatment-experienced patients with HCV genotype 1. In this study, though neither patient 1 nor patient 7 showed drug-resistant variants, it would be very important for the selection of therapy to identify resistant or minor variants prior to treatment. Additionally, when treatment has failed, it is necessary to consider viral factors as a cause.

In conclusion, deep sequencing technologies are a powerful tool for obtaining more profound insight into the dynamics of variants in the HCV quasispecies in human serum. Although the cost of deep sequencing is still much greater than the reagent costs for Sanger sequencing, it is still attractive in clinical medicine because deep sequencing is able to generate much more information on the viral genome sequences in internal organs. The cost will decrease in the future as the technology of deep sequencing develops. As DAA combination treatment of HCV infection is developed, obtaining sequence information on variants in individual cases by use of deep sequencing will be feasible for determining optimal antiviral treatment.

ACKNOWLEDGMENTS

We thank M. Tsuda, N. Koshita, and K. Kuroda for technical assistance. We also acknowledge the support of the Biomedical Research Core of the Tohoku University Graduate School of Medicine.

This study was supported in part by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (assignment no. 22790627) and by grants from the Ministry of Health, Labor, and Welfare of Japan.

REFERENCES

1. **Akuta N, et al.** 2007. Prediction of response to pegylated interferon and ribavirin in hepatitis C by polymorphisms in the viral core protein and very early dynamics of viremia. Intervirology **50**:361–368.
2. **Akuta N, et al.** 2005. Association of amino acid substitution pattern in core protein of hepatitis C virus genotype 1b high viral load and non-virological response to interferon-ribavirin combination therapy. Intervirology **48**:372–380.
3. **Ali S, et al.** 2008. Selected replicon variants with low-level in vitro resistance to the hepatitis C virus NS5B polymerase inhibitor PSI-6130 lack cross-resistance with R1479. Antimicrob. Agents Chemother. **52**:4356–4369.
4. **Barbotte L, et al.** 2010. Characterization of V36C, a novel amino acid substitution conferring hepatitis C virus (HCV) resistance to telaprevir, a potent peptidomimetic inhibitor of HCV protease. Antimicrob. Agents Chemother. **54**:2681–2683.
5. **Cannon NA, Donlin MJ, Fan X, Aurora R, Tavis JE.** 2008. Hepatitis C virus diversity and evolution in the full open-reading frame during antiviral therapy. PLoS One **3**:e2123.
6. **Choo QL, et al.** 1991. Genetic organization and diversity of the hepatitis C virus. Proc. Natl. Acad. Sci. U. S. A. **88**:2451–2455.
7. **Clavel F, Hance AJ.** 2004. HIV drug resistance. N. Engl. J. Med. **350**:1023–1035.
8. **Domingo E, et al.** 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene **40**:1–8.
9. **Enomoto N, et al.** 1995. Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. J. Clin. Invest. **96**:224–230.
10. **Enomoto N, et al.** 1996. Mutations in the nonstructural protein 5A gene and response to interferon in patients with chronic hepatitis C virus 1b infection. N. Engl. J. Med. **334**:77–81.
11. **Feng H, Shuda M, Chang Y, Moore PS.** 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science **319**:1096–1100.
12. **Finkbeiner SR, et al.** 2009. Identification of a novel astrovirus (astrovirus VA1) associated with an outbreak of acute gastroenteritis. J. Virol. **83**:10836–10839.
13. **Flint M, et al.** 2009. Selection and characterization of hepatitis C virus replicons dually resistant to the polymerase and protease inhibitors HCV-796 and boceprevir (SCH 503034). Antimicrob. Agents Chemother. **53**:401–411.
14. **Fried MW, et al.** 2002. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N. Engl. J. Med. **347**:975–982.
15. **Gao M, et al.** 2010. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. Nature **465**:96–100.
16. **Ge D, et al.** 2009. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. Nature **461**:399–401.
17. **Honda M, Beard MR, Ping LH, Lemon SM.** 1999. A phylogenetically conserved stem-loop structure at the 5′ border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. J. Virol. **73**:1165–1174.
18. **Honda M, et al.** 1996. Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. Virology **222**:31–42.
19. **Ina Y.** 1994. ODEN: a program package for molecular evolutionary analysis and database search of DNA and amino acid sequences. Comput. Appl. Biosci. **10**:11–12.
20. **Kato N, et al.** 1990. Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc. Natl. Acad. Sci. U. S. A. **87**:9524–9528.
21. **Kieffer TL, et al.** 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Hepatology **46**:631–639.
22. **Kwong AD, McNair L, Jacobson I, George S.** 2008. Recent progress in the development of selected hepatitis C virus NS3.4A protease and NS5B polymerase inhibitors. Curr. Opin. Pharmacol. **8**:522–531.
23. **Lataillade M, et al.** 2010. Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. PLoS One **5**:e10952.
24. **Li H, et al.** 2009. The sequence alignment/map format and SAMtools. Bioinformatics **15**:2078–2079.
25. **Manns MP, et al.** 2001. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial. Lancet **358**:958–965.
26. **Martell M, et al.** 1992. Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. J. Virol. **66**:3225–3229.
27. **McCown MF, et al.** 2008. The hepatitis C virus replicon presents a higher barrier to resistance to nucleoside analogs than to nonnucleoside polymerase or protease inhibitors. Antimicrob. Agents Chemother. **52**:1604–1612.
28. **Moore RA, et al.** 2011. The sensitivity of massively parallel sequencing for detecting candidate infectious agents associated with human tissue. PLoS One **6**:e19838.
29. **Moradpour D, Penin F, Rice CM.** 2007. Replication of hepatitis C virus. Nat. Rev. Microbiol. **5**:453–463.
30. **Nakamura S, et al.** 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One **4**:e4219.
31. **Ninomiya M, Takahashi M, Shimosegawa T, Okamoto H.** 2007. Analysis of the entire genomes of fifteen torque teno midi virus variants classifiable into a third group of genus Anellovirus. Arch. Virol. **152**:1961–1975.
32. **Okamoto H, et al.** 1992. Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. Virology **190**:894–899.
33. **Okamoto H, et al.** 1992. Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology **188**:331–341.
34. **Okamoto H, et al.** 1993. Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolates and PCR primers for specific detection. J. Gen. Virol. **74**:2385–2390.
35. **Pawlotsky JM.** 2011. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. Hepatology **53**:1742–1751.
36. **Robinson JT, et al.** 2011. Integrative genomics viewer. Nat. Biotechnol. **29**:24–26.
37. **Sarrazin C, et al.** 2007. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. Gastroenterology **132**:1767–1777.
38. **Simmonds P, et al.** 1994. Identification of genotypes of hepatitis C virus

by sequence comparisons in the core, E1 and NS-5 regions. J. Gen. Virol. 75:1053–1061.
39. **Smith DB, et al.** 1997. The origin of hepatitis C virus genotypes. J. Gen. Virol. 78:321–328.
40. **Suppiah V, et al.** 2009. IL28B is associated with response to chronic hepatitis C interferon-α and ribavirin therapy. Nat. Genet. 41:1100–1104.
41. **Susser S, et al.** 2009. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. Hepatology 50: 1709–1718.
42. **Szpara ML, Parsons L, Enquist LW.** 2010. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. J. Virol. 84:5303–5313.
43. **Tanaka Y, et al.** 2009. Genome-wide association of IL28B with response to pegylated interferon-α and ribavirin therapy for chronic hepatitis C. Nat. Genet. 41:1105–1109.
44. **Tokita H, et al.** 1996. Hepatitis C virus variants from Jakarta, Indonesia classifiable into novel genotypes in the second (2e and 2f), tenth (10a) and eleventh (11a) genetic groups. J. Gen. Virol. 77:293–301.
45. **Verbinnen T, et al.** 2010. Tracking the evolution of multiple in vitro hepatitis C virus replicon variants under protease inhibitor selection pressure by 454 deep sequencing. J. Virol. 84:11124–11133.
46. **Victoria JG, et al.** 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J. Virol. 83:4642–4651.
47. **Xiang X, et al.** 2011. Viral sequence evolution in Chinese genotype 1b chronic hepatitis C patients experiencing unsuccessful interferon treatment. Infect. Genet. Evol. 11:382–390.

Gain-of-Function Mutations in *RIT1* Cause Noonan Syndrome, a RAS/MAPK Pathway Syndrome

Yoko Aoki,[1,*] Tetsuya Niihori,[1] Toshihiro Banjo,[2] Nobuhiko Okamoto,[3] Seiji Mizuno,[4] Kenji Kurosawa,[5] Tsutomu Ogata,[6] Fumio Takada,[7] Michihiro Yano,[8] Toru Ando,[9] Tadataka Hoshika,[10] Christopher Barnett,[11,12] Hirofumi Ohashi,[13] Hiroshi Kawame,[14] Tomonobu Hasegawa,[15] Takahiro Okutani,[16] Tatsuo Nagashima,[17] Satoshi Hasegawa,[18] Ryo Funayama,[19] Takeshi Nagashima,[19] Keiko Nakayama,[19] Shin-ichi Inoue,[1] Yusuke Watanabe,[2] Toshihiko Ogura,[2] and Yoichi Matsubara[1,20]

RAS GTPases mediate a wide variety of cellular functions, including cell proliferation, survival, and differentiation. Recent studies have revealed that germline mutations and mosaicism for classical RAS mutations, including those in *HRAS*, *KRAS*, and *NRAS*, cause a wide spectrum of genetic disorders. These include Noonan syndrome and related disorders (RAS/mitogen-activated protein kinase [RAS/MAPK] pathway syndromes, or RASopathies), nevus sebaceous, and Schimmelpenning syndrome. In the present study, we identified a total of nine missense, nonsynonymous mutations in *RIT1*, encoding a member of the RAS subfamily, in 17 of 180 individuals (9%) with Noonan syndrome or a related condition but with no detectable mutations in known Noonan-related genes. Clinical manifestations in the *RIT1*-mutation-positive individuals are consistent with those of Noonan syndrome, which is characterized by distinctive facial features, short stature, and congenital heart defects. Seventy percent of mutation-positive individuals presented with hypertrophic cardiomyopathy; this frequency is high relative to the overall 20% incidence in individuals with Noonan syndrome. Luciferase assays in NIH 3T3 cells showed that five RIT1 alterations identified in children with Noonan syndrome enhanced ELK1 transactivation. The introduction of mRNAs of mutant *RIT1* into 1-cell-stage zebrafish embryos was found to result in a significant increase of embryos with craniofacial abnormalities, incomplete looping, a hypoplastic chamber in the heart, and an elongated yolk sac. These results demonstrate that gain-of-function mutations in *RIT1* cause Noonan syndrome and show a similar biological effect to mutations in other RASopathy-related genes.

RAS GTPases are monomeric G proteins with a molecular mass of 20–40 kDa and cycle between a GTP-bound active and a GDP-bound inactive state. The members of the RAS superfamily are structurally classified into at least five subfamilies: RAS, Rho, Rab, Sar1/Arf, and Ran families.[1,2] The Ras subfamily consists of classical RAS proteins (HRAS, KRAS, and NRAS), RRAS, RRAS2 (TC21), RRAS3 (MRAS), RAPs, RAEB, RALs, RIT1, and RIT2 (RIN). RAS proteins interact with multiple effectors, including RAF kinases, phosphatidylinositol 3-kinase (PI-3 kinase), RalGDS, p120GAP, MEKK1, RIN1, AF-6, phospholipase C epsilon, and the Nore-MST1 complex, and activate multiple downstream signaling cascades.[1,2] Of these signaling pathways, the RAS/mitogen-activated protein kinase (RAS/MAPK) signaling pathway plays a central role in cellular proliferation and differentiation.

Noonan syndrome (MIM 163950) is an autosomal-dominant disorder characterized by short stature, distinctive facial features, and congenital heart defects.[3,4] The distinctive facial features include hypertelorism, downslanting palpebral fissures, ptosis, a webbed or short neck, and low-set, posteriorly rotated ears. Congenital heart defects, including pulmonary valve stenosis and atrial septal defects, occur in 50%–80% of individuals. Hypertrophic cardiomyopathy is observed in 20% of affected individuals. Other clinical manifestations include cryptorchidism, mild intellectual disability, bleeding tendency, and hydrops fetalis. The incidence of this syndrome is estimated to be between 1 in 1,000 to 1 in 2,500 live births. Individuals with Noonan syndrome are at risk of juvenile myelomonocytic leukemia (JMML), a myeloproliferative disorder characterized by excessive production of myelomonocytic cells.[4] Noonan syndrome exhibits phenotypic overlap

[1]Department of Medical Genetics, Tohoku University School of Medicine, Sendai 980-8574, Japan; [2]Department of Developmental Neurobiology, Institute of Development, Aging, and Cancer, Tohoku University, Sendai 980-8575, Japan; [3]Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi 594-1101, Japan; [4]Department of Pediatrics, Central Hospital, Aichi Human Service Center, Kasugai 480-0392, Japan; [5]Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama 232-8555, Japan; [6]Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu 431-3192, Japan; [7]Department of Medical Genetics, Kitasato University Graduate School of Medical Sciences, Sagamihara 252-0373, Japan; [8]Department of Pediatrics, Akita University School of Medicine, Akita 010-8543, Japan; [9]Department of Pediatrics, Municipal Tsuruga Hospital, Tsuruga 914-8502, Japan; [10]Department of Pediatrics, Tottori Prefectural Central Hospital, Tottori 680-0901, Japan; [11]South Australian Clinical Genetics Service, SA Pathology, Women's and Children's Hospital, North Adelaide, Adelaide, SA 5006, Australia; [12]School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, SA 5005, Australia; [13]Division of Medical Genetics, Saitama Children's Medical Center, Saitama 339-8551, Japan; [14]Department of Genetic Counseling, Ochanomizu University, Tokyo 112-8610, Japan; [15]Department of Pediatrics Keio University School of Medicine, Tokyo 160-8582, Japan; [16]Division of Neonatal Intensive Care Unit, General Perinatal Medical Center, Wakayama Medical University, Wakayama 641-8510, Japan; [17]Department of Pediatrics, Jikei University School of Medicine, Tokyo 105-8461, Japan; [18]Department of Pediatrics, Niigata Graduate School of Medical and Dental Sciences, Niigata 951-8510, Japan; [19]Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan; [20]National Research Institute for Child Health and Development, Tokyo 157-8535, Japan
*Correspondence: aokiy@med.tohoku.ac.jp
http://dx.doi.org/10.1016/j.ajhg.2013.05.021.

with Costello syndrome (MIM 218040) and cardiofaciocutaneous (CFC) syndrome (MIM 115150).

In 2001, Tartaglia et al. identified missense mutations in protein-tyrosine phosphatase, nonreceptor type 11 (*PTPN11* [MIM 176876]), which encodes the tyrosine phosphatase SHP-2 in 50% of individuals with Noonan syndrome.[5] In contrast, loss-of-function or dominant-negative mutations in *PTPN11* have been reported in individuals with Noonan syndrome with multiple lentigines[6] (formerly referred to as LEOPARD [multiple lentigines, electrocardiographic conduction abnormalities, ocular hypertelorism, pulmonic stenosis, abnormal genitalia, retardation of growth, and sensorineural deafness] syndrome [MIM 151100]). To date, germline mutations in *PTPN11*, *KRAS* (MIM 190070), *SOS1* (MIM 182530), *RAF1* (MIM 164760), and *NRAS* (MIM 164790) have been identified in individuals with Noonan syndrome[7–12] (NS1 [MIM 163950], NS3 [MIM 609942], NS4 [MIM 610733], NS5 [MIM 611553], and NS6 [MIM 613224]), and mutations in *SHOC2* (MIM 602775) and *CBL* (MIM 165360) have been identified in two Noonan-syndrome-like syndromes[13–16] (NSLH [MIM 607721] and NSLL [MIM 613563], respectively) (Figure S1, available online). Moreover, we and another group have identified germline mutations in *HRAS* (MIM 190020) in individuals with Costello syndrome[17] and germline mutations in *KRAS*, *BRAF* (MIM 164757), *MAP2K1* (MIM 176872), and *MAP2K2* (MIM 601263) in individuals with CFC syndrome.[18,19] Mutations in *BRAF* have been also identified in a small percentage of individuals with Noonan syndrome (NS7 [MIM 613706]). A line of studies have shown that a group of the above genetic disorders result from dysregulation of the RAS and downstream signaling cascade (RAS/MAPK pathway syndromes, or RASopathies).[20,21] Recently, mosaicism for *KRAS* and *HRAS* mutations has been reported in nevus sebaceous and Schimmelpenning syndrome,[22] further extending a spectrum of diseases with a dysregulated RAS/MAPK pathway.

To identify genetic causes of Noonan syndrome, we recruited 180 individuals with Noonan syndrome or a related phenotype; they were negative for all coding exons in *PTPN11*, *KRAS*, *HRAS*, and *SOS1*; exons 6 and 11–16 in *BRAF*; exons 7, 14, and 17 in *RAF1*; exons 2 and 3 in *MAP2K1* and *MAP2K2*; and exon 1 in *SHOC2*. Further genetic analysis has been conducted according to their first diagnoses.[17,23–29] This study was approved by the ethics committee of Tohoku University School of Medicine. We obtained informed consent from all subjects involved in the study. We sequenced the exomes of 14 individuals whose clinical manifestations had been confirmed to be consistent with Noonan syndrome by trained dysmorphologists. Targeted enrichment was performed with the Agilent SureSelect Human All Exon v.1 Kit for four individuals and with the SureSelect Human All Exon 50Mb kit for ten individuals. Exon-enriched DNA libraries from these 14 individuals were sequenced on the Illumina Hiseq 2000 for 91 bp (v.1 kit) or 101 bp (50Mb kit). The Burrows-Wheeler Aligner (BWA) was used to align the sequence reads to the human genome (UCSC Genome Browser hg19);[30] all BWA parameters were kept at the default settings. After the removal of duplicates from the alignments, realignment around known indels, recalibration, and SNP and indel calling were performed with the Genome Analysis Toolkit (v.1.5).[31] ANNOVAR was used for annotation against the RefSeq database and dbSNP.[32] We identified approximately 10,000 nonsynonymous, nonsense, and splice-site variations and coding indels per individual (Table S1). Filtering steps using variant databases (dbSNP132 and the 1000 Genome Project database) and in-house exome data were carried out, resulting in the identification of 122–282 variants per individual. By visual inspection of the generated data, four heterozygous *RIT1* (MIM 609591; RefSeq accession number NM_006912.5) variants (c.246T>G [p.Phe82Leu], c.265T>C [p.Tyr89His], c.270G>T [p.Met90Ile], and c.284G>C [p.Gly95Ala]) were found in four individuals. Sanger sequencing validated the heterozygous state of the four variants. We did not find any other strong candidate genes in the results of exome sequencing.

RIT1 shares approximately 50% sequence identity with RAS, has an additional N-terminal extension, and does not possess a C-terminal CAAX motif, a specific motif for post-translational modification.[33,34] *RIT1* is located in chromosomal region 1q22 and consists of six exons. We analyzed an additional 166 individuals diagnosed with Noonan syndrome or a related disorder but without mutations in known genes.[17,23–29] Sanger sequencing of all coding exons in *RIT1* in the 166 individuals showed that 13 in 166 individuals had changes. Combining with the 4 in 14 individuals from exome sequencing, a total of nine missense, nonsynonymous mutations were identified in 17 of 180 (9%) individuals who were suspected to have Noonan syndrome or a related disorder (Table 1 and Figures 1A–1L). The identified germline *RIT1* mutations encode alterations located in the G1 domain (c.104G>C [p.Ser35Thr]); the switch I region, involving the G2 domain (c.170C>G [p.Ala57Gly]); and the switch II region, corresponding to RAS (c.242A>G [p.Glu81Gly], c.244T>G [p.Phe82Val], c.246T>G [p.Phe82Leu], c.247A>C [p.Thr83Pro], c.265T>C [p.Tyr89His], c.270G>T [p.Met90Ile], and c.284G>C [p.Gly95Ala]) (Figure S2). Amino acids where alterations are located are conserved among species (Figure S3). The *RIT1* mutations encode alterations clustered in the switch II region. In contrast, *HRAS* germline mutations identified in Costello syndrome are clustered at codon 12 and 13 in the region encoding the G1 domain[17] (Figure 1M). Mutations in parents were not identified in seven families. These mutations are apparently de novo, but biologic confirmation of parentage was not performed. One mutation, c.104G>C, was inherited from a mother with a Noonan syndrome phenotype (Table 1). None of these mutations were identified in 480 controls.

To assess the functional consequences of *RIT1* mutations identified in affected individuals, we introduced a

Table 1. Mutations in *RIT1*, Family Status, and Heart Defects of Mutation-Positive Individuals

Subject	Exon	Nucleotide Change[a]	Amino Acid Change[b]	Father	Mother	HCM[c]	PS[c]	Other Heart Defects[c]
NS414	2	c.104G>C	p.Ser35Thr	WT	p.Ser35Thr	+	–	MVP, MR
KCC27	2	c.104G>C	p.Ser35Thr	NA	NA	+	+	–
NS43	4	c.170C>G	p.Ala57Gly	NA	NA	+	–	MR, TR
NS185	4	c.170C>G	p.Ala57Gly	NA	NA	+	+	ASD, PDA
NS216	4	c.170C>G	p.Ala57Gly	NA	NA	+	–	–
NS402	4	c.170C>G	p.Ala57Gly	WT	WT	+	+	–
NS168	5	c.242A>G	p.Glu81Gly	NA	NA	–	+	VSD
NS410	5	c.244T>G	p.Phe82Val	WT	WT	+	–	–
NS358	5	c.246T>G	p.Phe82Leu	WT	WT	–	+	ASD
NS465	5	c.246T>G	p.Phe82Leu	NA	NA	–	+	VSD
NS276	5	c.247A>C	p.Thr83Pro	WT	WT	+	+	PVC
KCC8	5	c.265T>C	p.Tyr89His	NA	NA	+	+	–
KCC38	5	c.270G>T	p.Met90Ile	WT	WT	+	+	ASD, VSD, PDA
NS234	5	c.284G>C	p.Gly95Ala	WT	WT	–	–	ASD
NS265	5	c.284G>C	p.Gly95Ala	WT	WT	+	+	–
Og22	5	c.284G>C	p.Gly95Ala	NA	NA	–	–	–
Og45	5	c.284G>C	p.Gly95Ala	NA	NA	+	+	ASD

PCR primers used for sequencing are shown in Table S3. Nucleotide changes are not located in CpG dinucleotides, suggesting that they exhibit baseline mutation rates with a phenotypic filtering effect and that only these mutations lead to this phenotype. Abbreviations are as follows: WT, wild-type; HCM, hypertrophic cardiomyopathy; PS, pulmonic stenosis; MVP, mitral valve prolapse; MR, mitral regurgitation; TR, tricuspid regurgitation; ASD, atrial septal defect; PDA, patent ductus arteriosus; VSD, ventricular septal defect; PVC, premature ventricular contraction; and NA, not available.
[a]RefSeq NM_006912.5.
[b]RefSeq NP_008843.1.
[c]HCM and heart anomalies were diagnosed by echocardiography.

single-base substitution (p.Ser35Thr, p.Ala57Gly, p.Glu81Gly, p.Phe82Leu, or p.Gly95Ala) identified in individuals with Noonan syndrome into a pCAGGS expression vector[36] harboring *RIT1* cDNA. As an experimental control, cDNAs harboring *RIT1* c.89G>T (p.Gly30Val), c.104G>C (p.Ser35Asn), and c.236A>T (p.Gln79Leu) and *Braf* c.1910T>A (p.Val637Glu) (RefSeq NM_139294), which corresponds to oncogenic p.Val600Glu in humans, were also generated. RIT1 p.Gly30Val and p.Gln79Leu correspond to oncogenic RAS alterations p.Gly12Val and p.Gln61Leu, respectively. We introduced pFR-luc, pFA2-Elk1, phRLnull-luc, and wild-type (WT) or mutant expression constructs of *RIT1* into NIH 3T3 cells to examine the transcriptional activation by ELK1,[18,33] a transcription factor that is activated by MAPK. The results revealed that compared with the WT cDNA, all *RIT1* mutations exhibited significant activation. RIT1 p.Gln79Leu, followed by p.Gly95Ala, p.Ala57Gly, p.Phe82Leu, and p.Glu81Gly, showed the highest ELK1 transactivation, as also shown in a past study[37] (Figure 2A). The c.104G>C (p.Ser35Thr) substitution was identified in two affected individuals. RIT1 p.Ser35Asn, which corresponds to dominant-negative alteration p.Ser17Asn in RAS, has been used as a dominant-negative substitution in cell experiments.[38] To examine the functional consequence of p.Ser35Thr, identified in affected individuals, we compared the ELK1 transactivation in cells expressing p.Ser35Thr and those expressing p.Ser35Asn. Enhanced ELK1 transactivation was observed in cells expressing p.Ser35Thr, but not in cells expressing p.Ser35Asn (Figure 2B). These results suggest that *RIT1* mutations identified in affected individuals were gain-of-function mutations.

RIT1 is expressed ubiquitously in embryonic and adult tissues.[33,34] *Rit1*-null mice have been shown to grow to adulthood without any apparent abnormalities;[39] hence, physiological roles of RIT1 in development remain unknown. To examine the developmental effect of identified mutations, we introduced mRNA of the WT and three *RIT1* mutations (c.236A>T [p.Gln79Leu], c.242A>G [p.Glu81Gly], and c.284G>C [p.Gly95Ala]) into 1-cell-stage zebrafish embryos and observed the phenotype at 11 hr postfertilization (hpf). An oval-shaped egg sack, a typical manifestation of the gastrulation defect, was observed in embryos expressing RIT1 alterations (Figure 3A). This characteristic shape change was also observed in zebrafish expressing gain-of-function mutations of human *NRAS*[40]. Next, we observed the phenotype at later stages (48–52 hpf) (Figure 3B and Figure S4). The introduction of the WT mRNA did not interfere with the normal development, resulting in generally normal morphology

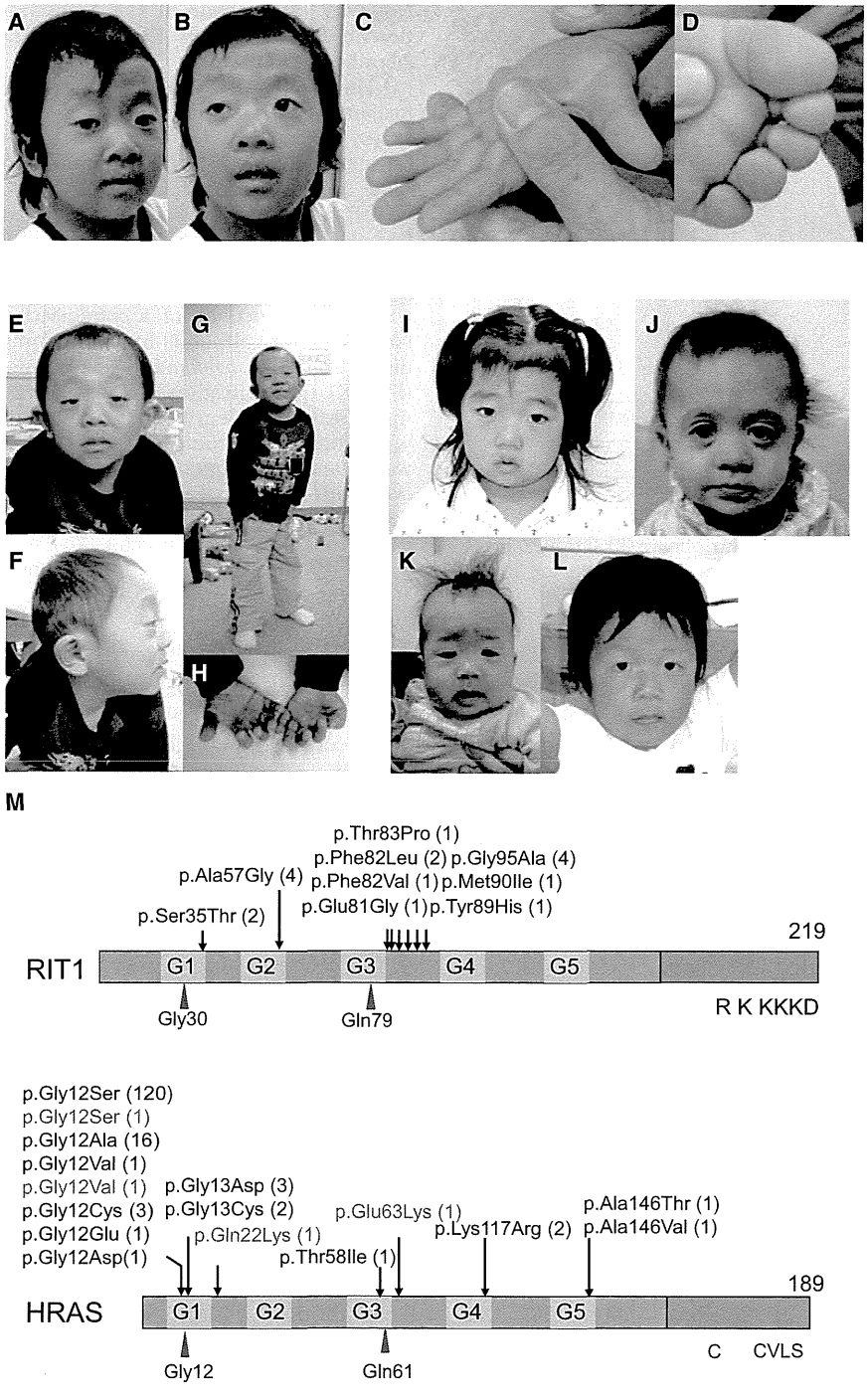

Figure 1. Photographs of Six Individuals in whom *RIT1* Mutations Were Identified
(A–D) KCC38 at 3 years of age. Broad forehead, sparse eyebrows, ptosis, hypertelorism, and hyperpigmentation were observed (A and B). Prominent finger pads were observed (C and D).
(E–H) NS358 at 4 years of age. Hypertelorism, epicanthus, sparse eyebrows, and low-set ears were observed.
(I) NS414 at 3 years of age.
(J) NS465 at 1 year of age.
(K) NS276 at 5 months.
(L) NS265 at 5 years of age.
(M) Structure and identified germline alterations in RIT1 and HRAS. HRAS alterations identified in individuals with Costello syndrome were described before[20] or shown in The RAS/MAPK Syndromes Homepage (see Web Resources). HRAS alterations identified in individuals with congenital myopathy with excess of muscle spindles[35] are indicated in purple.
We obtained specific consent for photographs from six individuals.

in 125/132 (94.7%) embryos; however, 7/132 (5.3%) embryos had limited mild craniofacial and heart abnormalities (Table 2). In contrast, a combined manifestation of craniofacial abnormalities, pericardial edema, and an elongated yolk sac was observed in 66.1%, 52.4%, and 40.5% of embryos expressing p.Gln79Leu, p.Glu81Gly, and p.Gly95Ala, respectively. Development was severely retarded in approximately 7% of embryos expressing RIT1 alterations; these embryos displayed the formation of a disorganized round body shape with a dysmorphic head and body trunk. In the head region, a hypoplastic brain, especially in the telencephalic area, was observed and resulted in misshapen morphology. In the ventral part of the head, the jaw structure was also hypoplastic, and the eyes were translocated medially. These morphological changes gave a cyclopia-like appearance. The ventral sides of the eyes were small, and coloboma along with a loss of pigment was evident (Figure 3B). These phenotypic changes are compatible with the gastrulation defect observed at 11 hpf (Figure 3A). Because the Fgf/Ras/MAPK signaling cascade plays an essential role in the convergent and extension cell movement during gastrulation,[41] perturbation by the RIT1 alterations could cause abnormal cell movement in the axial portions and thus lead to an elongated shape of the egg and the hypoplastic ventral side of the head. Detailed inspection of the morphology in mutant-injected embryos revealed abnormal cardiogenesis, namely, incomplete looping, hypoplastic chambers, and stagnation of blood flow in the yolk sac (Figure 3B). Although the atrium of these hearts beat regularly, the ventricle seemed to twitch passively by the contraction of the atrium (Movies S1, S2, S3, S4, S5, and S6). These results indicate that activating mutations in *RIT1* induce abnormal craniofacial and heart defects in zebrafish.

RIT1-mutation-positive individuals showed a distinct facial appearance, congenital heart defects, and skeletal

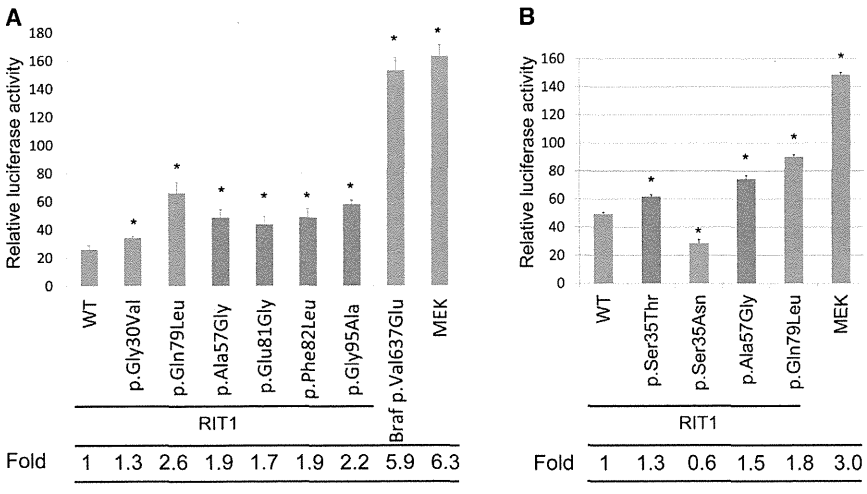

Figure 2. Stimulation of ELK Transcription in NIH 3T3 Cells Expressing *RIT1* Germline Mutations

(A) The ELK1-GAL4 vector and the GAL4 luciferase *trans*-reporter vector were transiently transfected with various *RIT1* germline mutations and activating mutations in *BRAF* and *MAP2K1* in NIH 3T3 cells. c.1910T>A (p.Val637Glu) in mouse *Braf* corresponds to oncogenic c.1799T>A (p.Val600Glu) in human *BRAF*. Relative luciferase activity was calculated by normalization to the activity of a cotransfected control vector, phRLnull-luc, containing distinguishable *R. reniformis* luciferase.

(B) ELK1 transactivation in cells expressing p.Ser35Thr, identified in individuals with Noonan syndrome, and p.Ser35Asn, were examined. p.Ser35Asn corresponds to dominant-negative alteration p.Ser17Asn in RAS.

Results are expressed as the means of quadruplicate (A) and triplicate (B) samples. Error bars represent the SDs of mean values. Red bars indicate germline *RIT1* mutations identified in Noonan syndrome. The following abbreviation is used: WT, wild-type. *$p < 0.01$ by t test.

abnormalities and were diagnosed with Noonan syndrome by diagnostic criteria developed by van der Burgt (Figures 1A–1L and Table 1).[4] Two individuals (NS358 and KCC38) were suspected to have CFC syndrome in the infantile period because of curly, sparse hair, a high cranial vault, and hypoplasia of the supraorbital ridges. Nine individuals showed perinatal abnormality, including polyhydramnios, nuchal translucency, and chylothorax (Table S2). It is of note that one individual (Og45) showing severe pleural effusion, hypertrophic cardiomyopathy, and hepatomegaly that ended in severe body edema and compromised circulation died 53 days after birth. Seven individuals showed high birth weight, probably as a result of subcutaneous edema, which is a typical manifestation observed in individuals with Noonan syndrome.[4] Out of 17 affected individuals, 16 (94%) had heart defects (Table 1): hypertrophic cardiomyopathy (HCM) in 12 (71%) individuals, pulmonary stenosis in 11 (65%) individuals, and atrial septal defects in 5 (29%) individuals. The incidence of pulmonic stenosis and mild cognitive defects is close to the overall incidence of these features in Noonan syndrome cohorts. By contrast, the incidence of HCM is far greater than in individuals with Noonan syndrome overall (25/118 in Noonan syndrome[42] versus 12/17 in individuals with *RIT1* mutations; $p < 0.0001$ by Fisher's exact test). It is of note that a high frequency of HCM (70%) was also reported in individuals with *RAF1* mutations.[10,11,24] It is possible that RIT1 interacts with RAF1 and that gain-of-function mutations in *RIT1* and *RAF1* exert similar effects in heart development.

Somatic alterations in classical RAS have been identified in approximately 30% of tumors.[43] Noonan syndrome and related disorders confer an increased risk of developing malignant tumors.[20,44] In a summary of the literature, it has been reported that 45 of 1,151 (3.9%) individuals with Noonan syndrome (but with an unknown mutation status) developed malignant tumors.[44] Since molecular analysis became available, gene-specific association with malignant tumors has been revealed. The association with JMML, a myeloproliferative disorder characterized by the excessive production of myelomonocytic cells, has been reported in individuals with *PTPN11*, *CBL*, and *KRAS* mutations. Recent reports showed that two individuals with *SOS1* mutations developed embryonal rhabdomyosarcoma.[45,46] A somatic *RIT1* variant, c.270G>A (p.Met90Ile), has been identified in lung cancer (COSMIC database). In the present cohort, 1 (NS168) of 17 individuals with *RIT1* c.242A>G (p.Glu81Gly) developed acute lymphoblastic leukemia at the age of 5 years. The child was treated by a standard protocol and has remained in complete remission. Examining whether gain-of-function mutations in *RIT1* cause tumorigenesis will require further study.

RIT1 has been isolated as a cDNA encoding highly conserved G3 and G4 domains of RAS proteins[33] or identified as a gene encoding a protein related to *Drosophila Ric*, a calmodulin-binding RAS-related GTPase.[34] RIT1 p.Gln79Leu, which corresponds to RAS p.Gln61Leu, is implicated in transforming NIH 3T3 cells, neurite outgrowth in neuronal cells, and the activation of ERK and p38 MAPK in a cell-specific manner.[37,38,47] In this study, enhanced ELK1 transactivation was observed in cells expressing mutant *RIT1* cDNAs. Previous studies showed that enhanced ELK transactivation was observed in NIH 3T3 cells expressing *HRAS*, *KRAS*, *BRAF*, and *RAF1* mutations identified in individuals with Costello, CFC, and Noonan syndromes.[17,18,24] Gastrulation defects observed in zebrafish embryos expressing RIT1 alterations (p.Glu81Gly, p.Gly95Ala, or p.Gln79Leu) were also reported in zebrafish embryos expressing an activating mutation in *NRAS*, *BRAF*, *MAP2K1*, or *MAP2K2*.[40,48] Taken together, these

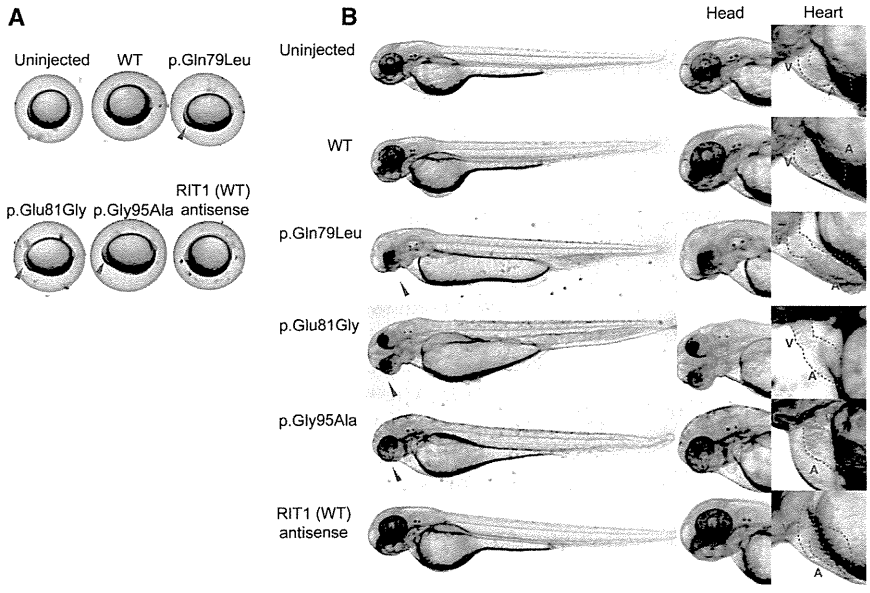

Figure 3. Morphology of Embryos Injected with the WT or Mutant *RIT1* mRNA

In vitro transcription of each mRNA was performed with the mMESSAGE mMACHINE kit (Applied Biosystems) according to the manufacturer's instructions. Synthesized mRNAs were purified with G-50 Micro Columns (GE Healthcare) and subsequently adjusted to a 300 ng/μl concentration for microinjection. Approximately 1 nl (300 pg) of RNA in water with 0.2% phenol red was injected into the cytoplasm of 1-cell-stage zebrafish embryos. Injected embryos were incubated at 28°C until observation.

(A) At 11 hpf, the shapes of the embryos injected with the WT sense or antisense mRNA were round, a normal morphology as observed in the uninjected embryos. In contrast, embryos expressing mutations (c.236A>T [p.Gln79Leu], c.242A>G [p.Glu81Gly], and c.284G>C [p.Gly95Ala]) are oval and compressed along the dorsal-ventral axis, indicative of a gastrulation defect. Note that cells have a hump in the head region at the anterior end of the body axis, the earliest manifestation of a craniofacial defect.

(B) Lateral views at 48 hpf are shown. Embryos expressing mutations (c.236A>T [p.Gln79Leu], c.242A>G [p.Glu81Gly], and c.284G>C [p.Gly95Ala]) formed swollen yolk sacs equally along the anterior posterior axis but did not show narrowing in the caudal half, which was clearly visible in the uninjected embryos and in those injected with the WT sense or antisense mRNA. In the craniofacial area, misshapen head and jaw structures and small eyes with hypoplasia on the ventral side were observed (middle panel); these phenotypes are consistent with the gastrulation defect. Shapes of the hearts (highlighted by red dotted lines) are shown in the right panel at a higher magnification. Normal looping of the heart tube and correct formation of two distinct chambers are observed in embryos injected with the WT sense or antisense mRNA. When mutations (c.236A>T [p.Gln79Leu], c.242A>G [p.Glu81Gly], and c.284G>C [p.Gly95Ala]) were expressed, looping was incomplete, resulting in stretched straight heart tubes. Constrictions at the atrial-ventricular canal are obscure, and the heart chambers are hypoplastic. Abbreviations are as follows: A, atrium; and V, ventricle.

results indicate that gain-of-function mutations in *RIT1* cause Noonan syndrome and show a similar effect to mutations in other RASopathy-related genes in human development.

Herein, we used whole-exome sequencing to identify germline *RIT1* mutations in individuals with Noonan syndrome, a disorder of the RASopathies. Mutations in *PTPN11*, *SOS1*, *RAF1*, *KRAS*, *BRAF*, and *NRAS* have been identified in 41%, 11%, 5%, 1%, 0.8%, and 0.2% of all cases, respectively,[3] and thus the frequency of *RIT1* mutations in Noonan syndrome might be similar to that of *RAF1* mutations. Our findings will improve diagnostic accuracy of Noonan syndrome and provide a clue to understanding the disorder's pathogenesis, including therapeutic approaches.

Supplemental Data

Supplemental Data include four figures, three tables, and six movies and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

The authors thank the families and the doctors who participated in this study. We are grateful to Jun-ichi Miyazaki at Osaka University for supplying the pCAGGS expression vector. We thank Yoko Narumi, Tomoko Kobayashi, Shoko Komatsuzaki, Yu Abe, Yuka Saito, Rumiko Izumi, Mitsuji Moriya, and Masako Yaoita for contributing to routine diagnostic work and Yoko Tateda, Kumi Kato, and Riyo Takahashi for their technical assistance. We are grateful to Eric Haan for sending samples of Noonan syndrome

Table 2. Morphologic Abnormality at 48–52 hpf of Zebrafish Embryos Injected with WT or Mutant RNA at the 1-Cell Stage

	No Abnormalities	Heart and Facial Abnormalities[a]	Severely Disorganized[b]	Total Number of Embryos
WT	125	7 (5.3%)	0 (0%)	132
p.Gln79Leu	31	78 (66.1%)	9 (7.6%)	118
p.Glu81Gly	42	55 (52.4%)	8 (7.6%)	105
p.Gly95Ala	44	34 (40.5%)	6 (7.1%)	84

[a]Craniofacial abnormalities, pericardial heart edema, and an elongated yolk sac were observed.
[b]Disorganized round body shape with a dysmorphic head and body trunk as shown in Figure S4.

and related disorders. We also acknowledge the support of the Biomedical Research Core of Tohoku University Graduate School of Medicine. This work was supported by the Funding Program for the Next Generation of World-Leading Researchers (NEXT Program) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) to Y.A. (LS004), by Grants-in-Aids from MEXT, from the Japan Society for the Promotion of Science, and from the Ministry of Health, Labor, and Welfare to Y.M. and T.N. This work was supported in part by the National Cancer Center Research and Development Fund (23-22-11).

Received: April 23, 2013
Revised: May 19, 2013
Accepted: May 23, 2013
Published: June 20, 2013

Web Resources

The URLs for data presented herein are as follows:

Catalogue of Somatic Mutations in Cancer (COSMIC), http://www.sanger.ac.uk/genetics/CGP/cosmic/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
The RAS/MAPK Syndromes Homepage, http://www.medgen.med.tohoku.ac.jp/RasMapk%20syndromes.html

References

1. Takai, Y., Sasaki, T., and Matozaki, T. (2001). Small GTP-binding proteins. Physiol. Rev. *81*, 153–208.
2. Giehl, K. (2005). Oncogenic Ras in tumour progression and metastasis. Biol. Chem. *386*, 193–205.
3. Romano, A.A., Allanson, J.E., Dahlgren, J., Gelb, B.D., Hall, B., Pierpont, M.E., Roberts, A.E., Robinson, W., Takemoto, C.M., and Noonan, J.A. (2010). Noonan syndrome: clinical features, diagnosis, and management guidelines. Pediatrics *126*, 746–759.
4. van der Burgt, I. (2007). Noonan syndrome. Orphanet J. Rare Dis. *2*, 4.
5. Tartaglia, M., Mehler, E.L., Goldberg, R., Zampino, G., Brunner, H.G., Kremer, H., van der Burgt, I., Crosby, A.H., Ion, A., Jeffery, S., et al. (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. Nat. Genet. *29*, 465–468.
6. Digilio, M.C., Conti, E., Sarkozy, A., Mingarelli, R., Dottorini, T., Marino, B., Pizzuti, A., and Dallapiccola, B. (2002). Grouping of multiple-lentigines/LEOPARD and Noonan syndromes on the PTPN11 gene. Am. J. Hum. Genet. *71*, 389–394.
7. Schubbert, S., Zenker, M., Rowe, S.L., Böll, S., Klein, C., Bollag, G., van der Burgt, I., Musante, L., Kalscheuer, V., Wehner, L.E., et al. (2006). Germline KRAS mutations cause Noonan syndrome. Nat. Genet. *38*, 331–336.
8. Roberts, A.E., Araki, T., Swanson, K.D., Montgomery, K.T., Schiripo, T.A., Joshi, V.A., Li, L., Yassin, Y., Tamburino, A.M., Neel, B.G., and Kucherlapati, R.S. (2007). Germline gain-of-function mutations in SOS1 cause Noonan syndrome. Nat. Genet. *39*, 70–74.
9. Tartaglia, M., Pennacchio, L.A., Zhao, C., Yadav, K.K., Fodale, V., Sarkozy, A., Pandit, B., Oishi, K., Martinelli, S., Schackwitz, W., et al. (2007). Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. Nat. Genet. *39*, 75–79.
10. Pandit, B., Sarkozy, A., Pennacchio, L.A., Carta, C., Oishi, K., Martinelli, S., Pogna, E.A., Schackwitz, W., Ustaszewska, A., Landstrom, A., et al. (2007). Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. Nat. Genet. *39*, 1007–1012.
11. Razzaque, M.A., Nishizawa, T., Komoike, Y., Yagi, H., Furutani, M., Amo, R., Kamisago, M., Momma, K., Katayama, H., Nakagawa, M., et al. (2007). Germline gain-of-function mutations in RAF1 cause Noonan syndrome. Nat. Genet. *39*, 1013–1017.
12. Cirstea, I.C., Kutsche, K., Dvorsky, R., Gremer, L., Carta, C., Horn, D., Roberts, A.E., Lepri, F., Merbitz-Zahradnik, T., König, R., et al. (2010). A restricted spectrum of NRAS mutations causes Noonan syndrome. Nat. Genet. *42*, 27–29.
13. Cordeddu, V., Di Schiavi, E., Pennacchio, L.A., Ma'ayan, A., Sarkozy, A., Fodale, V., Cecchetti, S., Cardinale, A., Martin, J., Schackwitz, W., et al. (2009). Mutation of SHOC2 promotes aberrant protein N-myristoylation and causes Noonan-like syndrome with loose anagen hair. Nat. Genet. *41*, 1022–1026.
14. Loh, M.L., Sakai, D.S., Flotho, C., Kang, M., Fliegauf, M., Archambeault, S., Mullighan, C.G., Chen, L., Bergstraesser, E., Bueso-Ramos, C.E., et al. (2009). Mutations in CBL occur frequently in juvenile myelomonocytic leukemia. Blood *114*, 1859–1863.
15. Niemeyer, C.M., Kang, M.W., Shin, D.H., Furlan, I., Erlacher, M., Bunin, N.J., Bunda, S., Finklestein, J.Z., Sakamoto, K.M., Gorr, T.A., et al. (2010). Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. Nat. Genet. *42*, 794–800.
16. Pérez, B., Mechinaud, F., Galambrun, C., Ben Romdhane, N., Isidor, B., Philip, N., Derain-Court, J., Cassinat, B., Lachenaud, J., Kaltenbach, S., et al. (2010). Germline mutations of the CBL gene define a new genetic syndrome with predisposition to juvenile myelomonocytic leukaemia. J. Med. Genet. *47*, 686–691.
17. Aoki, Y., Niihori, T., Kawame, H., Kurosawa, K., Ohashi, H., Tanaka, Y., Filocamo, M., Kato, K., Suzuki, Y., Kure, S., and Matsubara, Y. (2005). Germline mutations in HRAS proto-oncogene cause Costello syndrome. Nat. Genet. *37*, 1038–1040.
18. Niihori, T., Aoki, Y., Narumi, Y., Neri, G., Cavé, H., Verloes, A., Okamoto, N., Hennekam, R.C., Gillessen-Kaesbach, G., Wieczorek, D., et al. (2006). Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome. Nat. Genet. *38*, 294–296.
19. Rodriguez-Viciana, P., Tetsu, O., Tidyman, W.E., Estep, A.L., Conger, B.A., Cruz, M.S., McCormick, F., and Rauen, K.A. (2006). Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. Science *311*, 1287–1290.
20. Aoki, Y., Niihori, T., Narumi, Y., Kure, S., and Matsubara, Y. (2008). The RAS/MAPK syndromes: novel roles of the RAS pathway in human genetic disorders. Hum. Mutat. *29*, 992–1006.
21. Tidyman, W.E., and Rauen, K.A. (2009). The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation. Curr. Opin. Genet. Dev. *19*, 230–236.
22. Groesser, L., Herschberger, E., Ruetten, A., Ruivenkamp, C., Lopriore, E., Zutt, M., Langmann, T., Singer, S., Klingseisen, L., Schneider-Brachert, W., et al. (2012). Postzygotic HRAS and KRAS mutations cause nevus sebaceous and Schimmelpenning syndrome. Nat. Genet. *44*, 783–787.