recurrent fever and systemic inflammation. CINCA/NOMID is the most severe clinical phenotype in the spectrum of cryopyrin-associated periodic syndromes (CAPS), which also include two less severe but phenotypically similar syndromes, familial cold autoinflammatory syndrome (FCAS; MIM #120100), and Muckle–Wells syndrome (MWS; MIM #191900). CAPS are caused by mutations in the *NLRP3* gene, which is a member of the Nod-like receptor (NLR) family of the innate immune system.[4–6]

Approximately 60% of CINCA/NOMID patients carry heterozygous germline missense mutations in the *NLRP3* coding region (mutation-positive patients).[7] More than 80 different disease-causing mutations have been reported to date.[8] However, the remaining clinically diagnosed CINCA/NOMID patients (~40%) show no heterozygous germline *NLRP3* mutation based on conventional DNA sequencing-based genetic analyses (mutation-negative patients). In a previous international collaborative study, we found that there was a high incidence of somatic *NLRP3* mosaicism in mutation-negative CINCA/NOMID patients worldwide.[9] The level of mosaicism ranges from 4.2 to 35.8% (median = 10.2%). Rapidly diagnosing somatic *NLRP3* mosaicism is important to ensure proper treatment. However, the conventional approach used to identify somatic mosaicism of the *NLRP3* gene is time and labour intensive due to the subcloning of the *NLRP3* exon polymerase chain reaction (PCR) products, hereafter designated as amplicons, followed by capillary DNA sequencing of more than 100 subclones for each patient. Thus, this approach is not suitable to routinely diagnose somatic mosaicism of the *NLRP3* gene and additional labour and time will be required to reliably identify somatic mosaicism that occurs at a lower rate. The aim of the present study was to establish a new method that can be used to reliably diagnose somatic mosaicism using the *NLRP3* gene as a model. Massively parallel DNA sequencing (MPS) technology is an obvious method of choice to identify somatic mosaicism, and this approach has been already reported by other groups.[10–12] However, a well-known caveat of MPS is the high rate of sequencing errors, which cannot be disregarded when identifying low-level somatic mosaicism. To our knowledge, there have been no reports of a reliable method to discriminate MPS sequencing errors from somatic mosaicism with statistical confidence.

In this study, we first analysed the patterns of sequencing errors in *NLRP3* coding exons at a single-residue resolution by MPS using a Roche 454 GS-FLX sequencer and then constructed an error rate map for each base position in the *NLRP3* exons. Based on the error rate map, we could formulate a discrimination pipeline of somatic mosaicism from sequencing errors and thereby detect new somatic mosaicism in mutation-negative CINCA/NOMID patients, whose somatic mutations were subsequently confirmed by subcloning and Sanger sequencing. This approach can also be generally used to identify low-level somatic mosaicism in other genes.

2. Patients and methods

2.1. *Patients and DNA materials*

Patients were clinically diagnosed with CAPS by their referring physicians and the *NLRP3* gene was examined using the conventional Sanger sequencing method. DNA samples were obtained from Japanese *NLRP3* somatic mosaic patients ($n = 5$) who have been previously described,[9,13] CAPS patients ($n = 5$) with heterozygous *NLRP3* mutations, and healthy donors ($n = 50$). Genomic DNA samples from mutation-negative CINCA/NOMID patients ($n = 10$) were obtained from the National Institute of Health, Bethesda, USA. To generate DNA samples with no mosaicism, we constructed a set of subcloned plasmids containing each exon and its flanking intronic regions in the *NLRP3* gene from healthy donor genomic DNA using a Topo TA cloning kit (Invitrogen, San Diego, CA, USA). The cloned plasmids containing each exon and the flanking regions were validated by Sanger sequencing. Written informed consent was obtained from all the patients and their families. The study was approved by the ethical committees of Kyoto University and Kazusa DNA Research Institute and was conducted in accordance with the Helsinki Declaration.

2.2. *MPS of* NLRP3 *gene amplicons*

Genomic DNA samples were extracted from whole blood or peripheral blood mononuclear cells as previously described. We used a two-step PCR assay and pooled sample libraries for MPS. To cover the entire *NLRP3* coding exonic regions and flanking intronic regions, 14 amplicons were designed to be as long as an average read length for a 454 GS-FLX sequencer (up to 450 bases) and then amplified from each genomic DNA sample (Fig. 1A). The sequences of the PCR primers that were used to generate these 14 amplicons are provided in Supplementary Table S1. The upper and lower amplicon-specific primer sequences were flanked by common 15-base adapter sequences (TGTAAAACGACGGCC and GGAAA CAGCTATGAC for the upper and lower primers, respectively) at the 5′ end in order to fuse the primer-binding sequence for MPS in the second-step PCR. The first PCR amplifications were performed in 50-μl reactions using 30 ng of genomic DNA, 1× PrimeSTAR GXL buffer, 0.2 mM of each dNTP,

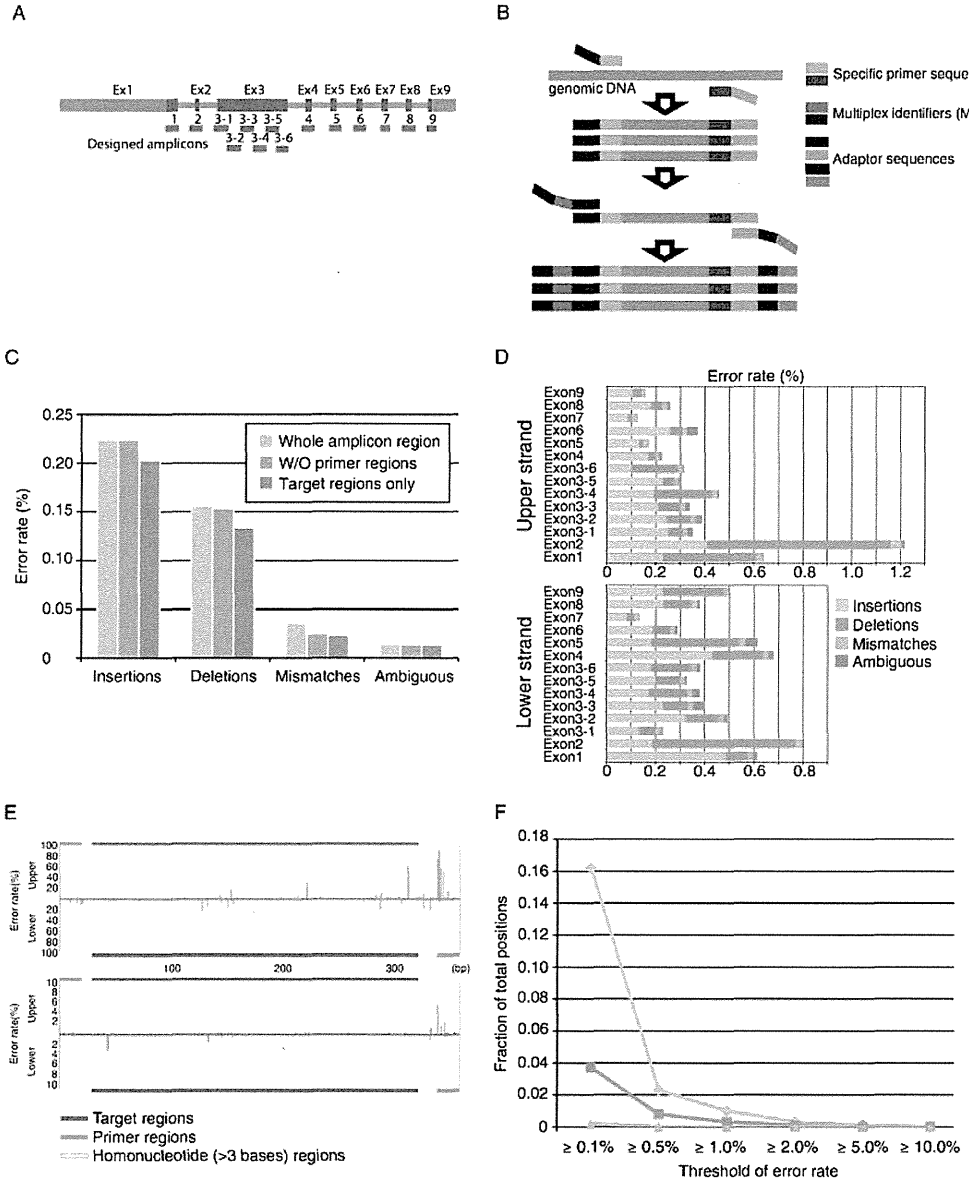

Figure 1. The amplicon analysis for *NLRP3* exons and its error rate. (A) Exon–intron structure of the *NLPR3* gene. Thick and thin rectangles depict exons and introns, respectively. Blue thick rectangles indicate the CDS region. The 14 designed amplicons (red) for nine exons are shown under the exon–intron structure. (B) Amplicon design schema. (C) Error rate for each error category in the region of entire amplicon (pale blue), that without designed primer regions (light blue), and the target regions (CDS + flanking intron; dark blue), respectively. (D) Strand-wise error rate for each amplicon. (E) Error rates along the amplicon sequence of exon 1 in each strand for insertions and deletions in the upper panel and mismatches and ambiguous base calls in the lower panel. The orange and blue lines depict the primer and target regions, respectively. The yellow shaded area depicts the homonucleotide ($n > 3$) region. The colour representation for the bars is the same as (D). (F) Co-occurrence error rate in both strands. The fraction of positions where a certain error occurred with the error rate for insertions, deletions, and mismatches. The colour representation is the same as in (D) and (E).

12.5 pmol of each forward and reverse primer, and 1.25 U of PrimeSTAR GXL DNA polymerase (Takara Bio, Shiga, Japan). The thermal cycling profile consisted of an initial denaturation step at 98°C for 1 min, followed by 28–32 cycles of 10 s denaturation at 98°C, 15 s of annealing at 60°C, and a 30 s extension at 68°C. The lengths of the PCR products ranged from 291 to 421 bp. The second PCR amplifications were performed using primers with adapter sequences at the 3′ end and Multiplex Identifier (MID) sequences at the 5′ end (Fig. 1B), which was used as a tag for each sample. The PCR reactions were performed in 50-μl volumes using 0.5 μl of the first PCR products, 1× PrimeSTAR GXL buffer, 0.2 mM of each dNTP, 12.5 pmol of each forward and reverse primer, and 1.25 U of PrimeSTAR GXL

DNA polymerase to attach the anchor sequences for MPS. The thermal cycling profile consisted of an initial denaturation step at 98°C for 20 s, followed by 20 cycles of 10 s denaturation at 98°C, 15 s of annealing at 60°C, and a 40 s extension at 68°C.

After confirming the amount and integrity of the PCR products by agarose gel electrophoresis, we mixed virtually equal amounts of the respective PCR amplicons that were generated using the same genomic DNA and applied the samples to a 454 Genome Sequencer (GS)-FLX system (Roche Diagnostics Corp., USA). All amplicons were amplified by emPCR and sequenced together in a multiplex fashion. MPS on this platform was performed as instructed by Roche. The sequencing reads from each of the pooled libraries were identified by their MID tags.

2.3. *Sequence data analysis*

The sequence read data were generated using GS RunProcessor ver.2.5.3 with default settings. Reads were sorted according to the MID tag sequences and were mapped to the reference amplicon sequences using the BLAT program[14] with the '-fine' option. In order to identify positions where the bases in a read differed from those in the reference sequence, each read was aligned to its reference sequence with the dpAlign module in the BioPerl package (http://www.bioperl.org/). The 454 pyrosequencing-related errors were categorized as insertions, deletions, mismatches, or ambiguous base calls. When aligning sequences, insertions/deletions are allocated based on the sequence context and strand orientation. To eliminate alignment artefacts due to insertion/deletion positions, the lower strand reads were converted to the reverse complement sequence, i.e. keeping the same strandness as the upper strand reads, when aligned with the reference sequence. A sequence error was defined as discordance in an equivalent position between the reference and control (from the 49 healthy individuals and a cloned plasmid vector). The error rate for a specified category was defined as the number of errors divided by the total number of bases in a read. The error rates of a base position on each strand were calculated from 50 control samples.

2.4. *Confirmation of somatic mosaicism of the* NLRP3 *gene by subcloning and subsequent capillary DNA sequencing*

To confirm the somatic mutational frequency that was identified based on the 454 sequencing data, we subcloned the PCR products and performed capillary DNA sequencing as previously described.[9] A Topo TA cloning kit (Invitrogen, San Diego, CA, USA) was used to subclone each of the 14 amplicons.

2.5. *Functional analysis*

To determine whether the identified *NLRP3* mutants are disease-causing, we assessed both ASC [apoptosis-associated speck-like protein containing a caspase recruitment domain; PYCARD, an approved symbol from the HUGO Gene Nomenclature Committee (HGNC) database]-dependent NF-κB activation in HEK293FT cells and transfection-induced cell death in THP-1 cells, a human monocytic cell line, as previously described.[9,13,15] cDNAs encoding carboxy-terminal green fluorescent protein (GFP)-tagged NLRP3 and its mutants were subcloned into pcDNA5/TO (Invitrogen). Before being introduced into THP-1 cells (10^6) using a Cell Line Nucleofector Kit V (Amaxa Biosystems, Cologne, Germany), phorbol myristate acetate (10 ng/ml) was added to enhance transient expression of *NLRP3* gene with minimizing spontaneous cell death.[15] Four hours after the introduction of plasmids (0.5 μg), cell death of GFP-positive THP-1 cells was measured by flow cytometry.

Expression plasmids for NLRP3 and ASC in the pEF-BOS vector background have been previously described.[13] HEK293FT cells (10^5) were transfected using TransIT-293 Transfection Reagent (Milus Bio, Madison, WI, USA) with an NF-κB reporter construct (pNF-κB-luc; 20 ng; BD Biosciences Clontech, Palo Alto, CA, USA), an internal control construct (pRL-TK; 5 ng; Toyo Ink, Tokyo, Japan), and wild-type or mutant NLRP3 expression plasmid (20 ng) in the presence or absence of ASC expression plasmid (20 ng). The amounts of total plasmid DNA used for transfection experiments were kept constant by adding pEF-BOS vector DNA. Twenty-four hours later, the transfected cells were harvested and subjected to dual luciferase assay by which the ability of each construct to induce NF-κB activation was assessed as previously described.[9]

3. Results

3.1. *Construction of base- and strand-specific error rate maps of* NLRP3 *exons from the MPS data of 50 control samples*

Errors in sequence reads generated by a Roche 454 GS-FLX sequencer are not randomly distributed along the sequence and depend on various factors.[16] Although this is a well-known characteristic of 454 sequencing, the occurrence pattern of these errors has not been explored in detail simply because these sequencing errors are considered noise that can be filtered out in most cases. However, it is highly critical to understand the occurrence pattern of sequencing errors on the MPS platform because low-level somatic mosaicism might appear at a rate close to that of sequencing errors. To address this, we collected

~1 million sequence reads using the 454 GS-FLX sequencer for 14 amplicons of *NLRP3* exons from 50 control samples that were thought to be free from somatic mosaicism, and ~94% of those reads were mapped to one of the reference *NLRP3* exon sequences. The number of sequencing depths for each amplicon of each sample on each strand was between 65 and 2139 (mean = 565.3, Supplementary Table S2). We found that the average error rate for each mutation category (insertion, deletion, mismatch, and ambiguous base calls) at each base position on each strand of the amplicons in the control samples was 0.22, 0.16, 0.036, and 0.014%, respectively (Fig. 1C). These values were consistent with those reported in a recent study on the error rates with 454 sequencing data.[16] The sequencing error in the 454 GS-FLX system tends to occur at the beginning and end of the reads,[11,16] and we confirmed this trend in our amplicon sequencing data (Supplementary Fig. S1). Moreover, after removing the end regions of the read sequences, we found that the error rates of the target regions for each category were 0.20, 0.134, 0.023, and 0.014%, respectively (Fig. 1C and Supplementary Table S3). When generating the amplicon sequences for the *NLRP3* exons, the target sequence (CDS region and flanking intron in 10-bp length) was designed to be 300–400 bp and not adjacent to primer sequences in order to obtain relatively low sequencing error rates (Fig. 1C). However, when the base- and strand-specific error rates of the respective amplicons were compared, we noticed that there were large variations in the error rate among amplicons in a strand-specific manner (Fig. 1D). We further examined the occurrence pattern of sequencing errors, as shown in Fig. 1E; the average sequencing error rates at each base in the 50 control amplicons are shown in a bar graph, where the bars in the upper or lower direction show the sequence error rates at the base position on the upper or lower strand of the amplicons, respectively. As evident in Fig. 1E, the error rates at most residues were low (<1%) with some hotspots for each type of error. Most of the insertion/deletion errors preferentially occurred at a homonucleotide region (yellow regions in Fig. 1E) as previously described,[17] but it was not always the case for all of homonucleotide regions. We could not find any tight relationship between other sequence patterns and the error rate. In addition, there was almost no position where sequencing errors occurred at a similar rate on both strands. This is more clearly shown in Fig. 1F, which indicates the numbers of positions with sequence variations in both strands that were higher than the threshold along the horizontal axis. These results indicate that the sequence errors can be discriminated from real genetic alterations when the sequence is read in both directions. However, it is important to keep in mind that PCR errors are not distinct from real genetic alterations. We did not observe any base substitution at a rate higher than 1% in our experiments (Fig. 1F), and the overall PCR error rate under MPS conditions was lower than 1% as long as a high-fidelity DNA polymerase was used to generate the amplicons.

Because Gilles *et al.*[16] recently reported that the occurrence of sequencing errors using the Roche 454 GS-FLX DNA sequencer depends on various factors, we first examined variations in the sequencing error rates of *NLRP3* exons among samples in the same run. For each mutation category, we found a similar trend in the error distribution rate in the amplicon sequences among the control samples (Supplementary Figs S2–S4). We confirmed that, for almost all residues, the error rate distributions among the 50 control samples fitted a Poisson distribution (data not shown). We next examined the run-to-run variation of the sequencing error rate for *NLRP3* exons. For this purpose, we performed an additional MPS run with seven amplicons (exons 3, 4, and 6) that were newly prepared and compared the number and positions of the sequencing errors between two independent sequencing runs. Out of 1993 base positions in the target regions, there was a low occurrence rate of mismatch errors in both runs and this seemed to fit a Poisson distribution. However, insertion/deletion errors (>1% error rate) were observed at ~100 base positions (<5% in the target regions) in each run, and only a half of these errors were shared between both runs (Table 1).

Table 1. Run-to-run variations in the error occurrence (>1% frequency)

Error category	Upper strand			Lower strand			All[a]
	First run	Second run	Overlap	First run	Second run	Overlap	
Insertions	63	73	42	76	96	52	10
Deletions	36	44	24	29	65	20	2
Mismatches	0	0	0	3	0	0	0
Ambiguous base calls	6	8	6	12	10	10	0

[a]The number of positions where the error rates in each category were commonly >1% for both strands in two independent runs.

This indicated that the occurrence of insertion/deletion errors was considerably affected by the run conditions (probably due to variations in the absolute signal strengths of pyrosequencing). Thus, as previously reported, the detection of insertion/deletion mutations by MPS on the 454 GS-FLX system was quite error-prone at least at a limited number of residues. However, the results also implied that false-positive mosaic mutations could be avoided by considering the sequencing data for both strands because these run-dependent insertion/deletion errors occur only in a single strand. Taken together, we conclude that the obtained sequence error map is stable and sufficiently robust to discriminate substitution sequencing errors from low-level mosaicism.

3.2. *Discrimination formula for detection of somatic mosaicism with statistical confidence*

We next examined known SNPs, known heterozygous mutations and somatic mosaic mutations of CAPS patients using MPS. All of these variations appeared on both strands at the expected allele frequencies as shown in Fig. 2, again indicating that filtering the strand-specific sequence variations is unlikely to eliminate real genetic variations.

Based on the experimentally observed sequencing errors with the 454 GS-FLX system described above, we established a discrimination formula to detect low-level somatic mosaicism as follows. In previous studies, the number of reads with the sequence error of a certain category in a sequence position was modelled based on the Poisson distribution with two parameters λ and k where the expected number of reads containing an error and the observed number of reads containing a sequence alteration, respectively, are as shown below[18]:

$$\mathrm{Pois}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \quad (1)$$

This model assumes that the error rate is constant across the different sequence regions but our data described above pointed out that the sequence error rate varies with the sequence content.[19] Thus, we introduced a position- and strand-specific error rate $q_{i,j,d}$ for a certain error category j in amplicon position i with strand d based on the sequencing data from 50 control samples. With the error rate $q_{i,j,d}$, the upper probability (P) that the number of reads (R) with a certain sequence alteration of category j in position i is equal or greater than the number of observed reads r out of N reads with a sequenced direction d for an unknown sample was defined as:

$$P(R \geq r_{i,j,d} | \lambda_{i,j,d}) = 1 - \sum_{k=0}^{r-1} \frac{\lambda_{i,j,d}^k e^{-\lambda_{i,j,d}}}{k!}, \quad (2)$$

where, $\lambda_{i,j,d} = N_{i,d} \times q_{i,j,d}$.

For the mismatch error rate, we did not consider the type of base substituted in an amplicon position in this study. We took $(1 - P)$ as a measure of the statistical confidence of the data and conventionally set a threshold of the statistical confidence to be 99.9%. In other words, if P-value was <0.001, the sequence alteration was considered to be a real sequence variation, not an error. For the final identification of real genetic variation with low-level somatic mosaicism, we determined that both of the P-values for the ith residue in the upper and lower strands must be smaller than the threshold.

To evaluate the lower detection limit for the allele frequencies of somatic mosaicism based on the statistical formulation shown above, we generated a series of known allele frequencies by diluting DNA from CAPS patients carrying heterozygous *NLRP3* mutations (c.1043C>T, c.1316C>T, and c.1985T>C) with DNA from normal donors carrying the wild-type *NLRP3* gene. In the dilution series, the mutant allele frequencies were adjusted to be 10, 5, 3, 2, 1, and 0.5% (Table 2). The data indicated that somatic mosaicism at these sites and at an allele frequency $\geq$1% could be convincingly detected with statistical significance ($P < 0.001$) if more than 350 reads for each strand were obtained for an amplicon. We also applied this statistical method to detect somatic mosaicism in patients with known low-level mosaic mutations described above and confirmed that all of

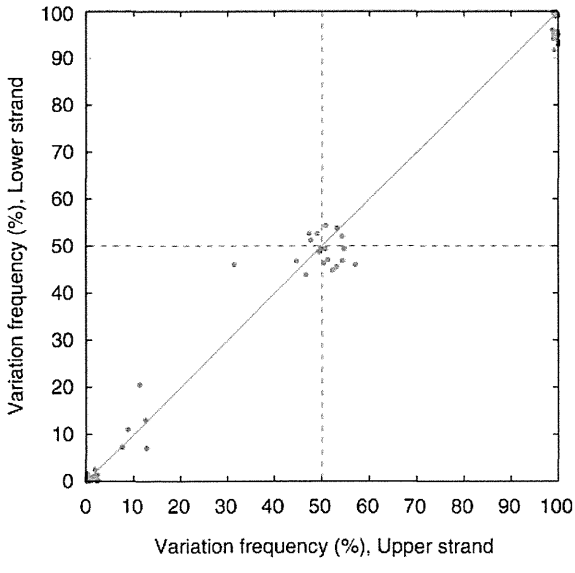

Figure 2. Scatter plot of the observed frequency variation in both strands. The colours depict known SNPs (green), heterozygous and mosaic mutations (orange) and errors (grey).

Table 2. Evaluation of the lower detection limit for mosaicism with three sets of dilution series

Mutation	Dilution (%)	Upper strand				Lower strand			
		Total reads	Mutant reads	%Mutant	*P*-value	Total reads	Mutant reads	%Mutant	*P*-value
c.1043C>T; p.Thr348Met	10.0	724	61	8.43	8.62E−130	520	57	10.96	1.73E−117
	5.0	453	24	5.30	2.86E−47	372	15	4.03	1.26E−25
	3.0	876	27	3.08	1.16E−46	757	21	2.77	6.83E−32
	2.0	737	10	1.36	1.05E−14	645	7	1.09	8.68E−09
	1.0	715	9	1.26	4.73E−13	624	4	0.64	1.11E−04
	0.5	1025	7	0.68	1.15E−14	756	3	0.40	3.22E−03[a]
c.1431C>A; p.Asn477Lys	10.0	542	65	11.99	1.22E−113	346	24	6.94	6.84E−49
	5.0	491	30	6.11	1.13E−44	356	17	4.78	2.42E−32
	3.0	487	21	4.31	1.26E−28	374	19	5.08	1.78E−36
	2.0	577	18	3.12	2.78E−22	495	9	1.82	4.57E−14
	1.0	491	4	0.82	9.17E−04	354	5	1.41	7.34E−08
	0.5	483	0	0	NA	424	3	0.71	NA
c.1985T>C; p.Met662Thr	10.0	658	79	12.01	1.13E−179	643	74	11.51	4.64E−167
	5.0	643	31	4.82	2.56E−59	608	33	5.43	9.96E−65
	3.0	777	27	3.48	4.65E−48	704	29	4.12	1.26E−53
	2.0	929	21	2.26	7.59E−34	835	15	1.80	3.92E−23
	1.0	735	17	1.09	2.74E−11	709	9	1.27	4.06E−13
	0.5	702	2	0.29	3.90E−03[a]	590	1	0.17	1.37E−01[a]

[a]Not significant.

Table 3. Potential mosaic mutations detected in patients with unknown mutations

Patient ID	Amplicon #	Variation		% Variation frequency		*P*-value		dbSNP	State
				Forward	Reverse	Forward	Reverse		
P1	Exon3_2	c.907G>C	p.Asp303His	7.12	11.56	3.0E−44	1.7E−84	rs121908153	Known
P2	Exon3_5	c.1699G>A	p.Glu567Lys	5.94	5.79	2.0E−69	8.9E−47	—	Known
P3	Exon3_5	c.1699G>A	p.Glu567Lys	18.28	15.33	0.0E+00	1.0E−312	—	Known
P4	Exon3_2	c.906C>A	p.Phe302Leu	9.78	9.70	1.7E−86	2.2E−122	—	Novel

the mutations could be detected with statistical significance without any false positives (data not shown).

3.3. *Detection and characterization of* NLRP3 *somatic mosaicism using the MPS platform*

To demonstrate the power of this approach in practice, we applied our new pipeline for 10 CINCA/NOMID patients in whom we failed to detect mutations in the *NLRP3* gene using a conventional direct DNA sequencing approach. The mutations detected by the analysis formulated using the MPS platform in this study are listed in Table 3. We successfully identified four out of the 10 patients with *NLRP3* somatic mosaicism, which was confirmed by subcloning and Sanger sequencing. The nucleotide substitutions were as follows (parentheses indicate the corresponding amino acid change): c.907G>C (p.Asp303His), c.1699G>A (p.Glu567Lys) in two patients, and c.906C>A (p.Phe302Leu). The frequencies of mosaicism identified in these patients by the MPS approach were consistent with those that were identified by the subcloning and subsequent capillary DNA sequencing method (data not shown). Both c.907G>C and c.1699G>A variants were reported as CINCA/NOMID-associated mutations in Infevers database (http://fmf.igh.cnrs.fr/ISSAID/infevers/) and in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/).[8]

Because the NLRP3 p.Phe302Leu mutation was novel and not detected in the 50 healthy controls, we performed an *in vitro* functional analysis to see the effect of p.Phe302Leu on the protein function. We used two different *in vitro* transfection experiments,

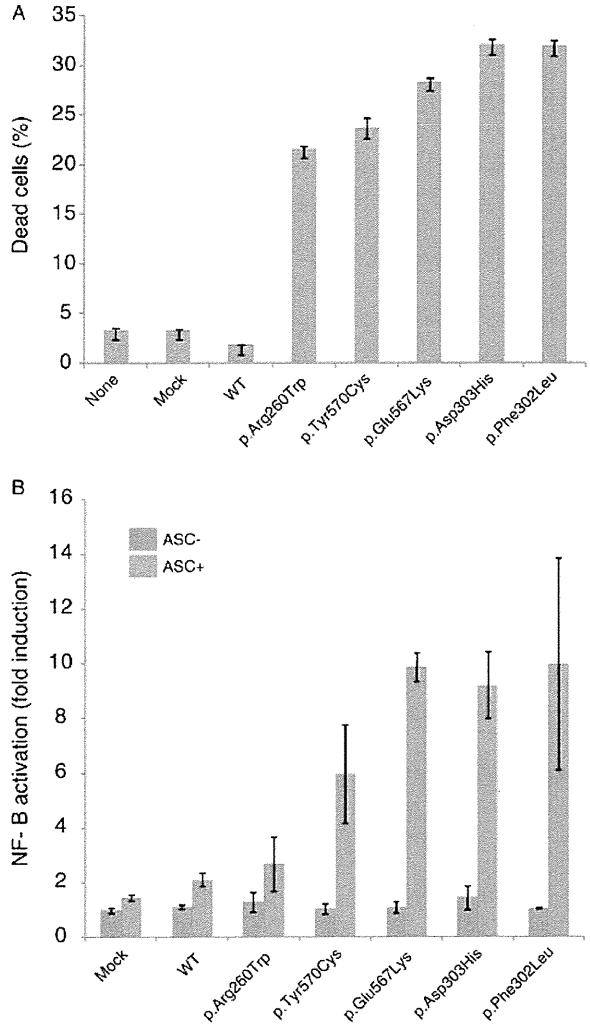

Figure 3. *In vitro* functional analysis of the identified *NLRP3* mosaic mutations. (A) Rapid cell death in transfected THP-1 cells. A GFP-fused wild-type or mutant *NLRP3* was transfected into THP-1 cells and incubated with PMA (10 ng/ml) for 4 h. The percentage of dead cells (7-amino-Actinomycin D [7-AAD]-positive) among the GFP-positive cells is shown. Data represent the means ± SD of triplicate experiments and are representative of two independent experiments. The data for previously reported mutations as well as the mutations found in this study are shown. (B) ACS-dependent NF-κB activation in transfected HEK293FT cells. HEK293FT cells were co-transfected with wild-type or mutant *NLRP3* in the presence or absence of ASC. NF-κB induction is shown as the fold-change compared with cells that were transfected with a control vector without ASC (set equal to one). Values are the means ± SD of triplicate experiments, and the data are representative of three independent experiments. The data for previously reported mutations (p.Arg260Trp and p.Tyr570Cys) and the mutations found in this study are shown. For each mutation, the data obtained in the presence and absence of ASC are shown. These findings identified p.Phe302Leu as a novel disease-causing mutation.

the rapid cell death in transfected THP-1 cells and the ASC-dependent NF-κB activation in transfected HEK293FT cells (Fig. 3A and B, respectively). Both assays clearly showed that p.Phe302Leu was a disease-causing mutation similar to known CINCA/NOMID-associated pathogenic mutations (p.Asp303His and p.Glu567Lys).[9]

4. Discussion

Although the somatic mutation rate at the nucleotide level *in vivo* was difficult to quantitatively measure due to the complexity of the genome and laborious molecular detection processes, recent advances in MPS technologies have allowed us to directly quantitate somatic mutations in human genome.[20–22] The current estimate for the somatic (*de novo*) mutation rate is $1-2 \times 10^{-8}$ residues/generation/haploid, and this estimate is sufficiently low that we would expect to never observe somatic mosaicism in the *NLRP3* gene by chance; although the error rate of the high-fidelity DNA polymerase used to produce the amplicons is two orders of magnitude larger than the somatic mutation rate,[23,24] we could not detect PCR-generated mosaicism higher than 1% in the 454 sequencing error maps. Based on the literature, the single base substitutions are the most frequent type of somatic mutations (~500 times more frequent than short insertions/deletions)[25] and protein-coding sequences are less mutagenic than sequences in non-coding regions, assuming that the somatic mutation spectrum in malignant cells is the same as in normal cells. Somatic mosaicism is thought to result from *de novo* gain-of-function-type mutations that are introduced at a very early and limited stage of development, and it is reasonable to focus our efforts on detecting base substitutions for somatic mosaicism in the *NLRP3* gene.

It is challenging but highly important in many areas of research, such as cancer, to detect low-level somatic mutations, which we designated as somatic mosaicism in this study, from apparently mutation-negative samples by conventional sequencing. Subcloning followed by the capillary DNA sequencing has been a *de facto* standard to identify somatic mosaicism, but this is not the method of choice for routine diagnostics because it is laborious, time consuming, and costly. Thus, it is reasonable for us to explore MPS as a new tool for this purpose. Although previous studies have used MPS technology to detect somatic mosaicism, it was unclear how sensitive this method is to detect a low-level somatic mosaicism using the MPS platform because this platform is generally error-prone. To address this challenge, we developed a new pipeline to detect low-level somatic mosaicism with statistical confidence using base position- and strand-specific error rate maps for the *NLRP3* amplicons to be studied. Whereas the

detection limit of somatic mosaicism depends on the base position and the read depth of the amplicons, the limit of detection could be as low as 1% allele frequency with no false positives for substitutions (the precision is higher than 99.9%). Our error map shows that 98.1% of base positions (3343 out of 3407 target positions) in the *NLRP3* exonic amplicons can be detected with ~1% mosaicism when more than ~350 reads were accumulated for each strand. Although the remaining region (64 base positions out of 3407 target positions) was too error-prone (the error rate ranged from 0.1 to 1.7% in either the upper or lower strand) to detect low-level mosaicism by MPS, medium-level mosaicism (5% or high) could be identified in all base positions in the target region with the same significance level. Based on this pipeline, we successfully identified four cases of somatic mosaicism among 10 apparently mutation-negative CINCA/NOMID patients. These results were subsequently confirmed by functional analysis and subcloning followed by capillary DNA sequencing method.

As described above, we revealed that a read depth of ~350 for each strand of each amplicon would be sufficient to detect somatic mosaicism as low as 1% with statistical confidence. This means that an analysis of somatic mosaicism (detection limit of 1% allele frequency) of the *NLRP3* gene for one sample requires $350 \times 2 \times 14 = 9800$ reads with the 454 GS-FLX sequencer, which has a capacity to obtain 1 000 000 reads per run. Thus, we could analyse ~100 patient samples with a single run (~10 h) using this MPS platform. For this purpose, a miniaturized 454 sequencer might be more convenient because it could analyse 10 patient samples at once with a reasonably reduced running cost.

The approach used to detect somatic mosaicism is very similar to that for low-frequency alleles in pooled DNA samples, for which MPS applications have been reported by many groups.[18,26,27] However, the main aim of these previous studies was to screen for a rare allele in a population. Thus, the discovery phase on the MPS platform must be followed by an evaluation phase using conventional methods. Therefore, when diagnosing somatic mosaicism of the *NLRP3* gene based solely on the MPS platform, we could not use the same approach to detect rare alleles in a population due to its low accuracy. The sequencing error rate on the Roche MPS platform was sufficiently stable and low enough as shown in this study. Using our pipeline, we were able to detect 1% somatic mosaicism in the *NLRP3* gene with 99.9% confidence. Although another research group recently used a similar approach with a short-read MPS,[28] the Roche long-read MPS is more suitable as a diagnostic tool mainly because of the short run time. If we could diagnose somatic mosaicism of the *NLRP3* gene within a reasonable time with low labour and costs as shown in this study, the success rate of CINCA/NOMID genetic diagnosis will increase from 60 to 80% or higher,[9] which will greatly advance the health and care of these patients and prevent irreversible bone and neurological complications of disease.

This pipeline would also be efficient to detect somatic mosaicism in mutation-negative patients with other diseases, including cancer. The error rate map for a given gene should be constructed from authentic plasmids, and used to detect somatic mosaicism of other genes as well as rare alleles in various populations.

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This study was supported by the Japanese Ministry of Education, Science, Sports, and Culture, and the Japanese Ministry of Health, Labor, and Welfare.

Acknowledgements: We thank all patients who participated in the study. We are grateful to Ms. Yuki Takaoka at the Department of Pediatrics, Kyoto University Graduate School of Medicine and Mr. Takashi Watanabe at the Department of Human Genome Research, Kazusa DNA Research Institute for their technical assistance.

References

1. Prieur, A.M. and Griscelli, C. 1981, Arthropathy with rash, chronic meningitis, eye lesions, and mental retardation, *J. Pediatr.*, **99**, 79–83.
2. Hassink, S.G. and Goldsmith, D.P. 1983, Neonatal onset multisystem inflammatory disease, *Arthritis Rheum.*, **26**, 668–73.
3. Torbiak, R.P., Dent, P.B. and Cockshott, W.P. 1989, NOMID—a neonatal syndrome of multisystem inflammation, *Skeletal Radiol.*, **18**, 359–64.
4. Feldmann, J., Prieur, A.M., Quartier, P., et al. 2002, Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in CIAS1, a gene highly expressed in polymorphonuclear cells and chondrocytes, *Am. J. Hum. Genet.* United States, 198–203.
5. Aksentijevich, I., Nowak, M., Mallah, M., et al. 2002, De novo CIAS1 mutations, cytokine activation, and evidence for genetic heterogeneity in patients with neonatal-onset multisystem inflammatory disease (NOMID)—a new member of the expanding family of pyrin-associated autoinflammatory diseases, *Arthritis Rheum.*, **46**, 3340–8.

6. Hoffman, H.M., Mueller, J.L., Broide, D.H., Wanderer, A.A. and Kolodner, R.D. 2001, Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome, *Nat. Genet.*, **29**, 301–5.
7. Goldbach-Mansky, R. 2011, Current status of understanding the pathogenesis and management of patients with NOMID/CINCA, *Curr. Rheumatol. Rep.*, **13**, 123–31.
8. Milhavet, F., Cuisset, L., Hoffman, H.M., et al. 2008, The infevers autoinflammatory mutation online registry: Update with new genes and functions, *Hum. Mutat.*, **29**, 803–8.
9. Tanaka, N., Izawa, K., Saito, M.K., et al. 2011, High incidence of NLRP3 somatic mosaicism in patients with chronic infantile neurologic, cutaneous, articular syndrome: results of an International Multicenter Collaborative Study, *Arthritis Rheum.*, **63**, 3625–32.
10. Qin, W., Kozlowski, P., Taillon, B.E., et al. 2010, Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex, *Hum. Genet.*, **127**, 573–82.
11. Campbell, P.J., Pleasance, E.D., Stephens, P.J., et al. 2008, Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing, *Proc. Natl. Acad. Sci. USA*, **105**, 13081–6.
12. Rohlin, A., Wernersson, J., Engwall, Y., Wiklund, L., Bjoerk, J. and Nordling, M. 2009, Parallel sequencing used in detection of mosaic mutations: Comparison with four diagnostic DNA screening techniques, *Hum. Mutat.*, **30**, 1012–20.
13. Saito, M., Nishikomori, R., Kambe, N., et al. 2008, Disease-associated CIAS1 mutations induce monocyte death, revealing low-level mosaicism in mutation-negative cryopyrin-associated periodic syndrome patients, *Blood*, **111**, 2132–41.
14. Kent, W.J. 2002, BLAT—the BLAST-like alignment tool, *Genome Res.*, **12**, 656–64.
15. Fujisawa, A., Kambe, N., Saito, M., et al. 2007, Disease-associated mutations in CIAS1 induce cathepsin B-dependent rapid cell death of human THP-1 monocytic cells, *Blood*, **109**, 2903–11.
16. Gilles, A., Meglecz, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. 2011, Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing, *BMC Genomics*, **12**, 245.
17. Margulies, M., Egholm, M., Altman, W.E., et al. 2005, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, **437**, 376–80.
18. Altmann, A., Weber, P., Quast, C., Rex-Haffner, M., Binder, E.B. and Muller-Myhsok, B. 2011, vipR: variant identification in pooled DNA using R, *Bioinformatics*, **27**, i77–84.
19. Dohm, J.C., Lottaz, C., Borodina, T. and Himmelbauer, H. 2008, Substantial biases in ultra-short read data sets from high-throughput DNA sequencing, *Nucleic Acids Res.*, **36**, e105.
20. Lee, W., Jiang, Z., Liu, J., et al. 2010, The mutation spectrum revealed by paired genome sequences from a lung cancer patient, *Nature*, **465**, 473–7.
21. Awadalla, P., Gauthier, J., Myers, R.A., et al. 2010, Direct Measure of the de novo mutation rate in Autism and Schizophrenia Cohorts, *Am. J. Hum. Genet.*, **87**, 316–24.
22. Conrad, D.F., Keebler, J.E.M., DePristo, M.A., et al. 2011, Variation in genome-wide mutation rates within and between human families, *Nat. Genet.*, **43**, 712–4.
23. Cha, R.S. and Thilly, W.G. 1993, Specificity, efficiency, and fidelity of PCR, *PCR Methods Appl.*, **3**, S18–29.
24. Vandenbroucke, I., Marck, H.V., Verhasselt, P., et al. 2011, Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications, *BioTechniques*, **51**, 167–77.
25. Pleasance, E.D., Cheetham, R.K., Stephens, P.J., et al. 2010, A comprehensive catalogue of somatic mutations from a human cancer genome, *Nature*, **463**, 191–6.
26. Fakhrai-Rad, H., Zheng, J.B., Willis, T.D., et al. 2004, SNP discovery in pooled samples with mismatch repair detection, *Genome Res.*, **14**, 1404–12.
27. Bansal, V. 2010, A statistical method for the detection of variants from next-generation resequencing of DNA pools, *Bioinformatics*, **26**, i318–24.
28. Flaherty, P., Natsoulis, G., Muralidharan, O., et al. 2012, Ultrasensitive detection of rare mutations using next-generation targeted resequencing, *Nucleic Acids Res.*, **40**, e2.

Rheumatol Int (2012) 32:3761–3764
DOI 10.1007/s00296-011-2225-z

ORIGINAL ARTICLE

Novel mutations of MVK gene in Japanese family members affected with hyperimmunoglobulinemia D and periodic fever syndrome

Takahisa Mizuno · Hidemasa Sakai · Ryuta Nishikomori · Koichi Oshima · Osamu Ohara · Ikue Hata · Yosuke Shigematsu · Takashi Ishige · Kazushi Tamura · Hirokazu Arakawa

Received: 12 July 2011 / Accepted: 22 October 2011 / Published online: 11 December 2011

Abstract Hyperimmunoglobulinemia D with periodic fever syndrome (HIDS) is a recessively inherited recurrent fever syndrome. We describe a family of eldest son and monozygotic twin younger sisters with characteristic syndrome of HIDS, but normal level of IgD. Mevalonate kinase (MK) activity was deficient in all of them, and analysis of the MVK gene revealed compound heterozygosity for 2 new mutations, one of which was the disease-causing splicing mutation and the other was a novel missense mutation. All the patients had the same compound heterozygous mutations c.227-1 G > A and c.833 T > C, which resulted in exon 4 skipping and p.Val278Ala. This is the first case in which exon skipping mutation of the MVK gene has been certainly identified at the genomic DNA level. In each case, in which HIDS is clinically suspected, despite normal IgD level, analysis of MK activity and the MVK gene should be performed.

Keywords HIDS · MVK gene · Novel mutation · Compound heterozygous mutation · Splicing mutation · Inherited recurrent fever syndrome

T. Mizuno · T. Ishige · K. Tamura · H. Arakawa (✉)
Department of Pediatrics and Developmental Medicine, Gunma University Graduate School of Medicine, 3-39-15 Showa-machi, Maebashi, Gunma 371-8511, Japan
e-mail: harakawa@gunma-u.ac.jp

H. Sakai · R. Nishikomori
Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan

K. Oshima · O. Ohara
Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, Kanagawa, Japan

I. Hata
Department of Pediatrics, Faculty of Medical Sciences, University of Fukui, Fukui, Japan

Y. Shigematsu
Department of Health Science, Faculty of Medical Sciences, University of Fukui, Fukui, Japan

Introduction

Hyperimmunoglobulinemia D and periodic fever syndrome (HIDS) is a rare autosomal recessive auto-inflammatory disorder characterized by recurrent febrile attacks with lymphadenopathy, abdominal distress, skin eruptions, and joint involvement [1–3]. Febrile attacks usually last for 3–7 days and are interrupted by asymptomatic intervals of several weeks' duration [4–6]. Symptoms appear in early infancy and may persist throughout life with gradual increases in serum IgD [7, 8]. The diagnostic hallmark of HIDS is a constitutively elevated level of serum IgD, although parts of the patients have been reported to have normal amount of serum IgD levels.

The HIDS is caused by mutations on mevalonate kinase gene (MVK), which encodes an enzyme involved in cholesterol and non-sterol isoprenoid biosynthesis. We present herein a Japanese family, eldest son and monozygotic twin younger sisters, with HIDS that had compound heterozygous mutations on MVK gene, one of which was the disease-causing splicing mutation and the other was a novel missense mutation. Serum concentrations of IgD were repeatedly within the normal range. These cases demonstrate that detail analysis with more specific diagnostic tests such as urinary excretion of mevalonic acid and MVK genetic analysis should be performed not to miss the correct diagnosis in patients, especially younger children with HIDS.

Case reports

Patients are the eldest son and monozygotic twin younger sisters of parents of Japanese origin. The eldest son (patient 1) had presented with recurrent fever from 5 months of age. The twin younger sisters (patient 2 and 3) presented with fever from 1 month of age. Vomiting and diarrhea were presented in the younger sister (patient 3). Febrile episodes appeared every 4–8 weeks and lasted for 3–5 days on all the three patients. During febrile episodes, peripheral blood leukocytosis and CRP elevations (more than 10 mg/dl) were observed. In intermittent period between fever episodes, serum CRP levels decreased, but did not always become negative. Their parents had no history of recurrent fever. Sepsis work-up did not show any foci and any pathogens causing the febrile episodes. The repeated bacterial cultures resulted in negative, and administration of the antimicrobial agents did not change the clinical courses of the febrile episodes, indicating that the fever was not induced by pathogen. In addition, immunological analysis such as serum IgA, IgM, IgG, and IgD, lymphocytes counts including T, B, NK cells, and mitogen proliferation assays of peripheral blood mononuclear cells (PBMCs) were normal.

Due to the recurrent high fevers caused most unlikely by pathogen and the heavy family history of the periodic fevers, we suspected hereditary periodic fever syndromes and performed genetic study. After written informed consents approved by institutional review board of the Kyoto University Hospital were obtained, peripheral blood samples were collected from the patients and their parents for isolating genomic DNA and total RNA.

First, we performed genomic DNA sequencing for MEFV gene for familial Mediterranean fever, MVK gene for HIDS, NLRP3 for cryopyrin-associated periodic syndrome, and TNFRSF1A for TNF receptor-associated periodic syndrome. Genomic DNA sequencing analysis of the MVK gene revealed the presence of heterozygous mutations of c.227-1 G > A at the exon/intron border of exon 4 and c.833T > C (p.Val278Ala). Subsequent amplification of the cDNA by RT-PCR showed that the former mutation caused deletion of exon 4 (Fig. 1a). Genomic DNA sequence analysis on their parents revealed that the parents inherited c.227-1 G > A from their father and c.833T > C from their mother, indicating that the three patients were compound heterozygous for MVK gene (Fig. 1b). The patients had markedly elevated excretion of mevalonic acid in urine, especially in febrile periods, and their mevalonate kinase enzyme activities were very low, which confirmed that all the three patients suffered from HIDS (Table 1).

While the patients did not have any mutations on TNFRSF1A and NLRP3, we identified MEFV non-synonymous nucleotide alterations on the elder brother, who was a heterozygote for L110P, E148Q, and R202Q, and the younger twin, who was a heterozygote for R202Q in addition to MVK gene mutations. These MEFV gene nucleotide alterations were regarded as SNPs, and the clinical diagnosis of FMF was not compatible with the patients, although the complex MEFV gene alterations of L110P/E148Q/R202Q have been reported on the clinically-diagnosed FMF patients.

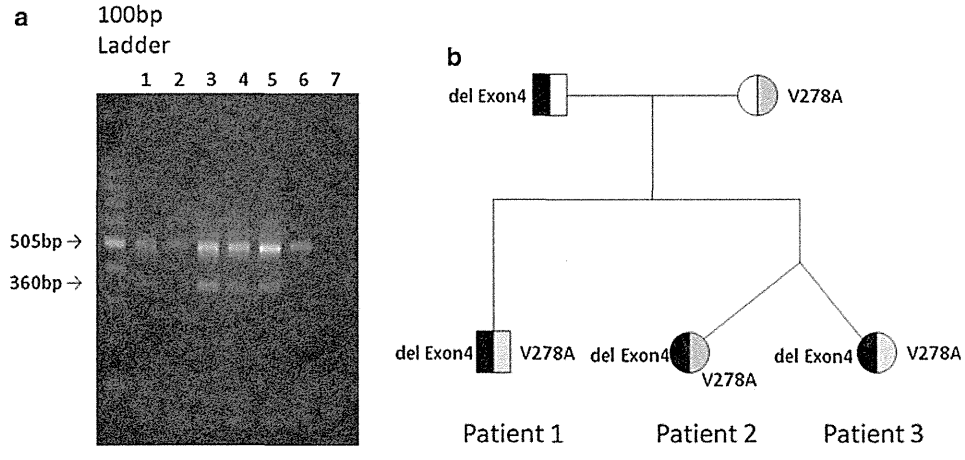

Fig. 1 Molecular genetic findings in the study patients. **a** Agarose gel electrophoresis of RT-PCR products for exon 2 to exon 5 of MVK shows the normal 505-bp alleles in samples from normal healthy control (*lane 6*) and mother (*lane 2*), as well as both the normal allele and the mutant 362-bp allele in the sample from father (*lane 1*), patient 1 (*lane 3*), patient 2 (*lane 4*), and patient 5 (*lane 5*). Subsequent cDNA sequencing confirmed that this 144-bp deletion in cDNA corresponds to codon 303–407 (exon 4). The molecular size marker was a 100-bp ladder. *Lane 7* represents PCR with distilled water added but not with DNA, indicating that there was no background amplification. **b** Pedigree of the affected family. The three patients are heterozygous for del exon 4 and V278A

Table 1 Urinary mevalonic acid and mevalonate kinase levels in the study patients

Patient no.	Mevalonic acid in urine (μg/mgCr)		Mevalonate kinase (pmol/minute/mg)
	Febrile period	Intermittent period	
1	67.9	11.3	3
2	55.6	17.7	2
3	58.8	18.5	2
Control		0.078 ± 0.012[a]	214 ± 62[a]

Control data are given as mean ± SD

[a] Values form healthy subjects were used to obtain a control range for urinary mevalonic acid levels (mean ± SD) and mevalonate kinase levels (mean ± SD)

Discussion

We present herein a sibling of HIDS that demonstrated compound heterozygous for two novel mutations of MVK gene. All the patients had the same compound heterozygous mutations c.227-1 G > A and c.833T > C, which resulted in exon 4 skipping and p.Val278Ala. The mutations are novel, especially the splicing mutation of MVK gene was identified at the genomic DNA level.

Cuisset et al. [9] reported that HIDS mutations were evenly distributed along the coding region of the MVK gene, in contrast to mutations causing MA, which clustered between 243 and 334. The sequence variations seen in MA are missense mutations that are in the same region as the variants described in HIDS. Further studies will be needed to clarify the association of phenotypical differences with MVK gene mutations. Over 80% of patients with HIDS were reported to have compound heterozygous mutation in the MVK gene. To our knowledge, both the skipping of exon 4 and V278A mutation have not been reported previously in HIDS. Moreover, this is the first case in which exon skipping mutation of the MVK gene has been certainly identified at the genomic DNA level. Only few groups reported HIDS patients with the skipping of exon in the cDNA of the MVK gene [10, 11]. They suggested that these exon skipping was probably due to the presence of a potential splice site mutation, but could not identify mutations responsible for these altered splicing through the sequence analysis at the genomic level. Most MVK mutations in patients with HIDS and MA have only been determined at the cDNA level; however, analysis of cDNA sometimes appeared troublesome, probably due to instability of the MVK mRNA. More detailed studies through the sequence analysis at the genomic level lead us to elucidate the role of MVK mutations in HIDS and MA, and expression studies in *E. coli* will be necessary to evaluate the effect of each mutation.

HIDS is classically defined as a high concentration of mevalonic acid in the urine and is characterized by a high serum IgD concentration during each febrile episode, but some reports from the Netherlands stated that high levels of serum IgD levels were not seen and affirmed that other diseases also showed high serum IgD levels [12]. In our cases, the analysis of enzymes and molecular genetics of MVK gene yielded the correct diagnosis, although serum concentrations of IgD were within the normal range. Thus, it should be now common practice to examine the MVK gene in order to diagnose this disease.

In conclusion, we present a Japanese family with HIDS that appeared to have novel mutations of MVK gene. Most of the HIDS cases were reported from European, especially Dutch, whereas only one HIDS case of Japanese patient was reported by Naruto et al. [13], which is only one report of Asian patient. Cases of HIDS may so far have been overlooked or misdiagnosed as infectious diseases or autoimmune disorders in Japan, besides there may be difference in race. It is necessary that accumulation of case in hereditary mutation and in other race leads to solve a detailed cause of HIDS.

Acknowledgments The authors thank Dr. Georg F. Hoffmann for measurement regarding the mevalonic kinase activity.

Conflict of interest There is no financial or other potential conflict of interest for each author.

References

1. van der Meer JW, Vossen JM, Radl J, Van Nieuwkoop JA, Meyer CJ, Lobatto S et al (1984) Hyperimmunoglobulinaemia D and periodic fever: a new syndrome. Lancet 1:1087–1090
2. Church LD, Churchman SM, Hawkins PN, McDermott MF (2006) Hereditary auto-inflammatory disorders and biologics. Semin Immunopathol 27:494–508
3. Simon A, van der Meer JW (2007) Pathogenesis of familial periodic fever syndromes or hereditary autoinflammatory syndromes. Am J Physiol Regul Integr Comp Physiol 292:R86–R98
4. Fenkel J, Houten SM, Waterham HR, Wanders RJ, Rijkers GT, Duran M et al (2001) Clinical and molecular variability in childhood periodic fever with hyperimmunoglobuninaemia D. Rheumatology (Oxford) 40:579–584
5. Stojanov S, Lohse P, Lohse P, Hoffmann F, Renner ED, Zellerer S et al. (2004) Molecular analysis of the MVK and TNFRSF1A genes in patients with a clinical presentation typical of the hyperimmunoglobulinemia D with periodic fever syndrome: a low-penetrance TNFRSF1A variant in a heterozygous MVK carrier possibly influences the phenotype of hyperimmunoglobulinemia D with periodic fever syndrome or vice versa. Arthritis Rheum 50:1951–1958
6. Touitous I, Lesage S, McDermott M, Cuisset L, Hoffman H, Dode C et al (2004) Infevers: an evolving mutation database for autoinflammatory syndromes. Hum Mutat 24:194–198
7. Pouchot J, Sampalis JS, Beaudet F, Carette S, Decary F, Salusinsky-Sternbach M et al (1991) Adult Still's disease: manifestations, disease course, and outcome in 62 patients. Medicine (Baltimore) 70:118–136

8. Haas D, Hoffmann GF (2006) Mevalonate kinase deficiencies: from mevalonic aciduria to hyperimmunoglobulinemia D syndrome. Orphanet J Rare Dis 1:13
9. Cuisset L, Drenth JP, Simon A, Vincent MF, Van der Velde Visser S, Van der Meer JW (2001) Molecular analysis of MVK mutations and enzymatic activity in hyper-IgD and periodic fever syndrome. Eur J Hum Genet 9(4):260–266
10. Drenth JP, Cuisset L, Grateau G, Vasseur C, van de Velde-Visser SD, de Jong JG et al (1999) Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. International Hyper-IgD Study Group. Nat Genet 22:178–181
11. Takada K, Aksentijevich I, Mahadevan V, Dean JA, Kelley RI, Kastner DL (2003) Favorable preliminary experience with etanercept in two patients with the hyperimmunoglobulinemia D and periodic fever syndrome. Arthritis Rheum 48:2645–2651
12. Drenth JP, Haagsma CJ, van der Meer JW (1994) Hyperimmunoglobulinemia D and periodic fever syndrome. The clinical spectrum in a series of 50 patients. International Hyper-IgD Study Group. Medicine (Bartimore) 73:133–144
13. Naruto T, Nakagishi Y, Mori M, Miyamae T, Imagawa T, Yokota S (2009) Hyper-IgD syndrome with novel mutation in Japanese girl. Mod Rheumatol 19(1):96–99

Human Molecular Genetics, 2012, Vol. 21, No. 7 **1496–1503**
doi:10.1093/hmg/ddr585
Advance Access published on December 13, 2011

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).

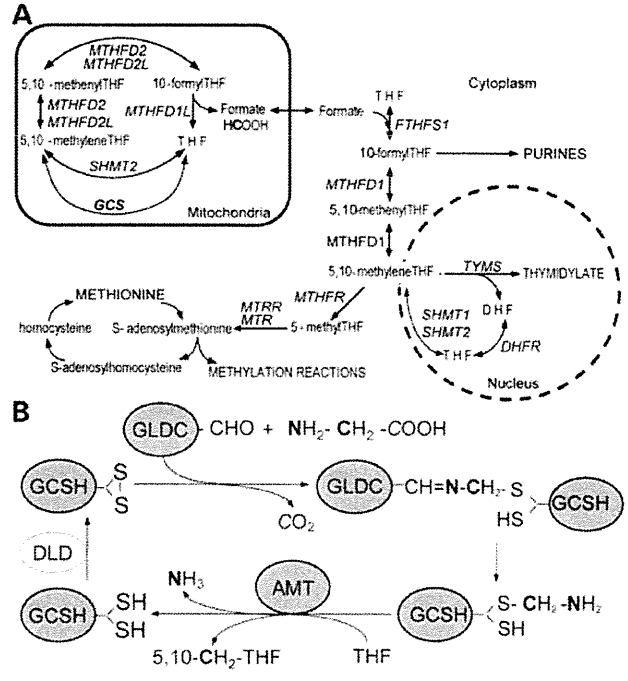

Figure 1. Schematic diagrams summarizing the key reactions of folate-mediated one-carbon metabolism and the GCS. (**A**) Folates donate and accept one-carbon units in the synthesis of purines, thymidylate and methionine. Mitochondrial FOCM supplies one-carbon units to the cytoplasm via formate. The GCS 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

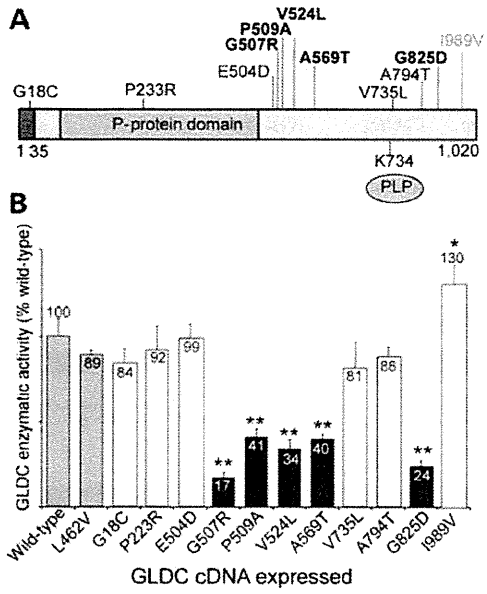

Figure 2. Characterization of *GLDC* missense mutations identified through DNA sequence analysis. (**A**) The schematic represents the 1020 amino acid residue GLDC polypeptide with the positions of the identified missense variants indicated. Mutations conferring significantly reduced activity (**B**) 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).

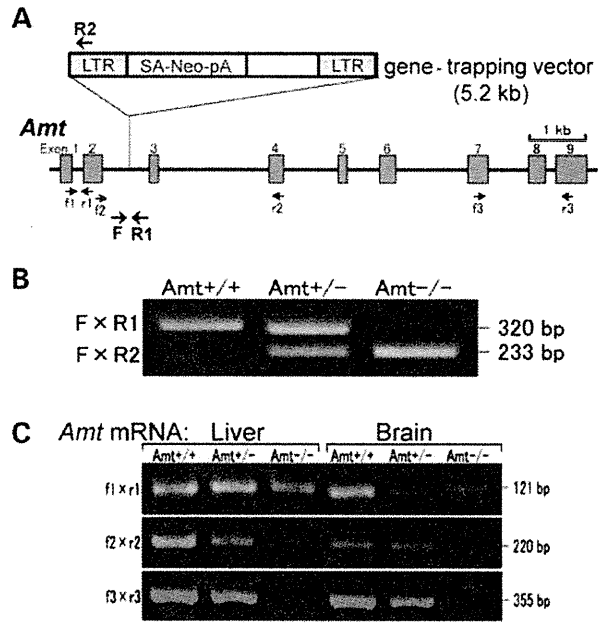

Figure 3. Generation of *Amt* knockout mouse by gene trapping. (**A**) The location of the gene-trap vector in *Amt* intron 2 in the ES cell line OST181110 was determined by inverse PCR. Mice carrying this mutation were generated using standard methods of blastocyst microinjection with OST181110 ES cells 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 $Amt^{+/-}$

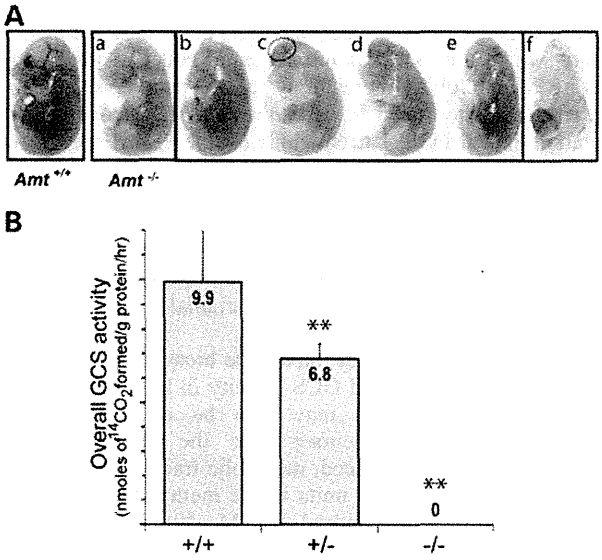

Figure 4. Mice lacking GCS activity exhibit NTDs. (**A**) Phenotypes of *Amt* mutant mice. NTDs were evident in the majority (88%) of $Amt^{-/-}$ fetuses (examples shown are at E17.5). Various types of NTDs were observed in $Amt^{-/-}$ fetuses, which principally affected the cranial region: a, no NTDs; b, small exencephaly (dotted circle); c–e, large exencephaly; f, craniorachischisis. (**B**) Enzymatic activity of the GCS in *Amt* knockout mice. $Amt^{+/-}$ and $Amt^{-/-}$ fetuses had significantly lower GCS activity in the liver than $Amt^{+/+}$ fetuses, with activity in $Amt^{-/-}$ samples below the level of detection ($^{**}P < 0.01$ compared with $Amt^{+/+}$).

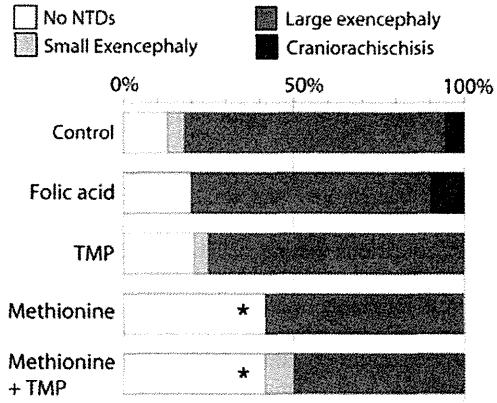

Figure 5. Maternal supplementation of *Amt* mutant embryos with folic acid, TMP or methionine. Maternal treatment with folic acid ($n = 10$ homozygous mutant fetuses) or TMP ($n = 12$) had no significant effect on NTD frequency, whereas the frequency of unaffected embryos was significantly increased following treatment with methionine ($n = 12$) or methionine plus TMP group ($n = 12$). The asterisk indicates significant difference compared with non-treated group ($P < 0.05$).

($n = 66$) fetuses. Defects mainly comprised exencephaly (82%), in which the cranial neural folds persistently failed to close (Fig. 4). There was also a low frequency of the more severe condition, craniorachischisis (5%), in which the neural tube remains open from the mid- and hindbrain, and throughout the spinal region (Fig. 4). Fetal liver samples were subjected to enzyme assay to determine overall activity of the GCS. In $Amt^{-/-}$ mice, overall GCS activity was effectively ablated being below the detection level of the assay (0.01 nmoles of $^{14}CO_2$ formed/gram protein/h), consistent with the Amt^{-} allele being a functional null (22) (Fig. 4). These findings confirm that AMT function is essential for GCS activity, and that the latter is necessary for successful neural tube closure.

Given that GCS is a component of FOCM (Fig. 1), we evaluated the possible prevention of NTDs by folate-related metabolites. Maternal supplementation was performed with folic acid, thymidine monophosphate (TMP), methionine or methionine plus TMP (23). Neither folic acid nor TMP significantly affected the frequency of NTDs among the homozygous $Amt^{-/-}$ offspring. However, we observed a significant protective effect of maternal supplementation with methionine or methionine plus TMP, compared with the non-treated group ($P < 0.05$; Fig. 5).

DISCUSSION

NTDs remain among the commonest human birth defects and understanding their genetic basis presents a considerable challenge owing to their multigenic inheritance and the potential influence of environmental factors, either predisposing or ameliorating. Several lines of evidence indicate a requirement for FOCM in neural tube closure and, therefore, GCS-encoding genes provide excellent candidates for possible involvement in NTD susceptibility. We identified putative mutations in *AMT* and *GLDC* which include a splice acceptor mutation and a number of non-synonymous variants that were absent from a large group of population-matched controls, as well as from public SNP databases. In the case of *GLDC*, enzymatic assay confirmed that several mutations resulted in significant loss of enzyme activity. Finally, *in vivo* functional evidence of a requirement for GCS function in neural tube closure was provided by the occurrence of NTDs in $Amt^{-/-}$ mice lacking GCS activity. Together these findings indicate that mutations in *GLDC* and *AMT* predispose to NTDs in both mice and humans.

Where parental samples were available (6 of the 11 NTD cases that involved putative mutations in *GLDC*), we demonstrated parent-to-child transmission (Supplementary Material, Table S2). Six were instances of maternal transmission and one involved paternal transmission. We hypothesize that absence of an overt NTD phenotype in parents who carry a deficient *GLDC* allele may result from incomplete penetrance, or lack of additional genetic or environmental factors which are predicted to be necessary for NTDs owing to their multifactorial aetiology. We also note that partial penetrance is a feature of numerous mouse models of NTDs (5,8).

Inherited GCS deficiency, owing to mutation of *AMT* and/or *GLDC*, has been shown to cause NKH in humans (17). NKH is a rare, autosomal recessive, inborn error of metabolism, characterized by accumulation of glycine and encephalopathy-like neurological signs, including coma and convulsive seizures in neonates. GCS activity is greatly diminished in NKH patients and they would, therefore, be predicted to be at increased risk of NTDs. It is possible that NTDs may occur in combination with NKH but as anencephaly is a lethal condition, co-existing

NKH would go undetected. Lack of NTDs in NKH patients may also reflect the multigenic nature of NTDs, which require the presence of additional risk alleles in non-GCS genes. NKH is a relatively rare condition, with a prevalence of 1/63 000 births in British Columbia (24) and 1/250 000 in the USA (25). It is therefore possible that an increased risk of NTDs among carriers of GCS mutations in NKH families may not have been noted and this possibility is worthy of investigation. Based on estimated carrier frequency and the incidence of mutations among NTD patients, we predict that NTDs might be expected among 1/150 of the siblings of NKH patients (see Supplementary Material, Table S3 for estimate calculation). One case report of an NKH patient with a *GLDC* mutation describes the additional presence of spinal cord hydromyelia (19). This condition is often associated with low spinal defects (involving secondary neurulation), but it is also possible that the expanded spinal canal was also present at a higher level and might indicate a limited defect in primary neurulation.

The mutations described in the current study were all present in heterozygous form and, therefore, are hypothesized to be insufficient to cause NKH while predisposing to NTDs. For example, in the current study we found four NTD patients and one control individual to be heterozygous for the A569T mutation, which is shown to result in reduced enzyme activity. This mutation was previously identified in a Caucasian patient with typical NKH, in combination with a second mutation, P765S (26), confirming that it is deleterious *in vivo*. Hence, we predict that, depending on the co-existing genetic milieu, the A569T variant may cause NKH, predispose to NTDs or be compatible with normal development.

The high incidence of NTDs in *AMT* mutant mice is particularly notable as NTDs have not previously been found to be a common feature of mouse models deficient for folate-metabolizing enzymes. This includes null mutants that have been reported for eight other genes that encode enzymes in FOCM (Fig. 1A) (27). Four have normal morphology at birth (*Cbs*, *Mthfd1*, *Mthfr* and *Shmt1*) (28–31). *Mthfd2* null embryos die by E15.5 but neural tube closure is complete (32) and null mutants for *Mtr*, *Mtrr* and *Mthfs* die before E9.5, prior to neural tube closure (33–35). Although analysis of mouse mutants has not supported a role for single-gene mutations in FOCM as major causes of NTDs, a requirement for cellular uptake of folate for neural tube closure has been demonstrated in *Folr1* null embryos, in which NTDs occur when rescued from early lethality by folic acid supplementation (36). There is also considerable evidence for possible involvement of gene–environment and/or gene–gene interactions in NTDs. For example, in *Pax3* mutant (*splotch*) embryos, which exhibit a defect of thymidylate biosynthesis, dietary folate-deficiency increases the frequency of cranial NTDs (23,37). Similarly, a diet deficient in folate and choline causes NTDs in *Shmt1* mutant embryos, whereas *Shmt1* and *Pax3* mutations exhibit genetic interaction (38).

Regarding the mechanisms by which GCS mutations affect neural tube closure, a key question is whether NTDs are caused by impairment of FOCM or by another cause such as glycine accumulation. Modelling of hepatic FOCM, based on biochemical properties of folate-metabolizing enzymes (39), predicts that loss of the mitochondrial GCS reaction would reduce the efflux rate of formate to the cytosol by ~50%. This results in reduced synthesis of purines and thymidylate, which are essential for the rapid cell division in the closing neural folds. Interestingly, a UK patient with anencephaly who was found to carry the *GLDC* loss-of-function mutation P509A in the current study (Table 1) was previously found to have impaired thymidylate biosynthesis, assayed in cultured fibroblasts (14). These findings support the hypothetical link between diminished GLDC function, reduced thymidylate biosynthesis and development of NTDs. Reduced thymidylate biosynthesis and diminished cellular proliferation are proposed to underlie folate-related cranial NTDs in *splotch* (*Pax3*) mouse mutants (37,38).

As well as impairment of nucleotide biosynthesis, the predicted effect of diminished GCS activity in reducing production of methionine (39) may also be of relevance as methionine is the precursor for the methyl donor *S*-adenosylmethionine. Indeed, metabolic tracing experiments suggest that ~80% of 1C units in the methylation cycle are generated within mitochondrial FOCM (40). Impairment of the methylation cycle and/or DNA methylation is known to cause NTDs in mice (41) and is proposed as a possible cause of human NTDs (7,42). It was therefore notable that we found a preventive effect of methionine supplementation in $Amt^{-/-}$ mice. Together, these findings suggest that FOCM, required for both thymidylate biosynthesis and methylation reactions that are essential for neural tube closure, may be functionally deficient in individuals who have mutations in *GLDC* or *AMT*.

MATERIALS AND METHODS

Patient cohorts and sequencing

Mutation analysis by DNA sequencing was performed on all exons of *AMT*, *GCSH* and *GLDC* as described (26). Cases comprised Japanese patients with anencephaly ($n = 14$) and two separate cohorts of UK patients with a diagnosis of anencephaly (combined $n = 24$), spina bifida ($n = 122$) or craniorachischisis ($n = 22$). In addition, the exons of *AMT*, *GCSH* and *GLDC* were sequenced in 76 Swedish patients with spina bifida. Unaffected controls, completely sequenced for these genes, comprised 36 Japanese and 189 unrelated UK subjects. Exons found to contain missense mutations were also sequenced in a further cohort of 192 well-characterized UK controls (43) and in 145 Swedish controls. This study was approved by the Ethical Committees of Tohoku University School of Medicine, UCL Institute of Child Health, Newcastle University and the Karolinska Institute.

Enzymatic assay of GCS activity and GLDC activity

GCS activity was measured in mouse liver samples by a decarboxylation reaction using [1-^{14}C]glycine as described (22). For analysis of GLDC activity, wild-type and mutant *GLDC* cDNAs were cloned into pCAG expression vector, kindly provided by Professor Jun-ichi Miyazaki (Osaka University, Japan) (44). Constructs were transfected into COS7 cells, which were harvested as described previously and cell pellets stored at −80°C prior to analysis (45). GLDC

enzymatic activity was determined, in triplicate, by exchange reaction between carbon dioxide and glycine using $NaH^{14}CO_3$ in the presence of excess recombinant bovine GCSH protein as described (22). An expression system of lipoylated bovine GCSH protein in *Escherichia coli* was kindly provided by Dr Kazuko Fujiwara (Tokushima University, Japan) (46). Statistical analysis was performed using SPSS software version 11.0 (SPSS, Inc., Chicago, IL, USA).

Knockout of Amt by insertion of a gene-trap vector

Mice carrying a gene-trap allele of *Amt* (here denoted Amt^{-}) were generated at Lexicon Genetics, Inc. (Houston, TX, USA) using the OST181110 ES cell line. The genomic insertion site of the gene-trap vector was determined by inverse PCR and localized to intron 2 (Supplementary Material, Fig. S1). Total RNA was prepared from the mouse liver and brain at E18 for RT-PCR analysis (Supplementary Material, Fig. S1 and Table S1). $Amt^{+/-}$ mice were backcrossed with wild-type C57BL/6 mice for nine generations to generate a congenic line of mice on the C57BL/6 background, for use in biochemical and histological analyses. This study was approved by the Animal Experiment Committee of Tohoku University.

Maternal supplementation with folic acid and related metabolites

Dams were treated with folic acid (25 mg/kg), thymidine-1-phosphate (TMP; 30 mg/kg) or L-methionine (70 mg/kg) by intra-peritoneal injection, 2 h prior to mating and daily from E7.5–10.5. Doses were based on previous studies (23,47,48).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by a research grant from the Ministry of Education, Culture, Sports, Science and Technology and a Research Grant from the Ministry of Health, Labour and Public Welfare in Japan. Research at Newcastle University was funded by the Newlife Foundation. Research at UCL Institute of Child Health was supported by SPARKS, the Wellcome Trust, Medical Research Council, UCL Biomedical Research Centre and by Great Ormond Street Hospital Children's Charity. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

REFERENCES

1. Greene, N.D. and Copp, A.J. (2009) Development of the vertebrate central nervous system: formation of the neural tube. *Prenatal Diag.*, **29**, 303–311.
2. Harris, M.J. and Juriloff, D.M. (2007) Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birth Defects Res. A Clin. Mol. Teratol.*, **79**, 187–210.
3. Greene, N.D.E., Stanier, P. and Copp, A.J. (2009) Genetics of human neural tube defects. *Hum. Mol. Genet.*, **18**, R113–R129.
4. Au, K.S., Ashley-Koch, A. and Northrup, H. (2010) Epidemiologic and genetic aspects of spina bifida and other neural tube defects. *Dev. Disabil. Res. Rev.*, **16**, 6–15.
5. Harris, M.J. and Juriloff, D.M. (2010) An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birth Defects Res. A Clin. Mol. Teratol.*, **88**, 653–669.
6. Copp, A.J. and Greene, N.D.E. (2010) Genetics and development of neural tube defects. *J. Pathol.*, **220**, 217–230.
7. Blom, H.J., Shaw, G.M., Den Heijer, M. and Finnell, R.H. (2006) Neural tube defects and folate: case far from closed. *Nat. Rev. Neurosci.*, **7**, 724–731.
8. Molloy, A.M., Brody, L.C., Mills, J.L., Scott, J.M. and Kirke, P.N. (2009) The search for genetic polymorphisms in the homocysteine/folate pathway that contribute to the etiology of human neural tube defects. *Birth Defects Res. A Clin. Mol. Teratol.*, **85**, 285–294.
9. Beaudin, A.E. and Stover, P.J. (2009) Insights into metabolic mechanisms underlying folate-responsive neural tube defects: a mini review. *Birth Defects Res. A Clin. Mol. Teratol.*, **85**, 274–284.
10. Tibbetts, A.S. and Appling, D.R. (2010) Compartmentalization of mammalian folate-mediated one-carbon metabolism. *Annu. Rev. Nutr.*, **30**, 57–81.
11. Boyles, A.L., Hammock, P. and Speer, M.C. (2005) Candidate gene analysis in human neural tube defects. *Am. J. Med. Genet. C Semin. Med. Genet.*, **135**, 9–23.
12. Shaw, G.M., Lu, W., Zhu, H., Yang, W., Briggs, F.B., Carmichael, S.L., Barcellos, L.F., Lammer, E.J. and Finnell, R.H. (2009) 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. *BMC Med. Genet.*, **10**, 49.
13. Martinez, C.A., Northrup, H., Lin, J.I., Morrison, A.C., Fletcher, J.M., Tyerman, G.H. and Au, K.S. (2009) Genetic association study of putative functional single nucleotide polymorphisms of genes in folate metabolism and spina bifida. *Am. J. Obstet. Gynecol.*, **201**, 394–411.
14. Dunlevy, L.P.E., Chitty, L.S., Doudney, K., Burren, K.A., Stojilkovic-Mikic, T., Stanier, P., Scott, R., Copp, A.J. and Greene, N.D.E. (2007) Abnormal folate metabolism in foetuses affected by neural tube defects. *Brain*, **130**, 1043–1049.
15. Parle-McDermott, A., Pangilinan, F., O'Brien, K.K., Mills, J.L., Magee, A.M., Troendle, J., Sutton, M., Scott, J.M., Kirke, P.N., Molloy, A.M. and Brody, L.C. (2009) A common variant in MTHFD1L is associated with neural tube defects and mRNA splicing efficiency. *Hum. Mutat.*, **30**, 1650–1656.
16. Kikuchi, G. (1973) The glycine cleavage system: composition, reaction mechanism, and physiological significance. *Mol. Cell. Biochem.*, **1**, 169–187.
17. Kure, S., Tada, K. and Narisawa, K. (1997) Nonketotic hyperglycinemia: biochemical, molecular, and neurological aspects. *Jpn J. Hum. Genet*, **42**, 13–22.
18. Kure, S., Narisawa, K. and Tada, K. (1992) Enzymatic diagnosis of nonketotic hyperglycinemia with lymphoblasts. *J. Pediatr.*, **120**, 95–98.
19. Hayasaka, K., Tada, K., Kikuchi, G., Winter, S. and Nyhan, W.L. (1983) Nonketotic hyperglycinemia: two patients with primary defects of P-protein and T-protein, respectively, in the glycine cleavage system. *Pediatr. Res.*, **17**, 967–970.
20. Ichinohe, A., Kure, S., Mikawa, S., Ueki, T., Kojima, K., Fujiwara, K., Iinuma, K., Matsubara, Y. and Sato, K. (2004) Glycine cleavage system in neurogenic regions. *Eur. J. Neurosci.*, **19**, 2365–2370.
21. Toone, J.R., Applegarth, D.A., Kure, S., Coulter-Mackie, M.B., Sazegar, P., Kojima, K. and Ichinohe, A. (2002) Novel mutations in the P-protein (glycine decarboxylase) gene in patients with glycine encephalopathy (non-ketotic hyperglycinemia). *Mol. Genet. Metab.*, **76**, 243–249.
22. Sakata, Y., Owada, Y., Sato, K., Kojima, K., Hisanaga, K., Shinka, T., Suzuki, Y., Aoki, Y., Satoh, J., Kondo, H. *et al.* (2001) Structure and expression of the glycine cleavage system in rat central nervous system. *Brain Res. Mol. Brain Res.*, **94**, 119–130.
23. Fleming, A. and Copp, A.J. (1998) Embryonic folate metabolism and mouse neural tube defects. *Science*, **280**, 2107–2109.