Journal of Clinical Microbiology

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

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

J. Clin. Microbiol. **2012, 50(3):857. DOI: 10.1128/JCM.05715-11.**
Published Ahead of Print 28 December 2011.

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).

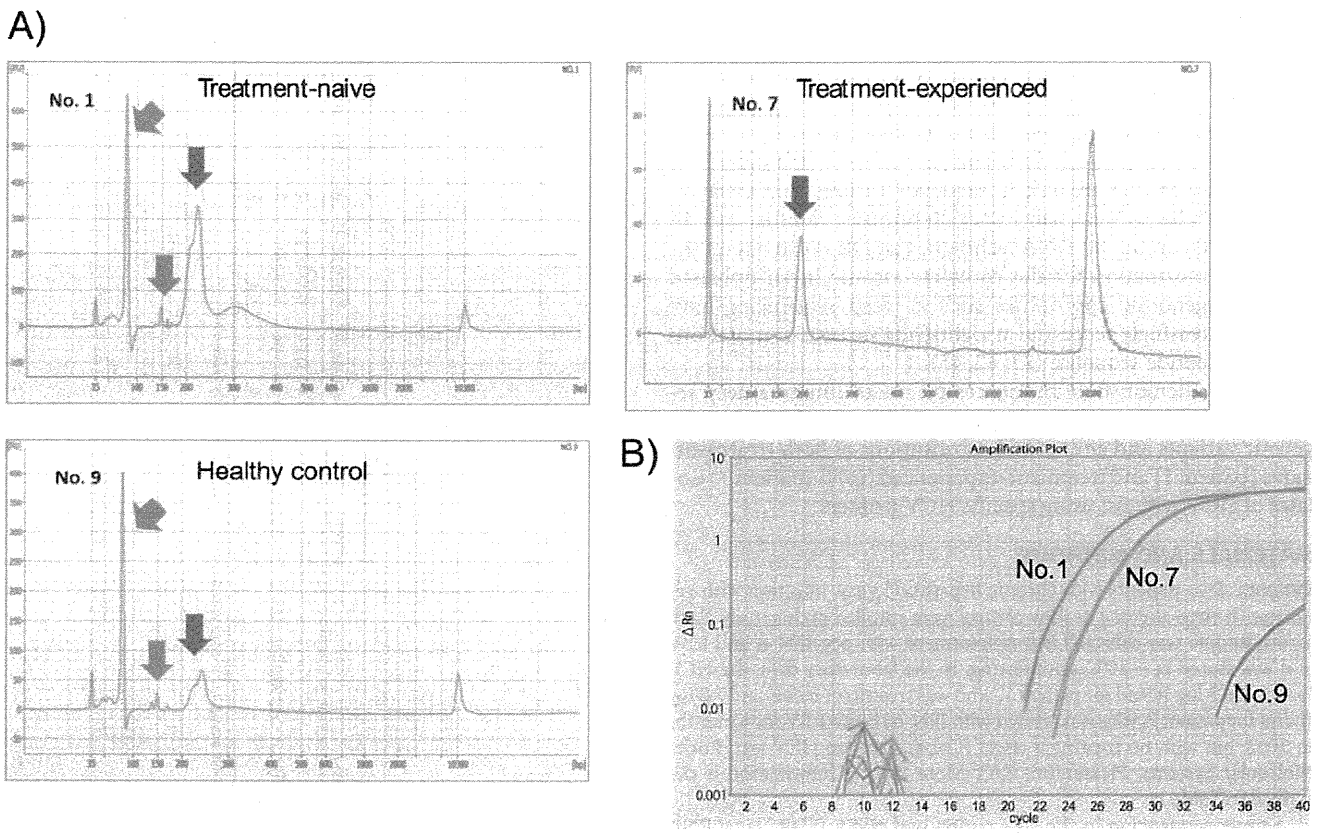

FIG 1 Evaluation of the quality of libraries. (A) The library was well refined for patient 7, but the primer and adaptor dimers were mixed in the libraries of patients 1 and 9, obtained using Bioanalyzer (Agilent). The horizontal axis shows the DNA size, and the vertical axis shows the quantity of DNA. 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

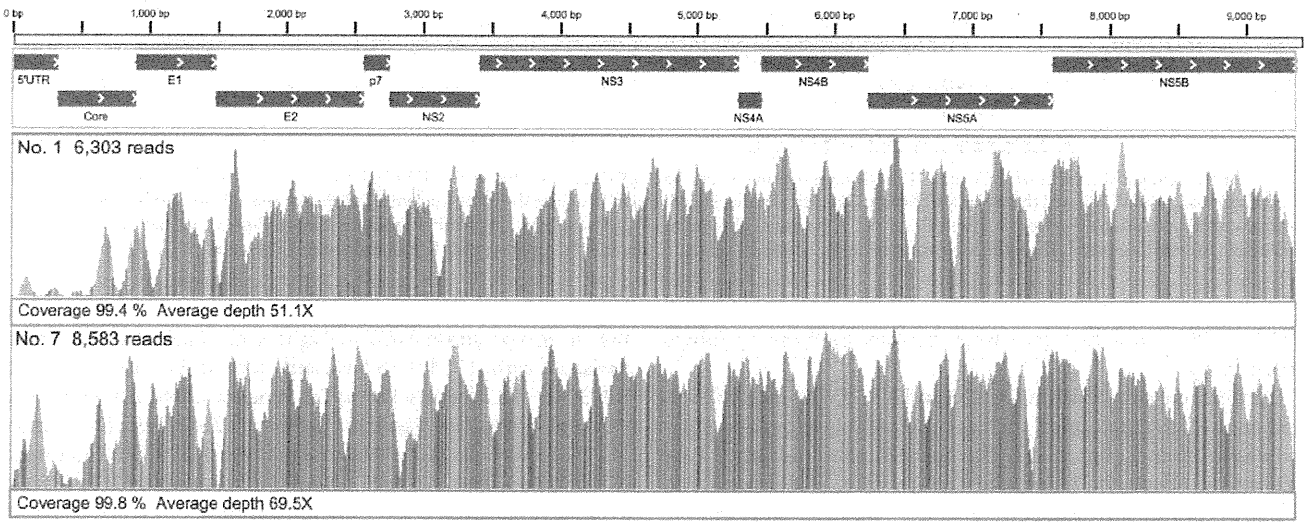

FIG 2 Mapping to the HCV reference genome. For patient 1, 6,303 reads were mapped to MD5-1. The coverage was 99.4%, and the average depth was 51.1×. For patient 7, 8,583 reads were aligned to MD2-2. The coverage was 99.8%, and the average depth was 69.5%.

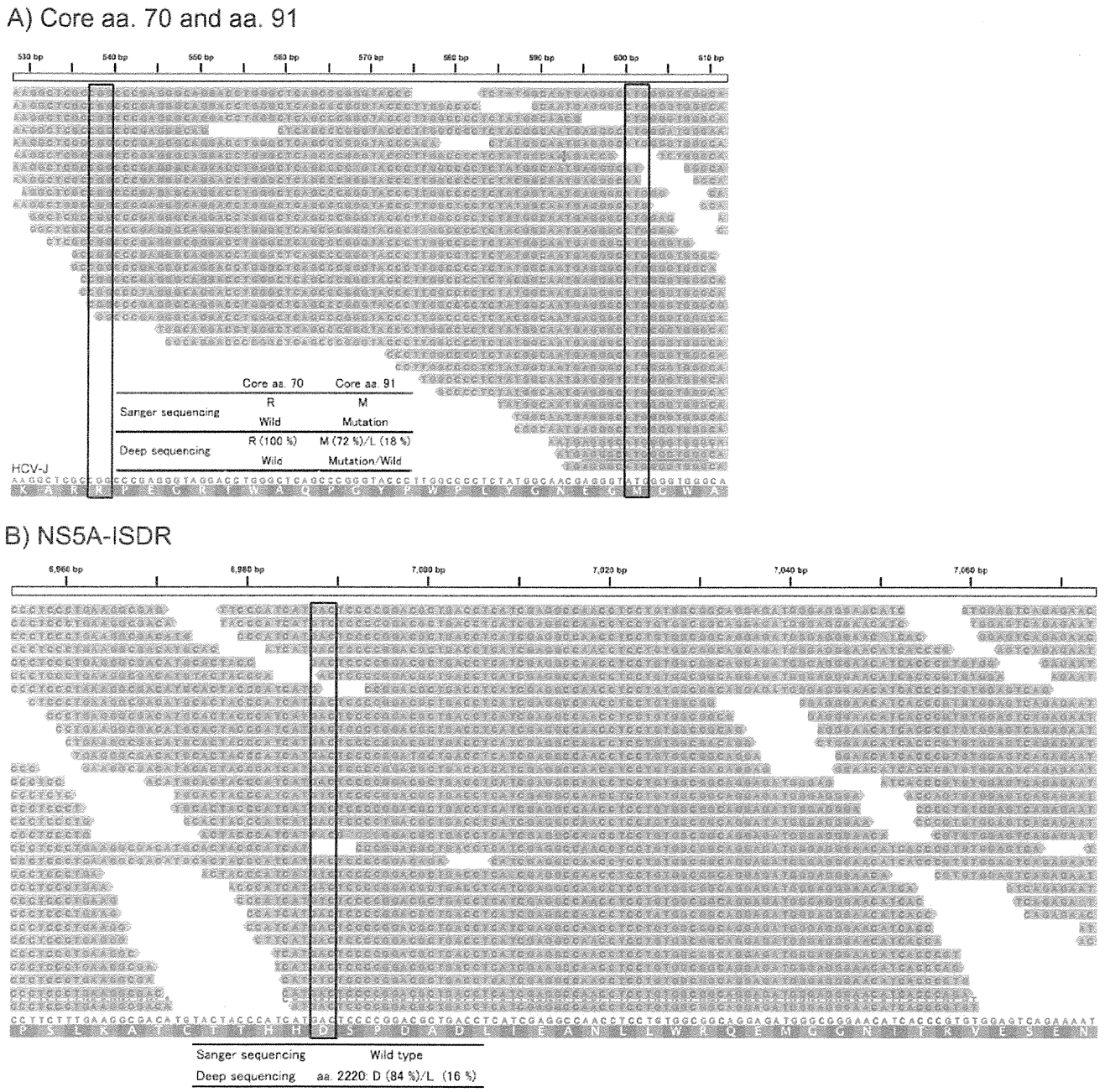

FIG 3 Amino acid substitutions of aa 70 and 91 in the core region (A) and aa 2209 to 2248 of the NS5A ISDR (B). Mutations in these regions were reported to affect the outcome of IFN-based therapies for chronic hepatitis C patients. The lower two lines show the nucleotide sequence and amino acid sequence 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.

Journal of the Neurological Sciences xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome

Mitsugu Uematsu [a,*], Kazuhiro Haginoya [b], Atsuo Kikuchi [a], Tojo Nakayama [a], Yousuke Kakisaka [a], Yurika Numata [a], Tomoko Kobayashi [a], Naomi Hino-Fukuyo [a], Ikuma Fujiwara [a], Shigeo Kure [a]

[a] Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan
[b] Department of Pediatric Neurology, Takuto Rehabilitation Center for Children, Sendai, Japan

ARTICLE INFO

Article history:
Received 20 August 2011
Received in revised form 11 November 2011
Accepted 15 November 2011
Available online xxxx

Keywords:
Brain–lung–thyroid syndrome
NKX2-1
Array CGH
ECD-SPECT
eZIS

ABSTRACT

Mutations in *NKX2-1* cause neurological, pulmonary, and thyroid hormone impairment. Recently, the disease was named brain–lung–thyroid syndrome. Here, we report three patients with brain–lung–thyroid syndrome. All patients were unable to walk until 24 months of age, and still have a staggering gait, without mental retardation. They have also had choreoathetosis since early infancy. Genetic analysis of *NKX2-1* revealed a novel missense mutation (p.Val205Phe) in two patients who were cousins and their maternal families, and a novel 2.6-Mb deletion including *NKX2-1* on chromosome 14 in the other patient. Congenital hypothyroidism was not detected on neonatal screening in the patient with the missense mutation, and frequent respiratory infections were observed in the patient with the deletion in *NKX2-1*. Oral levodopa did not improve the gait disturbance or involuntary movement. The results of ^{99m}Tc-ECD single-photon emission computed tomography (ECD-SPECT) analyzed using the easy Z-score imaging system showed decreased cerebral blood flow in the bilateral basal ganglia, especially in the caudate nuclei, in all three patients, but no brain magnetic resonance imaging (MRI) abnormalities. These brain nuclear image findings indicate that *NKX2-1* haploinsufficiency causes dysfunction of the basal ganglia, especially the caudate nuclei, resulting in choreoathetosis and gait disturbance in this disease.

1. Introduction

NK2 homeobox 1 (*NKX2-1* or *TITF-1*; MIM #600635), which maps on chromosome 14q13, is a member of the *NK-2* gene family of highly conserved homeodomain-containing transcription factors [1,2]. The gene is expressed in the thyroid, bronchial epithelium, and specific areas of the forebrain during development in the mouse [3–5]. Mice homozygous for the disrupted gene are born dead and lack a thyroid gland, lung parenchyma, and pituitary gland, while heterozygous mice develop normally [4]. An abnormality of the gene in humans was first reported in patients with congenital hypothyroidism [6]. Subsequently, heterozygous point mutations in *NKX2-1* were identified in affected members of a family with benign hereditary chorea [7]. Recently, *NKX2-1* was reported as the gene responsible for brain–lung–thyroid syndrome (MIM #610978), which involves symptoms of neurological impairment, pulmonary disorders, and hypothyroidism [8–13]. Respiratory distress during the neonatal period, recurrent respiratory tract infection, and hypothyroidism are common clinical findings. The neurological impairment is characterized by gait disturbance with delayed first walking and choreoathetosis, in the absence of mental retardation or brain magnetic resonance imaging (MRI) abnormalities [13]. However, some affected individuals have had low-average intelligence, learning problems, psychosis and seizures [14–16].

The pathological mechanism of *NKX2-1* haploinsufficiency has been clarified for the hypothyroidism [17] and pulmonary impairment [18,19], but it is still unclear for the neurological symptoms. Most of the neurological deficits, i.e., the gait disturbance and involuntary movements sometimes accompanied with dystonia, dysarthria, action tremor and saccadic abnormalities [20], reflect dysfunction of the control of movement. Therefore, the basal ganglia were considered to be the most important causal lesion [8,14]. The *NKX2-1* null mouse showed severe morphological changes in the basal ganglia, including absence of the globus pallidus and enlargement of the striatum [4]. *NKX2-1* gene expression has been identified as the origin of the pallidum in the mammalian and avian embryonic archistriatum. These studies indicated that *NKX2*-1 is essential for development of the striatum, especially the pallidum rather than the caudate nuclei [5,21,22].

Brain MRI of patients with brain–lung–thyroid syndrome showed no notable abnormalities, except one case report of reduced size and intensity in the pallidum [8]. Previous brain nuclear imaging studies described various findings regarding the basal ganglia, including reduced blood flow in the striatum and thalamus [23], and hypometabolism in the basal ganglia, more prominent in the caudate nuclei [15].

* Corresponding author at: Department of Pediatrics, Tohoku University School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan. Tel: +81 22 717 7287; fax: +81 22 717 7290.
E-mail address: uematsu@bk9.so-net.ne.jp (M. Uematsu).

0022-510X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jns.2011.11.025

Please cite this article as: Uematsu M, et al, Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome, J Neurol Sci (2011), doi:10.1016/j.jns.2011.11.025

Here, we report three patients with brain–lung–thyroid syndrome in whom the diagnosis was confirmed by genetic examinations. We performed brain nuclear image analysis to investigate the causal lesion for the neurological symptoms.

2. Method

2.1. Clinical findings

We studied three patients (5, 6, and 7 years old; one male and two females) with gait disturbance who visited Tohoku University Hospital between 2008 and 2009 (Table 1).

Patient 1 was the second female child of healthy non-consanguineous parents (Fig. 1). She was born at term without neonatal respiratory problems. Congenital hypothyroidism was noted on neonatal screening and she has been given thyroxin replacement therapy since then. After the age of 1.6 years, she developed recurrent respiratory infections and was admitted to hospital five times in one year. She had normal mental development, but delayed gross motor development. She could sit alone at the age of 12 months and first walked at 38 months. A staggering gait persists. Her trunk and extremes were mildly hypotonic and continuous choreoathetosis was observed during wakefulness and exacerbated by stress.

Patients 2 and 3 were cousins via their maternal families (Fig. 1). Patient 2 was the third female and Patient 3 was an only male child. Both sets of non-consanguineous parents were healthy fathers and affected mothers with mild involuntary movement and a history of delayed first walking. Both patients were born at term without any perinatal complications. Congenital hypothyroidism was diagnosed in the neonatal period by screening in Patient 2, but at the age of 5 years in Patient 3, despite a neonatal screening test. Unlike Patient 1, they had no severe respiratory infections during infancy. Similar to Patient 1, first walking was observed at 30 months in Patient 2 and at 24 months in Patient 3. They also have persistent gait disturbance and choreoathetosis without mental retardation. The neurological examinations in all three patients did not detect any abnormalities, such as muscle weakness, abnormal deep tendon reflexes, or cerebellar manifestations.

Brain MRI in all three patients showed normal brain size, form, and intensity, including the basal ganglia. Oral levodopa (20 mg/kg/day) was given to all three patients, but no obvious improvement in the neurological symptoms was observed.

2.2. Brain nuclear image analysis

All three patients underwent single photon emission computed tomography (SPECT) to evaluate brain function at Tohoku University Hospital using technetium-99 m ethyl cysteinate dimer (ECD, approximately 12 MBq/kg of body weight) as the radiotracer. Twenty minutes after the injection, SPECT images were acquired using a PRISM IRIX (Shimadzu, Kyoto, Japan), with a low-energy, high-resolution, fan-beam collimator. In total, 120 projection datum points in a 128×128 matrix were obtained in 20 min. Using an ODYSSEY computer (Shimadzu), tomograms two pixels thick (5.8 mm) were reconstructed after a high-frequency cutoff with a Butterworth filter.

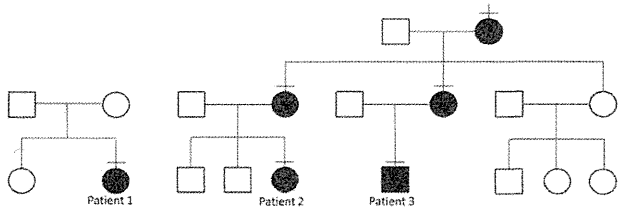

Fig. 1. Family pedigrees of three patients. Affected members are indicated by black squares and circles; unaffected members, white squares and circles. Patient 2 and 3 are cousins on mother's side.

The easy Z-Score Imaging System (eZIS; Fuji Film RI Pharma), used for the statistical analysis of SPECT images, standardizes brain images using Statistical Parametric Mapping (SPM99) [24]. Each SPECT image of the subjects after anatomical standardization followed by isotropic 12-mm smoothing was compared with the mean and SD of SPECT images of the age-matched healthy controls already incorporated in the eZIS program as a normal database using voxel-by-voxel Z-score analysis after voxel normalization to global mean values: Z score = (control mean − individual value)/control SD. These Z-score maps were overlain on tomographic sections and projection with an averaged Z-score of 14-mm thickness to surface rendering of the anatomically standardized MRI template.

Positron emission tomography (PET) was performed in Patients 2 and 3, 1 h after administering [^{18}F]-fluorodeoxyglucose (^{18}FDG) (approximately 3 MBq/kg of body weight) using a Biograph Duo, ECAT EXACT HR^{+} (Siemens, Hoffman Estates, IL) or SET-2400 W (Shimadzu) after fasting for at least 4 h. Emission scans were performed for 10 min for the entire brain. Attenuation was corrected. Fourteen 6-mm-thick slices parallel to the orbitomeatal line, encompassing virtually the entire brain, were analyzed visually by two investigators independently. When the interpretation was inconsistent, a third investigator was called to make a decision.

2.3. Gene analysis

Gene analyses were performed with the informed consent of the patients' parents. Genomic DNA was extracted from peripheral blood lymphocytes using a Sepa Gene kit (Sanko Junyaku, Tokyo, Japan). All coding exons and flanking introns in *NKX2-1* were amplified by PCR. All primers were based on the NCBI reference sequence (accession number NG_013365; the primer sequences are available upon request). The PCR products were separated on 3% agarose gels and purified with a QIAquick Gel Extraction kit (QIAGEN, Chatsworth, CA, USA). The PCR products were sequenced directly using a Big Dye Primer Cycle Sequencing kit and ABI 310 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA).

Subsequent array-based comparative genomic hybridization (CGH) analysis was performed using an Agilent 244 K oligonucleotide array (Agilent, Santa Clara, CA; www.agilent.com) with a resolution of approximately 15 kb following the protocols provided by Agilent. The array was analyzed with the Agilent scanner and the Feature Extraction software (v. 9.1.3).

Table 1
Clinical characteristics in three patients.

	Patient 1	Patient 2	Patient 3
Age/sex	7 years/female	5 years/female	6 years/male
Recurrence of respiratory infection	Yes	No	No
Neonatal respiratory problems	No	No	No
Hypothyroidism	Yes (neonatal screening)	Yes (neonatal screening)	Yes (diagnosed at 5 years)
Initiation of walking	3 years and 2 months	2 years and 6 months	2 years
Mental retardation	No	No	No
Choreoathetosis	Yes	Yes	Yes
Response to L-dopa	No	No	No
Brain MRI	Normal	Normal	Normal
NKX2-1 analysis	del 14q12–13	p.V205P	p.V205P

3. Results

From the raw nuclear image ECD-SPECT findings in all three patients (Fig. 2, lower figures) and FDG-PET in Patients 2 and 3

Please cite this article as: Uematsu M, et al, Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome, J Neurol Sci (2011), doi:10.1016/j.jns.2011.11.025

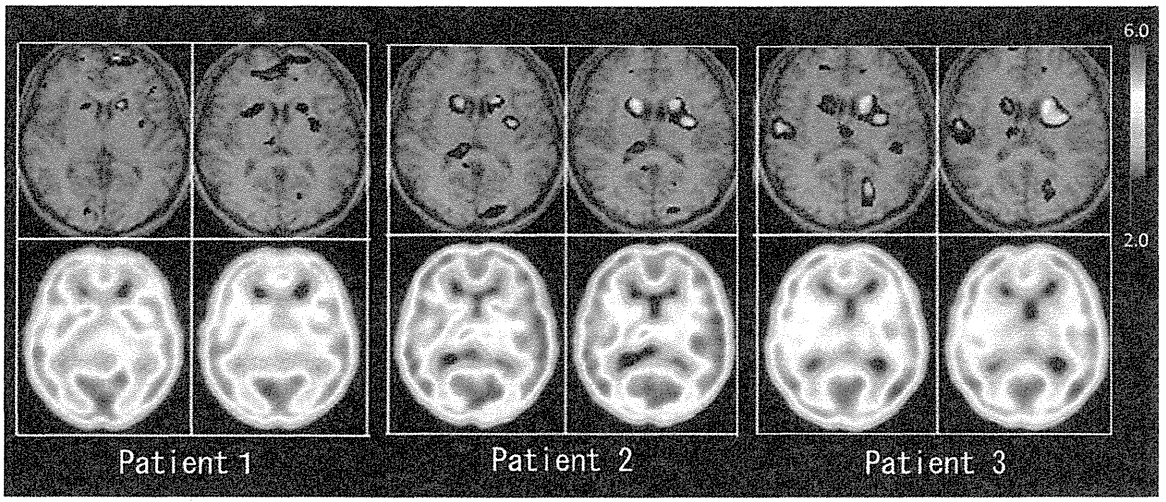

Fig. 2. ECD-SPECT and eZIS results.Usual color images of ECD-SPECT in three patients were shown on the lower figures in each patient. Analyzed images using eZIS were shown on the upper figures. The converted images indicate regions of decreased cerebral blood flow by colors from blue (2.0 standard deviation) to red (6.0 standard deviation). Reduction in cerebral blood flow was shown most common and prominent in the bilateral caudate nuclei in all three patients.

(data not shown), we could not discriminate visible areas of abnormal cerebral perfusion or glucose metabolism. However, the statistical analysis of the ECD-SPECT data using eZIS demonstrated significant declines in cerebral blood flow in the basal ganglia, especially in the caudate nuclei (Fig. 2, upper figure). Although several other brain regions were shown to have decreased blood flow, these were not shared in the three patients.

Array CGH analysis revealed that Patient 1 had an approximately 2.6-Mb hemizygous deletion including *NKX2-1* in 14q12–13 (Fig. 3, top). Direct sequencing analysis revealed a novel hemizygous mutation in the coding exons in Patients 2 and 3 (Fig. 3, bottom), but no mutation in Patient 1. A hemizygous G-to-T substitution at nucleotide position 613 (c.613G>T) in Patients 2 and 3 created an amino acid substitution at amino acid position 205 (p.Val205Phe) within exon 3, which is localized within the *NKX2-1* homeodomain. The mothers and grandmother of Patients 2 and 3 had the same missense mutation.

4. Discussion

In this study, we diagnosed three children with brain–lung–thyroid syndrome based on clinical findings of delayed walking, unsteady gait, choreoathetosis, and hypothyroidism. The diagnosis was confirmed by genetic analysis detecting a novel hemizygous deletion and missense mutations in *NKX2-1*. In addition, we performed nuclear image examinations and analyzed the results using statistical image analysis with eZIS. We found a significant reduction in the blood flow in the caudate nuclei in all three patients.

Recurrent respiratory infection was observed only in the patient with the deletion in *NKX2-1*, but not in the two patients with missense mutations. These phenotype-genotype correlations support previous reports [8,10–13] that large deletions and truncation mutations are related to the severe phenotype with the symptom triad, while missense mutations have a milder phenotype [13]. The existence of respiratory symptoms is very important for management because no deaths have been reported in patients without lung disease [12]. In one case with a missense mutation, hypothyroidism was not detected until the age of 5 years, while it was detected in his cousin with the same mutation at neonatal screening. This interfamilial heterogeneity, as described previously [25], indicates that simple haploinsufficiency cannot fully explain the spectrum of clinical presentations. Other modifying genes might contribute to the phenotype heterogeneity [9,12].

Central nervous impairment is the most common and essential symptom in brain–lung–thyroid syndrome [12,13]. Typically, mental retardation and brain MRI abnormalities are not associated with this disease. The characteristic presentation involves delayed walking, a staggering gait, and choreoathetosis. Since no obvious nervous system abnormalities were detected on neurological examinations, we speculate that the choreoathetosis in the lower limbs caused the delay in walking and unsteady gait.

To identify the brain region responsible for the neurological impairment, we examined brain ECD-SPECT and FDG-PET, but no obvious pathological abnormalities were detected on visual inspection of the raw images. Further statistical analysis of the ECD-SPECT data was performed using eZIS, while the FDG-PET was not analyzed because we did not have an appropriate analysis method. The eZIS method can detect a significant difference of regional cerebral blood flow by comparison with age-matched normal controls, and shows the result as color images. Previous reports described significant superiority of this program over visual inspection of raw SPECT images in several diseases [26–28]. We analyzed our patients and found a common, significant reduction in cerebral blood flow in the caudate nuclei. Although most reports describe expression of the *NKX2-1* gene in the pallidum [8,21,22], a recent study showed *NKX2-1* expression in the postnatal mouse striatum, including the caudate nuclei, in addition to the pallidum [29]. In humans, nuclear image studies indicated a reduction in blood flow [23] and glucose metabolism [15] in the basal ganglia. Hypoperfusion in the caudate nuclei was described in a patient with Huntington's disease [30,31], which usually involves chorea. From these reports and our ECD-SPECT findings using eZIS, we believe that the region responsible for the neurological symptoms in brain–lung–thyroid syndrome, pathologically, is the caudate nuclei. We speculate the possible mechanism that the mutation may impair developmental differentiation and organization of the striatum. Huntington's disease, which also involves the caudate nuclei, partially mimics brain–lung–thyroid syndrome clinically, although the latter is easily differentiated by the history of delayed walking.

In our cases, oral L-dopa [32], a dopamine agonist, and clonazepam failed to improve their neurological impairment. Only a few effective treatment for involuntary movement has been reported [13]. Although some reports described the choreatic movements tend to decrease over time [7,23], the movement disability causes severe trouble with daily life, especially writing difficulty resulting in a leaning impairment in

Please cite this article as: Uematsu M, et al, Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome, J Neurol Sci (2011), doi:10.1016/j.jns.2011.11.025

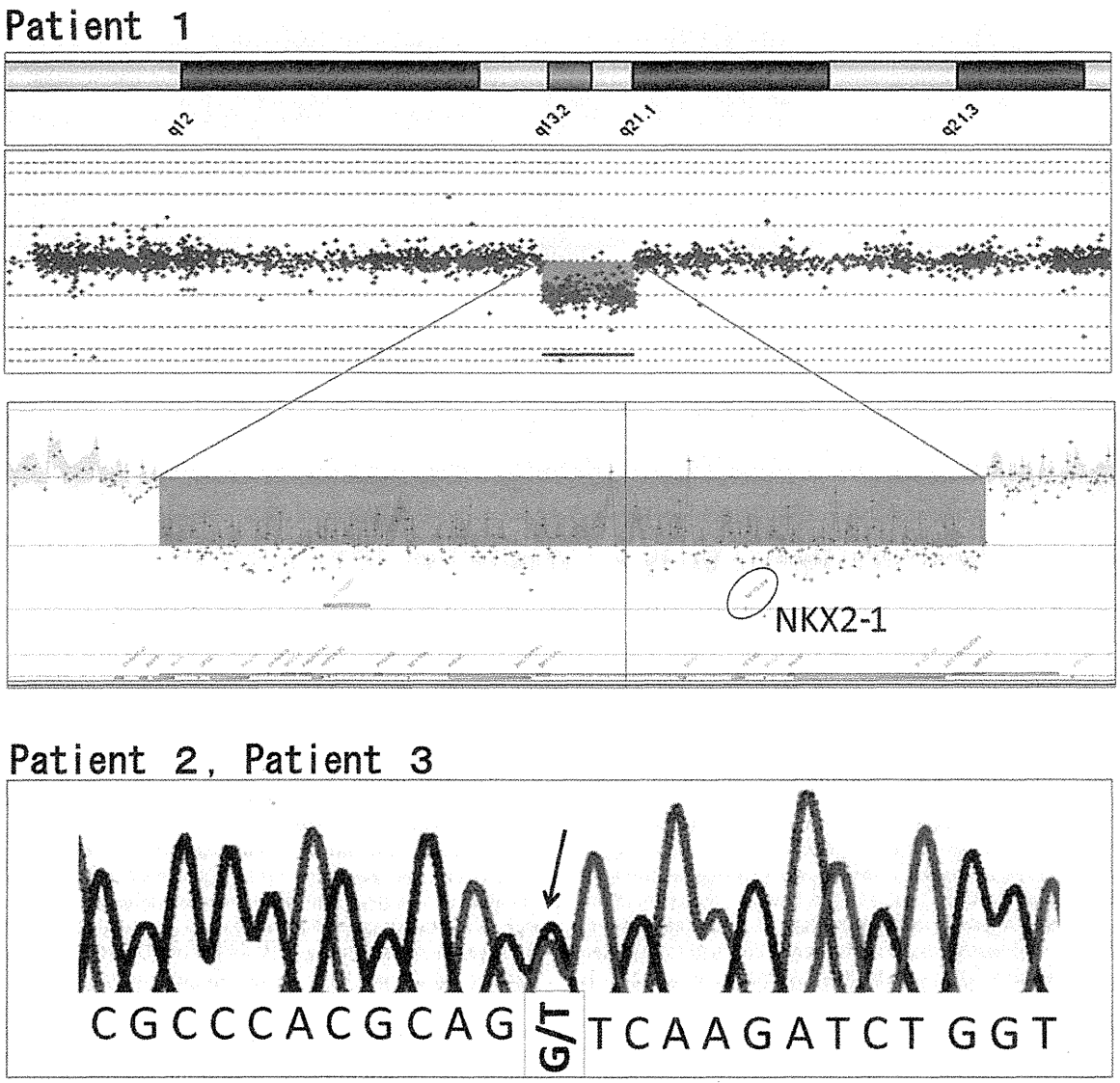

Fig. 3. Deletion and missense mutation in *NKX2-1*. Chromosome 14 profile and detail of 14q12–13 region generated by Cytogenomics (version 1.5, Agilent Technologies) showed a hemizygous 2.6-Mb deletion including *NKX2-1* in Patient 1. Sequencing analysis showed hemizygous missense mutations (c.613G>T) in *NKX2-1* in Patients 2 and 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

childhood. We postulate that deep brain stimulation might treat involuntary movement, such as in Huntington's disease [33]. The pathophysiology of this disease should be clarified to develop an effective treatment.

Acknowledgments

We are grateful to Ms. Yoko Chiba and Ms. Kumi Itou for their technical assistance.

References

[1] Hamdan H, Liu H, Li C, Jones C, Lee M, deLemos R, et al. Structure of the human Nkx2.1 gene. Biochim Biophys Acta 1998;1396:336–48.
[2] Ikeda K, Clark JC, Shaw-White JR, Stahlman MT, Boutell CJ, Whitsett JA. Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. J Biol Chem 1995;270:8108–14.
[3] Lazzaro D, Price M, de Felice M, Di Lauro R. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. Development 1991;113:1093–104.
[4] Kimura S, Hara Y, Pineau T, Fernandez-Salguero P, Fox CH, Ward JM, et al. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. Genes Dev 1996;10:60–9.
[5] Puelles L, Kuwana E, Puelles E, Bulfone A, Shimamura K, Keleher J, et al. Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. J Comp Neurol 2000;424:409–38.
[6] Acebron A, Aza-Blanc P, Rossi DL, Lamas L, Santisteban P. Congenital human thyroglobulin defect due to low expression of the thyroid-specific transcription factor TTF-1. J Clin Invest 1995;96:781–5.
[7] Breedveld GJ, van Dongen JW, Danesino C, Guala A, Percy AK, Dure LS, et al. Mutations in TITF-1 are associated with benign hereditary chorea. Hum Mol Genet 2002;11:971–9.
[8] Krude H, Schutz B, Biebermann H, von Moers A, Schnabel D, Neitzel H, et al. Choreoathetosis, hypothyroidism, and pulmonary alterations due to human NKX2-1 haploinsufficiency. J Clin Invest 2002;109:475–80.
[9] Breedveld GJ, Percy AK, MacDonald ME, de Vries BB, Yapijakis C, Dure LS, et al. Clinical and genetic heterogeneity in benign hereditary chorea. Neurology 2002;59:579–84.
[10] Willemsen MA, Breedveld GJ, Wouda S, Otten BJ, Yntema JL, Lammens M, et al. Brain–thyroid–lung syndrome: a patient with a severe multi-system disorder due to a de novo mutation in the thyroid transcription factor 1 gene. Eur J Pediatr 2005;164:28–30.
[11] Kleiner-Fisman G, Lang AE. Benign hereditary chorea revisited: a journey to understanding. Mov Disord 2007;22:2297–305 [quiz 452].
[12] Carre A, Szinnai G, Castanet M, Sura-Trueba S, Tron E, Broutin-L'Hermite I, et al. Five new TTF1/NKX2.1 mutations in brain–lung–thyroid syndrome: rescue by PAX8 synergism in one case. Hum Mol Genet 2009;18:2266–76.

Please cite this article as: Uematsu M, et al, Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome, J Neurol Sci (2011), doi:10.1016/j.jns.2011.11.025

[13] Inzelberg R, Weinberger M, Gak E. Benign hereditary chorea: an update. Parkinsonism Relat Disord 2011;17:301–7.
[14] Carmo Costa M, Costa C, Silva AP, Evangelista P, Santos L, Ferro A, et al. Nonsense mutation in TITF1 in a Portuguese family with benign hereditary chorea. Neurogenetics 2005;6:209–15.
[15] Salvatore E, Di Maio L, Filla A, Ferrara AM, Rinaldi C, Sacca F, et al. Benign hereditary chorea: clinical and neuroimaging features in an Italian family. Mov Disord 2010;25: 1491–6.
[16] Butt SJB, Sousa VH, Fuccillo MV, Hjerling-Leffler J, Miyoshi G, Kimura S, et al. The requirement of Nkx2-1 in the temporal specification of cortical interneuron subtypes. Neuron 2008;59:722–32.
[17] Trueba SS. PAX8, TITF1, and FOXE1 gene expression patterns during human development: new insights into human thyroid development and thyroid dysgenesis-associated malformations. J Clin Endocrinol Metabol 2004;90:455–62.
[18] Maeda Y, Davé V, Whitsett J. Transcriptional control of lung morphogenesis. Physiol Rev 2007;87:219–44.
[19] Guillot L, Carré A, Szinnai G, Castanet M, Tron E, Jaubert F, et al. NKX2-1mutations leading to surfactant protein promoter dysregulation cause interstitial lung disease in "Brain–Lung–Thyroid Syndrome". Hum Mutat 2010;31:E1146–62.
[20] Glik A, Vuillaume I, Devos D, Inzelberg R. Psychosis, short stature in benign hereditary chorea: a novel thyroid transcription factor-1 mutation. Mov Disord 2008;23:1744–7.
[21] Sussel L, Marin O, Kimura S, Rubenstein JL. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. Development 1999;126:3359–70.
[22] Flandin P, Kimura S, Rubenstein JLR. The progenitor zone of the ventral medial ganglionic eminence requires Nkx2-1 to generate most of the globus pallidus but few neocortical interneurons. J Neurosci 2010;30:2812–23.
[23] Mahajnah M, Inbar D, Steinmetz A, Heutink P, Breedveld GJ, Straussberg R. Benign hereditary chorea: clinical, neuroimaging, and genetic findings. J Child Neurol 2007;22:1231–4.
[24] Kanetaka H, Matsuda H, Asada T, Ohnishi T, Yamashita F, Imabayashi E, et al. Effects of partial volume correction on discrimination between very early Alzheimer's dementia and controls using brain perfusion SPECT. Eur J Nucl Med Mol Imaging 2004;31:975–80.
[25] Montanelli L, Tonacchera M. Genetics and phenomics of hypothyroidism and thyroid dys- and agenesis due to PAX8 and TTF1 mutations. Mol Cell Endocrinol 2010;322: 64–71.
[26] Matsuda H, Mizumura S, Nagao T, Ota T, Iizuka T, Nemoto K, et al. Automated discrimination between very early Alzheimer disease and controls using an easy Z-score imaging system for multicenter brain perfusion single-photon emission tomography. AJNR Am J Neuroradiol 2007;28:731–6.
[27] Imamura K, Wada-Isoe K, Kowa H, Tanabe Y, Nakashima K. The effect of donepezil on increased regional cerebral blood flow in the posterior cingulate cortex of a patient with Parkinson's disease dementia. Neurocase 2008;14:271–5.
[28] Sasaki M, Nakagawa E, Sugai K, Shimizu Y, Hattori A, Nonoda Y, et al. Brain perfusion SPECT and EEG findings in children with autism spectrum disorders and medically intractable epilepsy. Brain Dev 2010;32:776–82.
[29] Magno L, Catanzariti V, Nitsch R, Krude H, Naumann T. Ongoing expression of Nkx2.1 in the postnatal mouse forebrain: potential for understanding NKX2.1 haploinsufficiency in humans? Brain Res 2009;1304:164–86.
[30] Hasselbalch SG, Oberg G, Sorensen SA, Andersen AR, Waldemar G, Schmidt JF, et al. Reduced regional cerebral blood flow in Huntington's disease studied by SPECT. J Neurol Neurosurg Psychiatry 1992;55:1018–23.
[31] Reynolds Jr NC, Hellman RS, Tikofsky RS, Prost RW, Mark LP, Elejalde BR, et al. Single photon emission computerized tomography (SPECT) in detecting neurodegeneration in Huntington's disease. Nucl Med Commun 2002;23:13–8.
[32] Asmus F, Horber V, Pohlenz J, Schwabe D, Zimprich A, Munz M, et al. A novel TITF-1 mutation causes benign hereditary chorea with response to levodopa. Neurology 2005;64:1952–4.
[33] Kang GA, Heath S, Rothlind J, Starr PA. Long-term follow-up of pallidal deep brain stimulation in two cases of Huntington's disease. J Neurol Neurosurg Psychiatry 2010;82:272–7.

Please cite this article as: Uematsu M, et al, Hypoperfusion in caudate nuclei in patients with brain–lung–thyroid syndrome, J Neurol Sci (2011), doi:10.1016/j.jns.2011.11.025

HMG Advance Access published December 30, 2011

Human Molecular Genetics, 2011 **1–8**
doi:10.1093/hmg/ddr585

Mutations in genes encoding the glycine cleavage system predispose to neural tube defects in mice and humans

Ayumi Narisawa[1,2], Shoko Komatsuzaki[1], Atsuo Kikuchi[3], Tetsuya Niihori[1], Yoko Aoki[1], Kazuko Fujiwara[4], Mitsuyo Tanemura[5], Akira Hata[6], Yoichi Suzuki[6], Caroline L. Relton[7], James Grinham[8], Kit-Yi Leung[8], Darren Partridge[8], Alexis Robinson[8], Victoria Stone[8], Peter Gustavsson[9], Philip Stanier[8], Andrew J. Copp[8], Nicholas D.E. Greene[8,*], Teiji Tominaga[2], Yoichi Matsubara[1] and Shigeo Kure[1,3,*]

[1]Department of Medical Genetics, [2]Department of Neurosurgery and [3]Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan, [4]Institute for Enzyme Research, University of Tokushima, Tokushima, Japan, [5]Tanemura Women's Clinic, Nagoya, Japan, [6]Department of Public Health, Chiba University School of Medicine, Chiba, Japan, [7]Human Nutrition Research Centre, Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK, [8]Institute of Child Health, University College London, London, UK and [9]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

Received October 26, 2011; Revised November 25, 2011; Accepted December 6, 2011

Neural tube defects (NTDs), including spina bifida and anencephaly, are common birth defects of the central nervous system. The complex multigenic causation of human NTDs, together with the large number of possible candidate genes, has hampered efforts to delineate their molecular basis. Function of folate one-carbon metabolism (FOCM) has been implicated as a key determinant of susceptibility to NTDs. The glycine cleavage system (GCS) is a multi-enzyme component of mitochondrial folate metabolism, and GCS-encoding genes therefore represent candidates for involvement in NTDs. To investigate this possibility, we sequenced the coding regions of the GCS genes: *AMT*, *GCSH* and *GLDC* in NTD patients and controls. Two unique non-synonymous changes were identified in the *AMT* gene that were absent from controls. We also identified a splice acceptor site mutation and five different non-synonymous variants in *GLDC*, which were found to significantly impair enzymatic activity and represent putative causative mutations. In order to functionally test the requirement for GCS activity in neural tube closure, we generated mice that lack GCS activity, through mutation of *AMT*. Homozygous *Amt*$^{-/-}$ mice developed NTDs at high frequency. Although these NTDs were not preventable by supplemental folic acid, there was a partial rescue by methionine. Overall, our findings suggest that loss-of-function mutations in GCS genes predispose to NTDs in mice and humans. These data highlight the importance of adequate function of mitochondrial folate metabolism in neural tube closure.

INTRODUCTION

Neural tube defects (NTDs), such as spina bifida and anencephaly, are severe birth defects that result from failure of closure of the neural folds during embryonic development (1). Although NTDs are among the commonest birth defects in humans, the causes are still not well understood. This is most likely due to their complex, multifactorial causation

*To whom correspondence should be addressed at: Neural Development Unit, UCL Institute of Child Health, Guilford Street, London, WC1N 1EH, UK. Email: n.greene@ucl.ac.uk (N.D.E.G.); Department of Pediatrics, Tohoku University School of Medicine, 1-1 Seiryomachi, Aobaku, Sendai 980-8574, Japan. Email: kure@med.tohoku.ac.jp (S.Ku.)

which is thought to involve contributions from both genetic and environmental factors (2–4). The potential complexity of NTD genetics is illustrated by the fact that more than 200 different genes give rise to NTDs when mutated in mice (5,6). Moreover, inheritance patterns in humans suggest a multigenic model in which an affected individual may carry two or more risk alleles, which by themselves may be insufficient to cause NTDs (2).

Folate one-carbon metabolism (FOCM) is strongly implicated as a determinant of susceptibility to NTDs since suboptimal maternal folate status and/or elevated homocysteine are established risk factors, whereas periconceptional maternal folic acid supplementation can reduce the occurrence and recurrence of NTDs (7,8). Nevertheless, the precise mechanism by which folate status influences NTD risk remains elusive (7,9). FOCM comprises a network of enzymatic reactions required for synthesis of purines and thymidylate for DNA synthesis, and methionine, which is required for methylation of biomolecules (Fig. 1A) (9). In addition to the cytosol, FOCM also operates in mitochondria, supplying extra one-carbon units to the cytosolic FOCM as formate (Fig. 1A) (10).

Genes that are functionally related to folate metabolism have been subjected to intensive genetic analysis in relation to NTD causation, principally through association studies (reviewed in 3,4,11). In the most extensively studied gene, *MTHFR*, the c.677C>T SNP is associated with NTDs in some, but not all, populations. However, other FOCM-related genes have largely shown non-significant or only mild associations. Given the apparently complex inheritance of the majority of human NTDs, many association studies have been hampered by limitations on sample size. Moreover, although positive associations have been noted for other genes including *DHFR*, *MTHFD1*, *MTRR* and *TYMS* (12,13), these have not been replicated in all populations, and additional studies are required. The hypothesis that genetically determined abnormalities of folate metabolism may contribute to NTD susceptibility is supported by the observation of defects of thymidylate biosynthesis in a proportion of primary cell lines derived from NTDs (14). However, these defects do not correspond with known polymorphisms in FOCM-related genes. Overall, it appears likely that genetic influences on folate metabolism remain to be identified in many NTDs.

A potential link between mitochondrial FOCM and NTDs was suggested by the finding of an association of increased NTD risk with an intronic polymorphism in *MTHFD1L* (15). Another component of mitochondrial FOCM, the glycine cleavage system (GCS), acts to break down glycine to donate one-carbon units to tetrahydrofolate (THF), generating 5,10-methylenetetrahydrofolate (methylene-THF; Fig. 1B) (16,17). The GCS consists of four enzyme components, each of which is required for the glycine cleavage reaction (18,19). The components—glycine dehydrogenase (decarboxylating) (GLDC; P-protein), aminomethlyltransferase (AMT; T-protein), glycine cleavage system protein H (GCSH; H-protein) and dihydrolipoamide dehydrogenase (DLD; L-protein)—are encoded by distinct genes: *GLDC*, *AMT*, *GCSH* and *DLD*, respectively. The functions of *GLDC*, *AMT* and *GCSH* are specific to the GCS, whereas *DLD* encodes a housekeeping enzyme. GCS components have been found to be abundantly expressed in the neuroepithelium during embryogenesis in the rat (20).

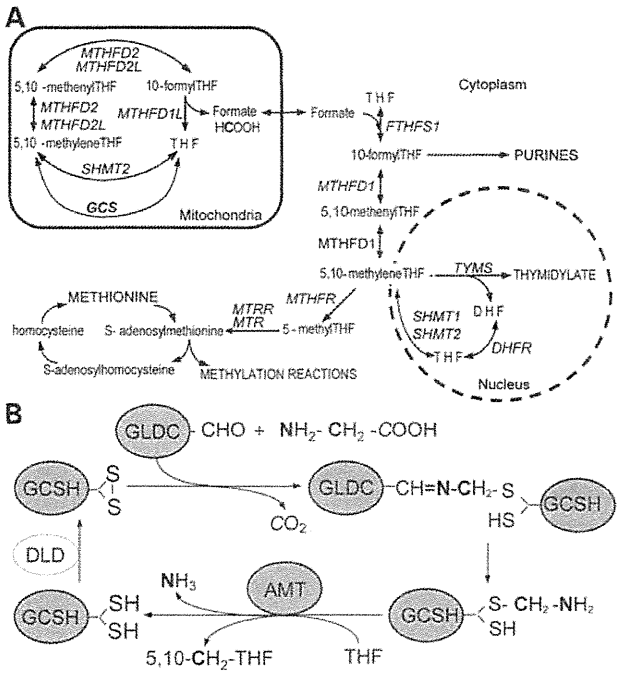

Figure 1. Schematic diagrams summarizing the key reactions of folate-mediated one-carbon metabolism and the GCS. (**A**) Folates donate and accept one-carbon units in the synthesis of purines, thymidylate and methionine. Mitochondrial FOCM supplies one-carbon units to the cytoplasm via formate. The GCS is a key component of mitochondrial FOCM that breaks down glycine and generates 5,10-methylene-THF from THF. Genes encoding enzymes for each reaction are indicated in italics. DHF, dihydrofolate; THF, tetrahydrofolate. (**B**) Summary of the GCS. The glycine cleavage reaction is catalysed by the sequential action of four individual enzymes: GLDC, GCSH, AMT and DLD. The first three of these (shaded grey) are specific to the GCS. Glycine is broken down into CO_2 and NH_3, and donates a one-carbon unit (indicated in bold) to THF, generating 5,10-methylene-THF. The other carbon in glycine (indicated in italics) enters CO_2.

We hypothesized that modulation of GCS activity has the potential to influence efficacy of cellular FOCM during the period of neural tube closure and, hence, susceptibility to NTDs. Therefore, in the current study, we screened genes encoding GCS components for possible mutations in NTD patients and controls. We tested variant proteins for loss of function by enzymatic assay and mice lacking GCS function were generated, to test the effect on embryonic development.

RESULTS

The hypothesis that genes of the GCS represent candidates for involvement in NTDs prompted us to screen for potential mutations in patient samples. Coding exons of *AMT* (9 exons), *GCSH* (5 exons) and *GLDC* (25 exons) were sequenced in a total of 258 NTD patients comprising cohorts from Japan, the UK and Sweden. Each of the major categories of NTDs was represented among study samples, including anencephaly ($n = 38$), spina bifida ($n = 198$) and craniorachischisis ($n = 22$).

Table 1. Nucleotide changes in NTD patients and controls identified by exon sequencing of AMT, GLDC and GCSH

Location	Nucleotide change	Effect	Number of mutation carriers in UK cohorts		Number of mutation carriers in the Japanese cohort		Number of mutation carriers in the Swedish cohort		Variant GLDC enzyme activity[a]
			NTD group (type[b]) ($n = 166$)[c]	Control group ($n = 189$)[c]	NTD group (type[b]) ($n = 14$)[c]	Control group ($n = 36$)[c]	NTD group (type[b]) ($n = 76$)[c]	Control group ($n = 145$)[c]	
AMT									
Exon 2	c.103A>C	p.R35R	0	1	0	0	0	—	
	c.214A>G	p.T72A	0	0	0	1	0	—	
Exon 6	c.623C>A	p.A208D	0	2	0	0	0	—	
	c.631G>A	p.E211K[d]	2 (SBA)	0	0	0	1	—	
	c.589G>C	p.D197H	0	0	1 (An)	0	0	—	
Exon 7	c.825T>A	p.N275K	0	1	0	0	0	—	
	c.850G>C	p.V284L	1 (SBA)	0	0	0	0	—	
GLDC									
Exon 1	c.52G>T	p.G18C	2 (SBO/SBA)	2	0	0	2 (SBA)	2	84%
Exon 5	c.668C>G	p.P223R	0	0	0	1	0	—	92%
Exon 12	c.1508A>C	p.E503A	1 (SBA)	0	0	0	0	0	—
	c.1512G>C	p.E504D	1 (SBA)	0	0	0	0	0	99%
	c.1519G>C	p.G507R	1 (An)	0	0	0	0	0	17%
	c.1525C>G	p.P509A[e]	1 (An)	0	0	0	0	0	41%
	c.1550G>C	p.S517T	0	0	0	0	1 (SBA)	0	—
	c.1570G>C	p.V524L	1 (SBA)	0	0	0	0	0	34%
Exon 14	c.1705G>A	p.A569T[f]	3 (SBA/SBO/SBO)	1	0	0	1 (SBA)	0	40%
Exon 17	c.1953T>C	p.H651H	0	1	0	0	0	—	—
Exon 19	c.2203G>T	p.V735L	0	2	0	0	0	—	81%
Intron 19	c.2316-1G>A	splice	1 (SBA)	0	0	0	0	—	
Exon 20	c.2380G>A	p.A794T	2 (SBASBA)	0	0	0	2 (SBA)	2	88%
	c.2406G>A	p.A802A	1 (An)	0	0	0	0	0	—
Exon 21	c.2474G>A	p.G825D	0	0	1 (An)	0	0	—	24%
	c.2487C>T	p.A829A	0	1	0	0	0	—	—
	c.2565A>C	p.A855A	1 (An)	0	0	0	0	—	—
Exon 23	c.2746C>T	p.L916L	1 (Crn)	0	0	0	0	—	—
Exon 25	c.2964G>A	p.R988R	0	0	0	0	1 (SBA)	0	—
	c.2965A>G	p.I989V	0	1	0	0	0	0	130%
GCSH									
Exon 1	c.53C>T	p.A18V	1 (An)	1	0	0	—	—	

All nucleotide changes were found in heterozygous form. One individual carried c.52G>T and c.1705G>A in *GLDC*, whereas no other individuals carried more than one of the nucleotide changes listed here. Eight silent polymorphisms and four missense variants present in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) are not listed in this table and include: *AMT*: c.954G>A (p.R318R, rs11715915); *GLDC*: c.249G>A (p.G83G, rs12341698), c.438G>A (p.T146T, rs13289273), c.501G>A (p.E167E, rs13289273), c.660C>T (p.L220L, rs2228095), c.666T>C (p.D222D, rs12004164), c.671G>A (p.R224H, rs28617412) and c.1384C>G (p.L462V, rs73400312); and for *GCSH*: c.62T>C (p.S21L, rs8052579), c.90C>G (p.P30P, rs8177847), c.159C>T (p.F53F, rs177876), c.218A>G (N73S, rs8177876), c.252T>C (Y84Y, rs8177907) and c.261C>G (L87L, rs8177908). Grey shading indicates loss-of-function mutations, based on enzymatic activity in the *in vitro* expression study or splicing defect.

[a]Residual enzymatic activity of GLDC mutant protein is expressed as %activity of the wild-type enzyme (Fig. 2).

[b]SBA, spina bifida aperta; SBO, spina bifida occulta; An, anencephaly; Crn, craniorachischisis.

[c]Total number of UK, Japanese or Swedish NTD patients.

[d]This variant was previously established as likely to be a non-functional polymorphism by segregation in an NKH family (21).

[e]A biochemical test of folate metabolism, the dU suppression test, was previously performed on primary fibroblasts derived from this patient and showed a defect of thymidylate biosynthesis to be present (14).

[f]p.A569T has previously been reported as a pathogenic mutation in a patient with typical NKH (21).

In *AMT*, we identified two novel sequence variants predicted to result in non-synonymous missense changes, c.589G>C (D197H) and c.850G>C (V284L), in anencephaly and spina bifida patients, respectively, from the UK cohort (Table 1). Neither variant was present in 526 UK or 36 Japanese control subjects or in the SNP databases dbSNP and 1000 Genomes. An additional missense variant, E211K, was also identified in three spina bifida patients, two from the UK and one from Sweden. Causative mutations in *AMT* have been found previously in an autosomal recessive inborn error of metabolism, non-ketotic hyperglycaemia (NKH) (17). The E211K variant had previously been identified in an NKH family but was established as likely to be a non-functional polymorphism by segregation (21). Therefore, this variant is considered unlikely to be causally related to NTDs.

Exon sequencing of *GCSH* revealed eight single-base substitutions, one of which (c.53C>T, p.A18V) was a novel change found in both an NTD and a single control (Table 1). The others all corresponded to known SNPs, which did not suggest a role for *GCSH* in NTDs.

Next we turned our attention to *GLDC*, in which we identified 27 single-base substitutions (Table 1), including 11 silent nucleotide changes, 15 non-synonymous changes and a splicing acceptor variant of intron 19 (c.2316-1G>A). The

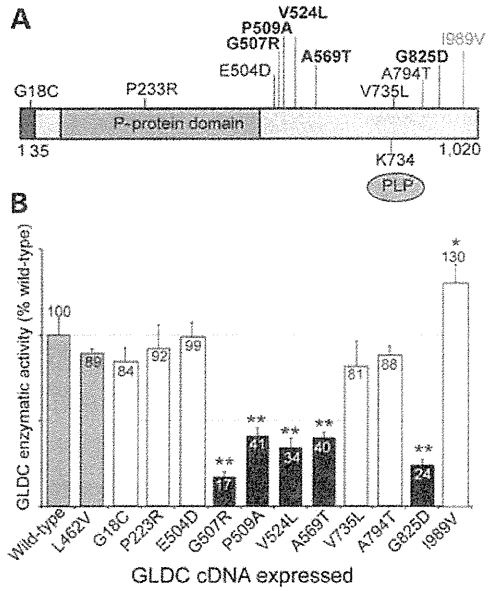

Figure 2. Characterization of *GLDC* missense mutations identified through DNA sequence analysis. (**A**) The schematic represents the 1020 amino acid residue GLDC polypeptide with the positions of the identified missense variants indicated. Mutations conferring significantly reduced activity (**B**) are indicated in bold. The leader peptide for mitochondrial import (shaded black) and the lysine 754-binding site for the co-factor pyridoxal phosphate (PLP) are indicated (49). (B) Enzymatic activity of GLDC missense variants. Expression vectors with wild-type and mutant *GLDC* cDNAs were transfected into COS7 cells for the evaluation of GLDC activity, which is expressed as relative activity (%) of cells expressing wild-type cDNA (shaded grey). The L462V GLDC enzyme (shaded grey) was tested as an example of a normally occurring variant (rs73400312). Variant proteins whose activities were significantly diminished compared with wild-type are indicated by black shading. The I989V variant, identified in a control parent, showed significantly elevated activity. Values are given as mean ± SD of triplicate experiments ($^{*}P < 0.05$; $^{**}P < 0.01$, compared with wild-type).

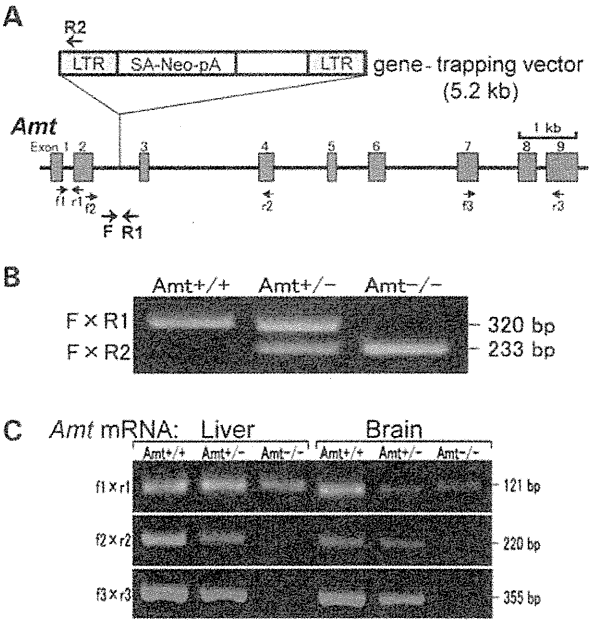

Figure 3. Generation of *Amt* knockout mouse by gene trapping. (**A**) The location of the gene-trap vector in *Amt* intron 2 in the ES cell line OST181110 was determined by inverse PCR. Mice carrying this mutation were generated using standard methods of blastocyst microinjection with OST181110 ES cells to generate chimeras, and germ-line transmission. LTR, long terminal repeats; SA, splicing acceptor site; Neo, neomycin phosphotransferase gene; pA, polyadenylation sequence. (**B**) For genotyping, mouse genomic DNA was subjected to allele-specific amplification with F, R1 and R2 primers (Supplementary Material, Table S1). A genomic fragment of 320 bp was amplified from the wild-type allele, whereas a 233 bp fragment was amplified from the *Amt*-mutant allele. (**C**) RT-PCR analysis of *Amt* mRNA expressed in the brain and liver of *Amt*-mutant mice. Primers in exon 1–2 generated a 121 bp band irrespective of mouse genotypes. RT-PCR in which either one (f2-r2) or both (f3-r3) primers were located in exons 3′ to the insertion site produced 220 and 355 bp cDNA fragments, respectively, in $Amt^{+/+}$ and $Amt^{+/-}$ mice, but not in $Amt^{-/-}$. The *Amt* mRNA in mice carrying the trap vector was, therefore, aberrantly spliced at the end of exon 2, resulting in truncation of *Amt* mRNA in $Amt^{-/-}$ mice.

latter is deduced to abolish normal splicing of the *GLDC* mRNA, with predicted skipping of exon 19 resulting in loss of the reading frame. Among the 15 missense variants identified in *GLDC*, 5 were unique to the NTD group, being absent from all 562 control individuals as well as from the SNP databases. A further three novel variants were found only in controls, whereas the remainder were found in both NTDs and controls, and included previously reported SNPs.

We investigated the possible functional effects of *GLDC* missense variants by expressing wild-type and mutant cDNA constructs in COS7 cells, followed by enzymatic assay of GLDC activity involving a decarboxylation reaction using [1-^{14}C]glycine (22). Twelve GLDC variants were tested, including those that were unique to NTD patients and, therefore, hypothesized to be potentially pathogenic (Fig. 2). The L462V variant, which corresponds to a known SNP (rs73400312), was included as an example of a known normally occurring form. Five of the missense changes, G507R, P509A, V524L, A569T and G825D, resulted in a significant reduction in GLDC activity compared with the wild-type protein ($P < 0.001$). Notably, all five of these deleterious variants were present solely in NTD cases, whereas none of the variants that were unique to controls (P223R, V735L and I989V) showed loss of enzymatic function. In the case of G18C and A794T, which occurred in both NTDs and controls, there was no significant loss of enzymatic activity, suggesting that these are unlikely to be causative mutations.

Having identified putative mutations in *AMT* and *GLDC* in NTD patients, we hypothesized that loss of GCS function could predispose to development of NTDs. In order to directly test the functional requirement for GCS activity in neural tube closure, we generated mice that lacked GCS activity, using a gene trap (OmniBank, OST181110) of the *Amt* gene. The vector was located in intron 2, resulting in a truncated transcript that lacked exons 3–9 (Fig. 3). The efficacy of the gene-trap vector in trapping expression of *Amt* (Amt^{-}) was confirmed by RT-PCR analysis (Fig. 3). Heterozygous $Amt^{+/-}$ mice were viable and fertile and exhibited no obvious malformations. Homozygous $Amt^{-/-}$ mice were not observed among post-natal litters from heterozygote intercrosses, and so fetuses were examined at embryonic day (E) 17.5. Strikingly, 87% of $Amt^{-/-}$ fetuses (34 out of 39) exhibited NTDs, whereas no malformations were observed in $Amt^{+/+}$ ($n = 33$) or $Amt^{+/-}$