Supplementary Information

Human nonsense mutations in primary hyperammonemia – Analysis of publicized patient mutations and variations in general populations in eight disease-causing genes

Nobuhiko Kamoshita[1,2*], Takafumi Hiramoto[1], Yuji Kashiwakura[1], Morisada Hayakawa[1,2], and Tsukasa Ohmori[1,2]

[1]Department of Biochemistry, Faculty of Medicine, and [2]Center for Gene Therapy Research, Jichi Medical University, 3311-1 Yakushiji, Shimotsuke-shi, Tochigi-ken 329-0498, Japan

*Corresponding author

Nobuhiko Kamoshita, nkamo@jichi.ac.jp

Running title: Nonsense variations in primary hyperammonemia

Contents

Supplementary Tables 2-1, 5 and 9 are provided in separate files.

Supplementary Information

Supplementary Materials and Methods

Data acquisition

Nucleotide and protein sequences were downloaded from NCBI GenBank.

Data exclusion

(1) Mutation sites: Four mutations in *SLC25A13* identified in cancerous tissues (Fig. S2 white lines) were listed in Dataset 1, but were excluded from the downstream analysis (Fig. 2 and 3). These were p.Gln169*_ochre from glioblastoma [1], p.Glu307*_ochre from esophageal squamous cell carcinoma [2], p.Gln311*_amber from breast adenocarcinoma [1], and p.Ser619*_opal from prostate adenocarcinoma [3].

During the literature and database search, we encountered three nucleotide changes causing a termination codon in a general population or the process of identification was not described (Fig. S2 orange lines). p.Glu373*_amber in *ASS1* and p.Ser599*_ochre in *SLC25A13* [4] were mutations solely reported from a general population. One mutation in OTC, p.Gly212*_opal was present in dbSNP but the process of identification was not clear. While the definition of nonsense mutation was strongly suggestive of disease-causing, since evidence to really evoke human disease was lacking for these sites, they were included in Dataset 1 but were excluded from the downstream analysis (Fig. 2 and 3).

Two nonsense nucleotide changes in the X chromosome, p.Glu271*_ochre and p.Glu273*_amber in *OTC*, which have recently been disclosed in jMorp 38KJPN, did not pass filters and were excluded.

(2) Evaluation of patient data

(i) inheritance: One case of *ASL* gene deficiency which is described as “normal parents” [5] was classified into “likely de novo.” One case of *CPS1* gene deficiency, parents of which were described as “unrelated and healthy,” but their first child died [6] were classified into “likely inherited.”

(3) Ethnicity match

Details are described in Table S7.

(i) Patient incidents from UK and US: In some cases, ethnicity data published from affiliations of these countries were anonymized. Unless specified in literature, subjects were regarded as an European (non-Finnish) origin.

(ii) Incidents from Brazil: Description of ethnicity was lacking and was excluded from the analysis in Figure 4.

Supplementary Results

Evaluation of clinical information

Owing to the sporadic nature of mutations (Fig. 2A), *s* was determined by gene. In SLC25A15, *s* was not determined because, in addition to incomplete penetrance (Table 1), most cases of NICCD (neonatal intrahepatic cholestasis caused by citrin deficiency) can survive to reproductive age. The rest of the seven genes were divided into three groups according to the location of the proteins they encode as below.

(1) Proximal/ mitochondrial enzymes (*NAGS*, *CPS1*, and *OTC*)

The common feature observed in the proximal three enzyme deficiencies was the high rate of neonatal-onset cases (Fig. 1B). Outcome was severe unless intervention started from the neonatal

period. In addition to five cases that received liver transplantation, pharmacological management using N-carbamyl-L-glutamate (NCG), which mimics NAG (N-acetyl glutamate, Fig. S1), an allosteric activator of *CPS1*, was effective for NAGS deficiency (Fig. 1C). In contrast, two and one post-neonatal-onset cases of *CPS1* and *OTC* have been reported to be associated with symptoms or death, respectively (Table S5). Calculated *s*, which is a minimal value based on an assumption that alive cases were reproductive, was 0.571, 0.825, and 0.833 in *NAGS*, *CPS1*, and male *OTC* deficiency, respectively.

(2) First two distal or cytoplasmic enzymes (*ASS1* and *ASL*)

While the incidence rate of *ASS1* deficiency in Japan and Germany was higher than that of *ASL* (Table 1), AF of nonsense alleles in a population was lower than that of *ASL*. The rate of cases assigned to the severe symptom was somehow low in this gene. Two cases of p.Tyr163*_ochre [7] and p.Arg279*_opal [8] diagnosed by newborn screening survived infancy. One late-onset case with p.Arg279*_opal mutation [9] was described as mild /asymptomatic at the age of 3 (Table S6). With an assumption that these cases were reproductive, *s* was calculated to be 0.500, although this value is influenced by the low denominator value (n=6) employed in the calculation (Fig. 1C).

In ASL, although not reflected in the population data values (Fig. 4A), 36 occurrences (50%), all homozygotic, were reported from Arabic populations, at four mutation sites: p.Gln116*_amber, p.Gln127*_amber, p.Gln354*_amber, and p.Gly351*_opal. According to the published description [10,11], the natural history of these cases was expected to be severe. Two neonatal cases survived infancy [10]. With an assumption that these cases are reproductive, *s* was calculated to be 0.714 (Fig. 1C).

(3) *ARG1* and ornithine transporter (*SLC25A15*)

Overall, 30 and 13 independent occurrences have been diagnosed in *ARG1* and *ORNT1* deficiency, with the fourth and second lowest occurrences among the eight genes. The peak period of diagnosis was in infancy or later (Fig. 1B), but without early diagnosis and intervention, severe neurological symptoms developed (Fig. 1C). Liver transplantation performed at the age of 1 year and 5 months cured one male patient of *ARG1* deficiency [12]. Calculated *s* was 0.840 and 0.885 in *ARG1* and *ORNT1* deficiencies, respectively.

(4) Aspartate transporter (*SLC25A13*)

Among 94 occurrences in 26 mutation sites (Dataset 2), 88 have been reported from East Asia, in which Japanese and Chinese occurrences accounted for 57 (61%) and 29 (31%), respectively, in Dataset 2.

One of the most notable variations, p.Ser225*_amber, did not form homozygotes in all 30 occurrences in Dataset 2. The phenotype of compound heterozygotes between p.Ser225* and c.1177+1G>A, a splice site mutation prevalent in 38KJPN, was compared with those of eight other genotypes [13]. The rates of neonatal symptoms of low birth weight and height [14] or abnormal laboratory findings in terms of elevated ammonium or transaminases were one of the worst three among the nine genotypes compared. The pathogenicity of p.Ser225*_amber in contributing to a severe phenotype in NICCD would be equivalent to that of c.852_855delTATG frameshift mutation or c.1311+1G>A, another splice site mutation prevalent in Japan.

Evaluation of patient frequency (Dataset 2)

Locations of mutations occurring more than once

Some mutations reported more than once are as follows. In three proximal or mitochondrial enzyme deficiencies, these were p.Trp324*_amber and p.Gln331*_amber in *NAGS*; p.Gln44*_amber, p.Ser430*_opal, p.Arg721*_opal, p.Arg787*_opal (underline indicates mutations with more than five independent events), p.Tyr1031*_ochre, p.Arg1174*_opal, p.Arg1262*_opal, and p.Leu1318*_amber in *CPS1*; and p.Arg23*_opal, p.Trp58*_opal, p.Arg92*_opal, p.Arg141*_opal, p.Ser146*_opal, p.Gln279*_amber, p.Glu310*_ochre, and p.Arg320*_opal in *OTC*.

In deficiencies of distal enzymes in cytoplasmic reactions, p.Gln27*_ochre and p.Arg279*_opal in *ASS1*; p.Gln116*_amber, p.Arg182*_opal, p.Arg213*_opal, p.Arg217*_opal, p.Gln354*_amber (double underline indicates mutations with more than 20 events), and p.Tyr430*_ochre in *ASL*; and

p.Gly12*_opal, p.Arg21*_opal, p.Lys75*_ochre, p.Trp122*_amber, and p.Arg291*_opal in *ARG1* were reported.
In transporter deficiencies, p.Arg179*_opal was the sole mutation in *SLC25A15*, while as many as 13 mutations of p.Glu16*_amber, p.Arg43*_opal, p.Gln159*_ochre, p.Arg184*_opal, p.Ser225*_amber, p.Gln259*_amber, p.Gly283*_opal, p.Arg319*_opal, p.Arg355*_opal, p.Arg360*_opal, p.Arg467*_opal, p.Glu601*_ochre, and p.Arg605*_opal were reported in *SLC25A13*.

Evaluation of allele frequency (AF) (Dataset 3)

Overall, 30 of the 60 variants were located at the reported patient mutation sites, which included 19 CGA>TGA sites (Fig. 2B). In contrast, 30 variants did not overlap with reported patient mutation sites. All of these non-overlapping variants were reported from a single population. Except for three ASJ (Ashkenazi Jewish) alleles of p.Gln247*_amber in the *NAGS* gene, the number of alleles at non-overlapping sites was at most two. These alleles would represent rare variations.
84 alleles (39%) were reported in the *SLC25A13* gene (Table S7), in which 68 were derived from the 38KJPN population (Dataset 3). This would be because over one-third of participants were Japanese (Table S1) and the penetrance of disease-causing variants in this gene is incomplete (Table 1).

Locations of 11 nonsense variations that overlap with known patient mutations and have been reported from more than one population

These were *CPS1*-27 (JPN+NFE), *ASS1*-08 (AFR+NFE), *ASL*-07, 09 (JPN+NFE, AMR+JPN+NFE, respectively), *ARG1*-02, 05, 06 (JPN+NFE, AFR+JPN, AFR+NFE), *SLC25A15*-03, 06 (AFR+EAS+JPN+NFE+OTH, AMR+JPN), and *SLC25A13*-03, 16 (JPN+NFE, JPN+NFE).

Supplementary Discussion

Examples of factor (iii)

Prevalent pathogenic variants and their AF in 38KJPN are as follows.

CPS1	c.1529delG	p.Gly510Alafs*5	0.000194
ASS1	c.421-2A>G		0.000671
	c.910C>T	p.Arg304Trp	0.000542
	c.1003C>T	p.Arg335Cys	0.000646
SLC25A13	c.852_855delTATG	p.Met285fs	0.002983
	c.1177+1G>A		0.004817
	c.1311+1G>A		0.001304

In Japan and China, the most prevalent disease-causing mutation in the *SLC25A13* gene is a splice site mutation c.1177+1G>A (p.A340_R392del) or a frameshift mutation c.852_855delTATG, p.Met285Profs*2, respectively. AF of the former in 38KJPN and the latter in EAS in gnomAD v.3.1.2 was described as 0.004817 and 0.005389, respectively. Overall, 20 and 30 incidences listed in Dataset 2 are compound heterozygotes with the former and the latter, respectively.

References for Supplementary Information

1 Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MDM, *et al*. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* 2015;**6**:10086.
2 Deng J, Weng X, Ye J, Zhou D, Liu Y, Zhao K. Identification of the Germline Mutation Profile in Esophageal Squamous Cell Carcinoma by Whole Exome Sequencing. *Front Genet* 2019;**10**:47.
3 Huang K-L, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, *et al*. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* 2018;**173**:355-370.e14.

4 Hou Y-CC, Yu H-C, Martin R, Cirulli ET, Schenker-Ahmed NM, Hicks M, *et al*. Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. *Proc Natl Acad Sci USA* 2020;**117**:3053–62.

5 Zhao M, Hou L, Teng H, Li J, Wang J, Zhang K, *et al*. Whole-Exome Sequencing Identified a Novel Compound Heterozygous Genotype in ASL in a Chinese Han Patient with Argininosuccinate Lyase Deficiency. *Biomed Res Int* 2019;**2019**:3530198.

6 Yan B, Wang C, Zhang K, Zhang H, Gao M, Lv Y, *et al*. Novel Neonatal Variants of the Carbamoyl Phosphate Synthetase 1 Deficiency: Two Case Reports and Review of Literature. *Front Genet* 2019;**10**:718.

7 Kido J, Matsumoto S, Sugawara K, Sawada T, Nakamura K. Variants associated with urea cycle disorders in Japanese patients: Nationwide study and literature review. *Am J Med Genet A* 2021;**185**:2026–2036.

8 Dimmock DP, Trapane P, Feigenbaum A, Keegan CE, Cederbaum S, Gibson J, *et al*. The role of molecular testing and enzyme analysis in the management of hypomorphic citrullinemia. *Am J Med Genet A* 2008;**146A**:2885–90.

9 Gao H-Z, Kobayashi K, Tabata A, Tsuge H, Iijima M, Yasuda T, *et al*. Identification of 16 novel mutations in the argininosuccinate synthetase gene and genotype-phenotype correlation in 38 classical citrullinemia patients. *Hum Mutat* 2003;**22**:24–34.

10 Al-Sayed M, Alahmed S, Alsmadi O, Khalil H, Rashed MS, Imtiaz F, *et al*. Identification of a common novel mutation in Saudi patients with argininosuccinic aciduria. *J Inherit Metab Dis* 2005;**28**:877–883.

11 AlTassan R, Bubshait D, Imtiaz F, Rahbeeni Z. A retrospective biochemical, molecular, and neurocognitive review of Saudi patients with argininosuccinic aciduria. *Eur J Med Genet* 2018;**61**:307–11.

12 Kido J, Matsumoto S, Takeshita E, Hayasaka C, Yamada K, Kagawa J, *et al*. Current status of surviving patients with arginase 1 deficiency in Japan. *Mol Genet Metab Rep* 2021**;29**: 100805.

13 Kido J, Häberle J, Sugawara K, Tanaka T, Nagao M, Sawada T, *et al*. Clinical manifestation and long-term outcome of citrin deficiency: Report from a nationwide study in Japan. *J Inherit Metab Dis* 2022;**45**:431–444.

14 Arai-Ichinoi N, Kikuchi A, Wada Y, Sakamoto O, Kure S. Hypoglycemic attacks and growth failure are the most common manifestations of citrin deficiency after 1 year of age. *J Inherit Metab Dis* 2021;**44**: 838-846.

15 Pronicka E, Piekutowska-Abramczuk D, Ciara E, Trubicka J, Rokicki D, Karkucińska-Więckowska A, *et al*. New perspective in diagnostics of mitochondrial disorders: two years' experience with whole-exome sequencing at a national paediatric centre. *J Transl Med* 2016;**14**:174.

16 Gobin-Limballe S, Ottolenghi C, Reyal F, Arnoux J-B, Magen M, Simon M, *et al*. OTC deficiency in females: Phenotype-genotype correlation based on a 130-family cohort. *J Inherit Metab Dis* 2021;**44**:1235–1247.

Figure S1

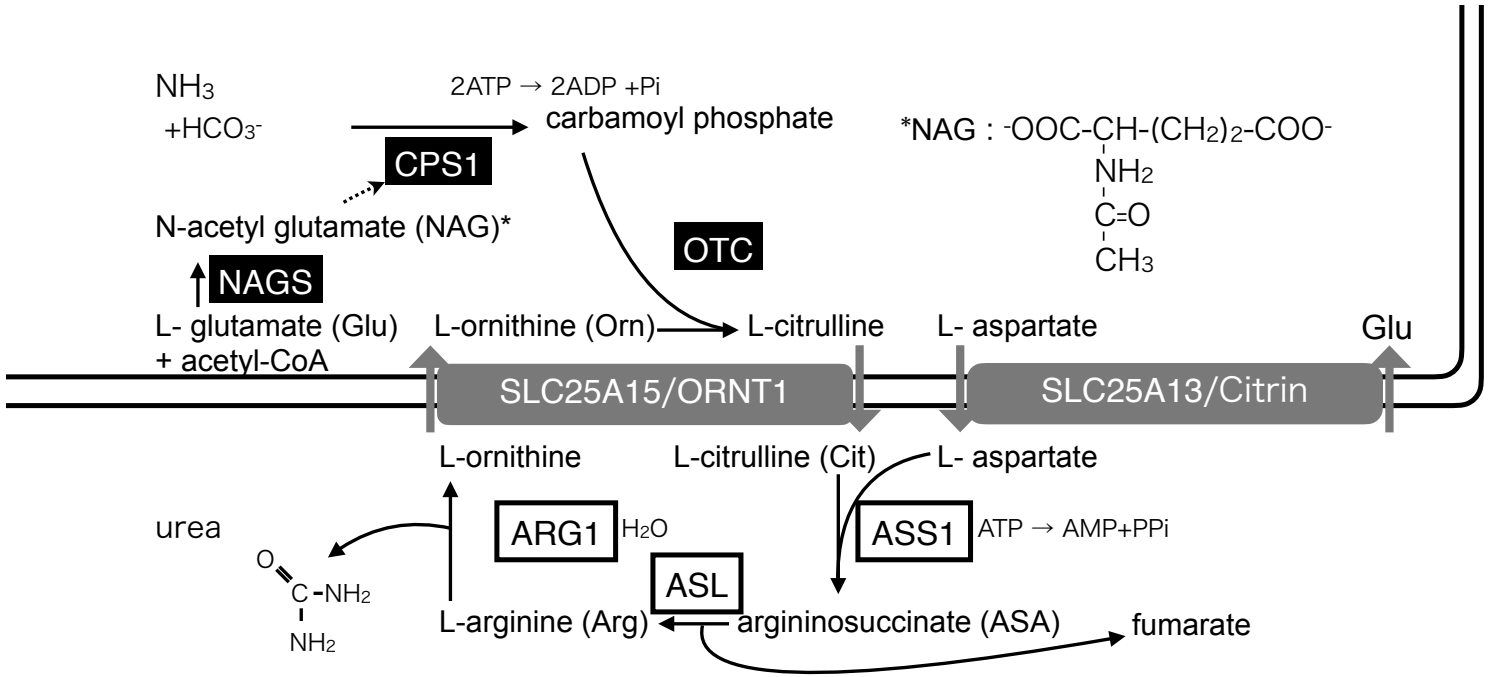

Figure S1. Enzymes and transporters involved in the urea cycle.
Initial substrates of ammonium and bicarbonate (top left) are shown with the final product of urea (bottom left) and major intermediates. Mitochondrial matrix is encircled by double lines. An asterisk indicates N-acetyl glutamate (NAG), a product of the first enzyme N-acetyl glutamate synthase (NAGS). Structure and allosteric activation to carbamoyl-phosphate synthase 1 (CPS1) are shown at the top right and by a dotted arrow at the top left, respectively. Mitochondrial and cytoplasmic enzymes are shown with closed and open rectangles, respectively. Mitochondrial transporters are shown in shaded orbitals. Abbreviations for gene products are shown in Roman font in this figure. OTC, ornithine transcarbamylase; ASS1, argininosuccinate synthase 1; ASL, argininosuccinate lyase; ARG1, arginase 1; SLC25A15, solute carrier family 25 member 15; ORNT1, ornithine transporter 1; SLC25A13, solute carrier family 25 member 13. Abbreviations of intermediates are shown within parentheses and also used in the laboratory findings in Dataset 2. Argininosuccinate, which is elevated in ASL deficiency or in argininosuccinuria, is measured via argininosuccinic acid (ASA).

Figure S2

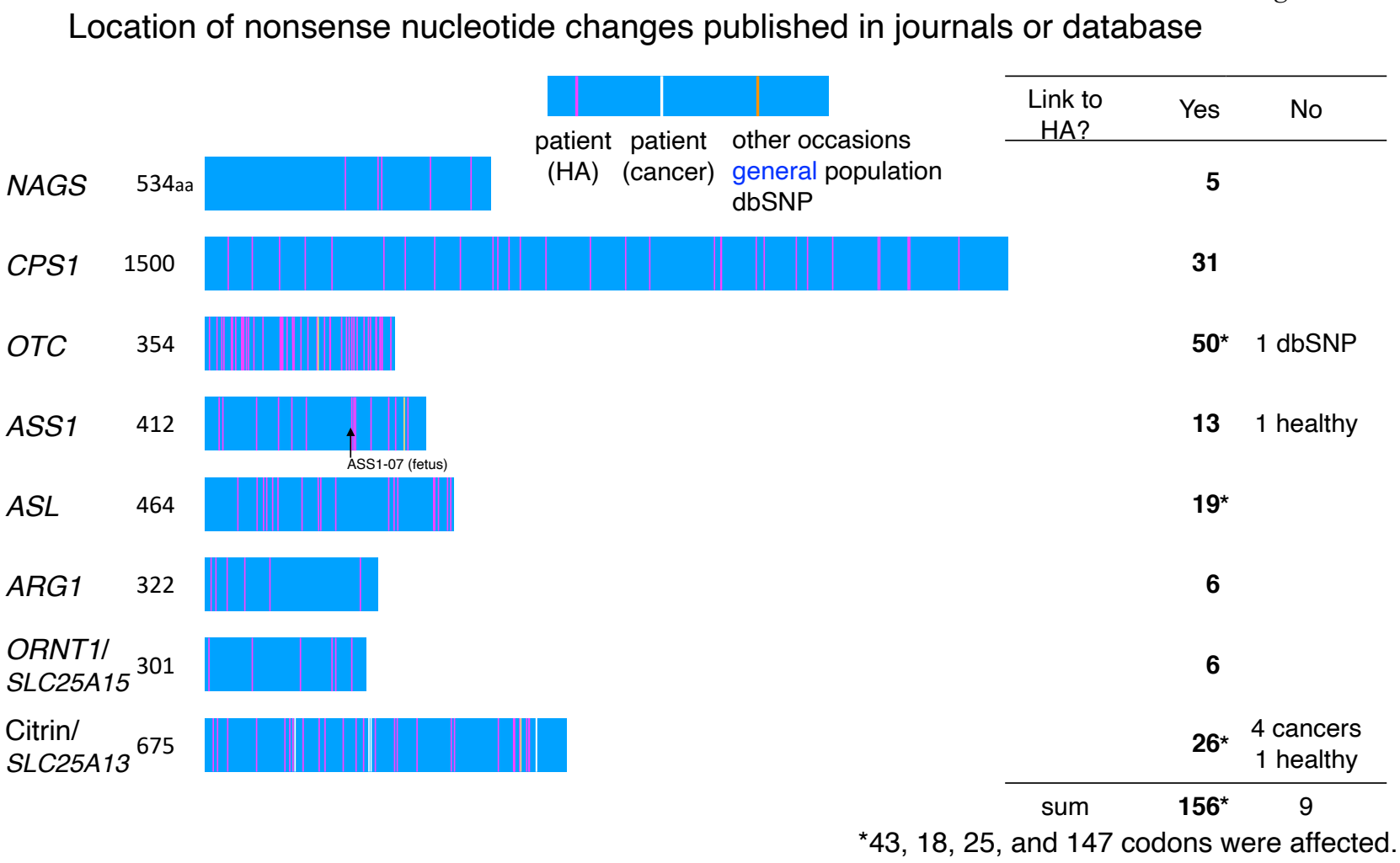

Figure S2. Nonsense nucleotide changes within the CDS of MANE Select transcripts (MANE Select CDS) of the eight genes causing primary hyperammonemia.
Locations of nucleotide changes identified in hyperammonemia patients are shown with magenta. Mutation in ASS1 identified in a fetus is indicated by an arrow. Locations of nucleotide changes in cancer or other conditions are shown with white or orange, respectively.
See Dataset 1 for details. A total of 155 locations linked to hyperammonemia (HA) are further analyzed.

Figure S3

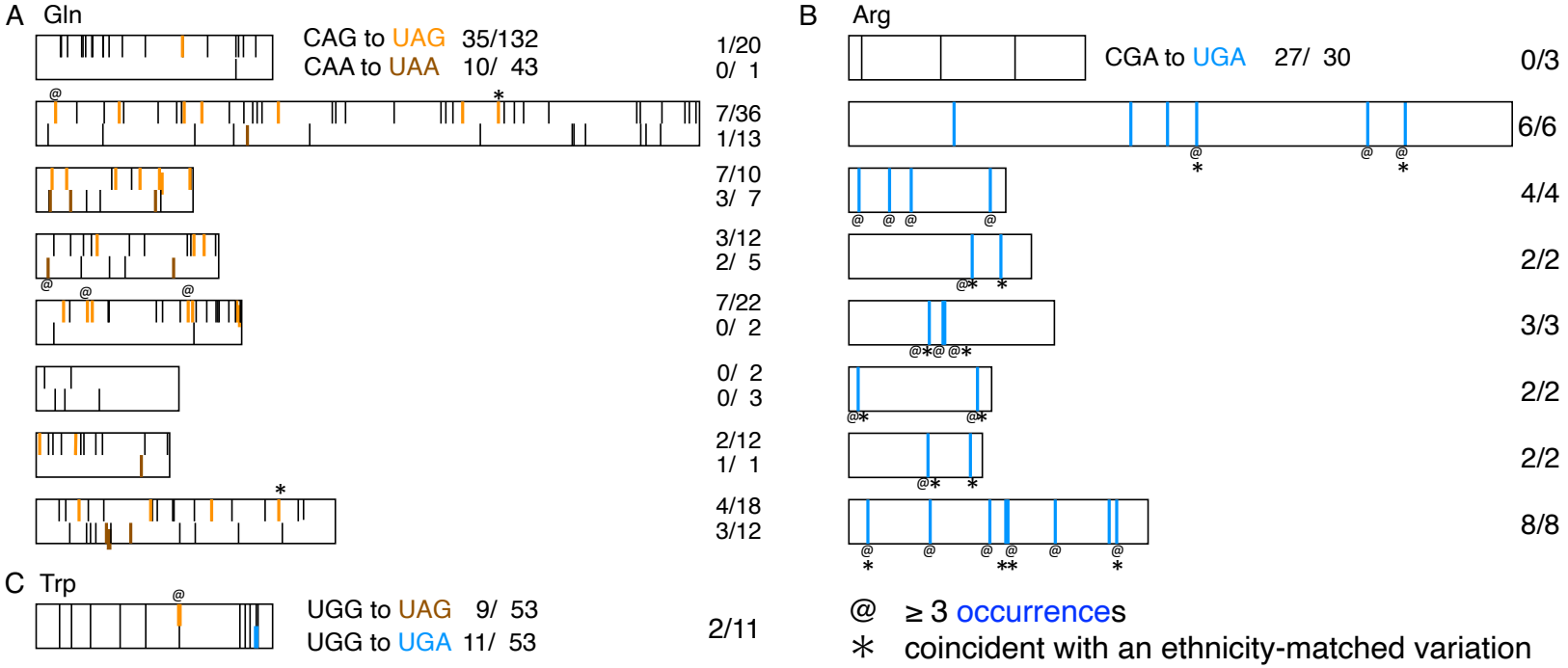

Figure S3. Distribution of nonsense mutations caused by base transitions. Codon positions of eight genes encoding glutamine (A), CGA arginine (B), and tryptophan (C) are shown by black bars. Patient mutations to TAG, TAA, or TGA are shown with an orange (amber), brown (ochre), or cyan (opal) bar, respectively. In A, mutations to TAG_amber and TAA_ochre are presented in the upper and lower sides of the rectangle, respectively. In C, mutations to TAG_amber and TGA_opal are shown in the upper and lower sides, respectively. Atsign indicates recurrent (≥3 occurrences) mutation sites. Asterisks indicate positions coincident with an ethnicity-matched nonsense variant in a general population. + (in C): One codon mutates into two different nonsense codons.

Figure S4

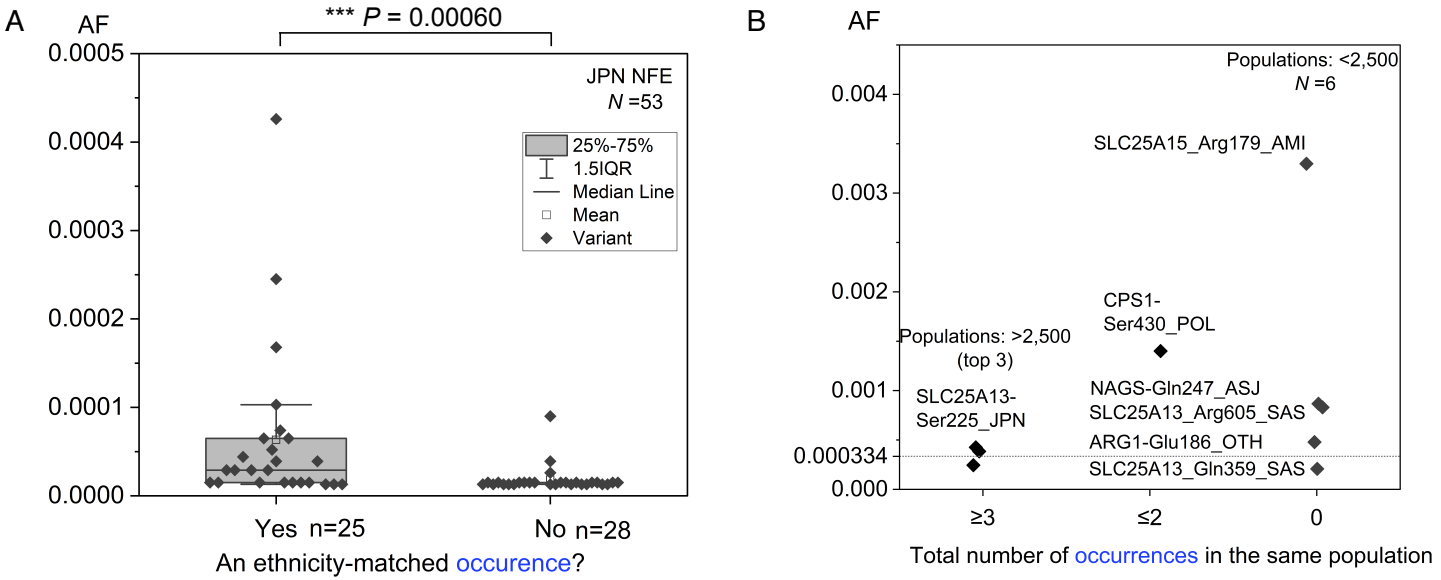

Figure S4. Box and scatter plots of nonsense variants.

(A) Variants in JPN or NFE populations were grouped according to the coincidence with an ethnicity-matched occurrences. Asterisks denote for a statistically significant difference assessed by Kruskal-Wallis test by ranks: ***, P <0.001.

(B) Nonsense variants reported from populations with <2,500 participants. Three variants with high AF values reported from >2,500 participants are shown for comparison in the column of the group "≥3." Variants in the groups "≤2" and "0" were reported in POL400 exomes [15] and gnomAD v3.1.2 databases, respectively. The level of an AF value expected from the equilibrium in a large population, 0.000334 is shown by a dot.

AF, allele frequency; IQR, interquartile range.

Table S1 Statistics in population studies [a)]

Database	Population		
	Name	Participants [b)]	Abbreviation [c)]
jMorp	38KJPN	38722	38KJPN
gnomAD v3.1.2.	European (non-Finnish)	34029	NFE
	African/African American	20744	AFR
	Latino/Admixed American	7647	AMR
	[European (Finnish)]	5316	[FE]
	East Asian	2604	EAS
	South Asian	2419	SAS
	Ashkenazi Jewish	1736	ASJ
	Other	1047	OTH
	Amish	456	AMI
	[Middle Eastern]	158	[ME]
Total		114878	

a) Data are the summation of values from jMorp 38KJPN and gnomAD v3.1.2, which were from https://jmorp.megabank.tohoku.ac.jp/202206/variants/statistics and https://gnomad.broadinstitute.org/help/what-populations-are-represented-in-the-gnomad-data, respectively. Population names, in which variants were not assigned are shown within brackets.

b) Percentage of population is shown in the circle chart on the right. The number of participants in the jMorp 14KJPN was 14,084.

c) Used in columns in Datasets.

Circle chart of participants.

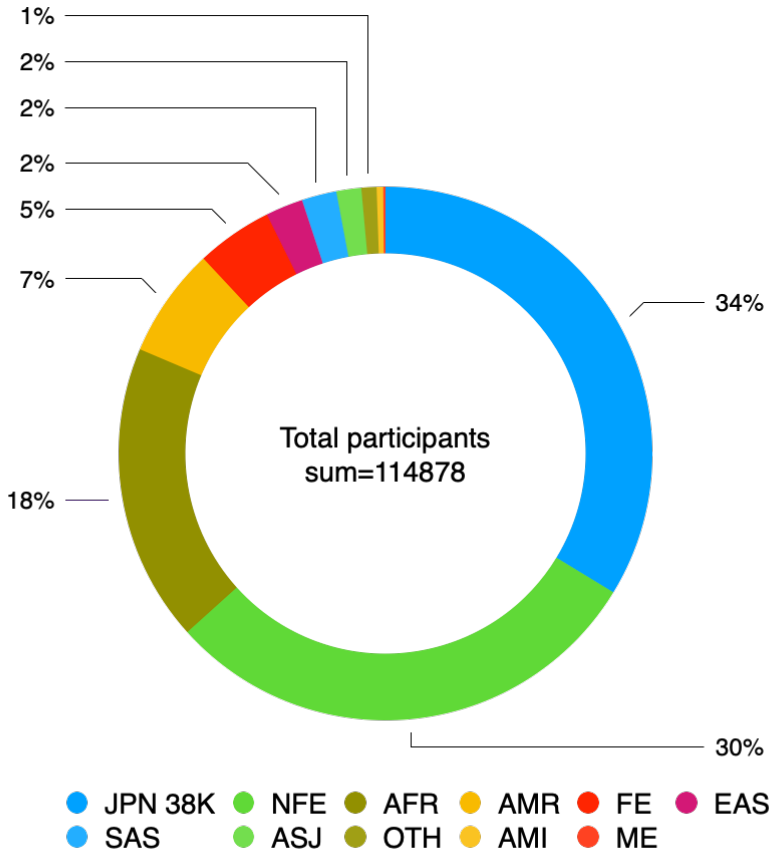

Table S2 Codon frequency of eight genes

Table S2-1 Codon frequency (see a separate file)

Table S2-2 Codon frequency for codons which can produce nonsense mutation [a)]

wild type	AAA	AAG	AGA	CAA	CAG	CGA	GAA	GAG	GGA	TAC	TAC	TAT	TAT	TCA	TCA	TCG	TGC	TGT	TGG	TGG	TTA	TTA	TTG	Total
amino acid	K	K	R	Q	Q	R	E	E	G	Y	Y	Y	Y	S	S	S	C	C	W	W	L	L	L	
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS	7.48	29.91	1.87	1.87	37.38	5.61	1.87	42.99	3.74	13.08	13.08	1.87	1.87	0.00	0.00	20.56	14.95	3.74	20.56	20.56	0.00	0.00	3.74	246.7
CPS1	32.64	36.64	10.66	8.66	23.98	4.00	34.64	29.98	23.32	14.66	14.66	10.66	10.66	15.99	15.99	1.33	5.33	7.99	9.99	9.99	7.99	7.99	18.65	356.4
OTC	42.25	30.99	14.08	19.72	28.17	11.27	33.80	22.54	22.54	11.27	11.27	19.72	19.72	14.08	14.08	0.00	2.82	5.63	14.08	14.08	14.08	14.08	33.80	414.1
ASS1	26.63	53.27	0.00	12.11	29.06	4.84	24.21	60.53	7.26	29.06	29.06	16.95	16.95	2.42	2.42	2.42	9.69	2.42	12.11	12.11	2.42	2.42	4.84	363.2
ASL	12.90	30.11	0.00	4.30	47.31	6.45	10.75	53.76	8.60	17.20	17.20	4.30	4.30	6.45	6.45	8.60	8.60	2.15	21.51	21.51	2.15	2.15	4.30	301.1
ARG1	24.77	49.54	21.67	9.29	6.19	6.19	46.44	12.38	55.73	18.58	18.58	9.29	9.29	9.29	9.29	0.00	0.00	9.29	6.19	6.19	3.10	3.10	18.58	352.9
SLC25A15	33.11	29.80	9.93	3.31	39.74	6.62	19.87	13.25	29.80	26.49	26.49	16.56	16.56	16.56	16.56	0.00	23.18	6.62	6.62	6.62	13.25	13.25	19.87	394.0
SLC25A13	29.54	28.06	13.29	17.73	26.59	11.82	39.88	17.73	39.88	13.29	13.29	14.77	14.77	11.82	11.82	5.91	5.91	4.43	4.43	4.43	17.73	17.73	17.73	382.6
8 genes [b)]	26.91	35.44	8.97	9.63	28.88	6.56	28.00	31.94	23.19	16.63	16.63	11.16	11.16	10.72	10.72	4.81	7.88	5.69	11.59	11.59	7.88	7.88	15.31	349.2
% to all 23 patterns [c)]	7.7%	10.2%	2.6%	2.8%	8.3%	1.9%	8.0%	9.1%	6.6%	4.8%	4.8%	3.2%	3.2%	3.1%	3.1%	1.4%	2.3%	1.6%	3.3%	3.3%	2.3%	2.3%	4.4%	100%

[a)] Permille values of codon frequency listed in Table S2-1. Columns are alphabetically aligned in the order of the nucleotide triplet from the left.

[b)] Codon numbers within the eight genes were divided by 4,571, a total number of codons including termination codons in the eight genes. Values are permille.

[c)] Percentage of values in the raw of '8 genes' relative to the summation of all 23 patterns, '349.2.'

Table S3 Evaluation of patients [a)]

Table S3-1 Inheritance

	de novo			inherited			sum	ND	total
	CGA	non-CGA	no family history	positive family history	non-CGA	CGA			
NAGS	0	0	2	4	1	0	7	2	9
CPS1	0	0	4	5	2	2	13	38	51
OTC	4	6	1	4	4	8	27	67	94
ASS1	0	0	1	1	0	0	2	18	20
ASL	0	0	1	1	1	0	3	69	72
ARG1	0	0	5	5	2	0	12	18	30
SLC25A15	0	0	1	4	0	0	5	8	13
SLC25A13	0	0	2	2	11	6	21	73	94

Table S3-2 Onset

	screening	neonatal	late	childhood	adult	symptomatic [b)]	non-symptomatic [b)]	ND	total
NAGS	0	11	0	0	0	0	0	1	12
CPS1	0	38	2	1	2	0	0	8	51
OTC (m)	0	53	0	1	0	0	0	4	58
OTC (f)	0	3	15	5	3	7	10	9	52
ASS1	1	12	0	1	0	0	0	6	20
ASL	19	20	6	0	0	0	0	27	72
ARG1	6	2	14	5	0	0	0	1	28
SLC25A15	0	3	11	3	0	0	0	0	17

Table S3-3 Outcome (upper, number of cases; lower, number of alleles [c])

	De: death				Sy: symptoms				Al: alive				Cu: cured				total
	neonatal	late	childhood	adult	neonatal	late	childhood	adult	neonatal	late	childhood	adult	neonatal	late	childhood	adult	
NAGS	6	0	0	0	0	0	0	0	4	0	0	0	1	0	0	0	11
CPS1	20	0	0	0	7	1	1	0	3	0	0	2	2	0	0	0	36
OTC (m)	9	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	12
OTC (f)	1	1	0	0	0	3	0	0	0	0	0	0	0	1	0	0	6
ASS1	2	0	0	0	1	0	0	0	2	0	1	0	0	0	0	0	6
ASL	3	1	0	0	3	1	0	0	2	0	0	0	0	0	0	0	10
ARG1	0	0	0	0	0	11	5	0	2	0	0	0	0	1	0	0	19
SLC25A15	0	0	1	0	2	9	2	0	1	1	0	0	0	0	0	0	16

	De: death				Sy: symptoms				Al: alive				Cu: cured				total
	neonatal	late	childhood	adult	neonatal	late	childhood	adult	neonatal	late	childhood	adult	neonatal	late	childhood	adult	
NAGS	**12**	0	0	0	0	0	0	0	**7**	0	0	0	**2**	0	0	0	**21**
CPS1	**23**	0	0	0	**8**	1	1	0	3	0	0	2	2	0	0	0	**40**
OTC (m)	9	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	12
OTC (f)	1	1	0	0	0	3	0	0	0	0	0	0	0	1	0	0	6
ASS1	2	0	0	0	1	0	0	0	2	0	1	0	0	0	0	0	6
ASL	**5**	1	0	0	3	1	0	0	**4**	0	0	0	0	0	0	0	**14**
ARG1	0	0	0	0	0	**13**	**8**	0	**3**	0	0	0	0	1	0	0	**25**
SLC25A15	0	0	**2**	0	2	**15**	**4**	0	1	**2**	0	0	0	0	0	0	**26**

[a] Inheritance, onset, and outcome of patients in Dataset 2 are classified as described in the main text and Supplementary Information.

[b] Classified according to authors' description for female *OTC* gene deficiency [16].

[c] Presence of nonsense alleles from homozygotes or compound heterozygotes was indicated with bold font.

ND, no data; *OTC* (m) or (f), male or female cases of *OTC* gene deficiency.

Table S4 Numbers of disease-causing nonsense mutations in the eight genes associated with primary hyperammonemia

S4-1. Patient mutation sites [a)]

wild type amino acid	AAA K	AAG K	AGA R	CAA Q	CAG Q	CGA R	GAA E	GAG E	GGA G	TAC Y	TAC Y	TAT Y	TAT Y	TCA S	TCA S	TCG S	TGC C	TGT C	TGG W	TGG W	TTA L	TTA L	TTG L	Total
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS					1			2											1	1				5
CPS1			1	1	7	6	2	2		2	1		1	1	1	1			2	1			2	31
OTC	3	1	1	3	7	4	4	3	3	1	1	1	1	3	2			1	3	5	1	2		50
ASS1				2	3	2			1	3			1				1							13
ASL					7	3		1	2	1		1	1							3				19
ARG1	1					2	1		1										1					6
SLC25A15	1			1	2	2																		6
SLC25A13				3	4	8	4	1	1		1					1			2	1				26
8 genes	5	1	2	10	31	27	11	9	8	7	3	2	4	4	3	2	1	1	9	11	1	2	2	156
percentage	3.2%	0.6%	1.3%	6.4%	19.9%	17.3%	7.1%	5.8%	5.1%	4.5%	1.9%	1.3%	2.6%	2.6%	1.9%	1.3%	0.6%	0.6%	5.8%	7.1%	0.6%	1.3%	1.3%	100%

[a)] Numbers of patient mutation sites identified in each gene are listed along with all of the possible 23 nucleotide change patterns that can cause nonsense mutation. Mutations causing TAA, TAG, and TGA number 40, 61, and 55, respectively. Base transition:base transversion = 88:68.

S4-2. Occurrence of mutation [b)]

wild type amino acid	AAA K	AAG K	AGA R	CAA Q	CAG Q	CGA R	GAA E	GAG E	GGA G	TAC Y	TAC Y	TAT Y	TAT Y	TCA S	TCA S	TCG S	TGC C	TGT C	TGG W	TGG W	TTA L	TTA L	TTG L	Total
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS					2			2											4	1				9
CPS1			1	1	9	21	2	2		3	1		1	1	2	1			2	1			3	51
OTC	3	1	1	3	9	40	5	3	4	1	1	1	1	3	3			1	4	7	1	2		94
ASS1				4	3	6			1	4			1				1							20
ASL					39	24		1	2	1		1	1							3				72
ARG1	1					22	1		2										4					30
SLC25A15	1			1	2	9																		13
SLC25A13				4	5	34	12	3	2		1					30			2	1				94
8 genes	5	1	2	13	69	156	20	11	11	9	3	2	4	4	5	31	1	1	16	13	1	2	3	383
percentage	1.3%	0.3%	0.5%	3.4%	18.0%	40.7%	5.2%	2.9%	2.9%	2.3%	0.8%	0.5%	1.0%	1.0%	1.3%	8.1%	0.3%	0.3%	4.2%	3.4%	0.3%	0.5%	0.8%	100%

[b)] Frequency of patients with mutation at each specific site. Mutation observed in the same family was counted as a single occurrence. Details of the count are shown in the column “family” in Dataset 2. Mutations causing TAA, TAG, and TGA number 54, 138, and 191, respectively. Base transition: base transversion = 267:116.

S4-3. Frequency of patient mutation sites (S4-1) adjusted by codon usage [c)]

wild type amino acid	AAA K	AAG K	AGA R	CAA Q	CAG Q	CGA R	GAA E	GAG E	GGA G	TAC Y	TAC Y	TAT Y	TAT Y	TCA S	TCA S	TCG S	TGC C	TGT C	TGG W	TGG W	TTA L	TTA L	TTG L	Total
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS					0.27			0.05											0.05	0.05				0.41
CPS1			0.09	0.12	0.29	1.50	0.06	0.07		0.14	0.07		0.09	0.06	0.06	0.75			0.20	0.10			0.11	3.71
OTC	0.07	0.03	0.07	0.15	0.25	0.36	0.12	0.13	0.13	0.09	0.09	0.05	0.05	0.21	0.14			0.18	0.21	0.36	0.07	0.14		2.91
ASS1				0.17	0.10	0.41			0.14	0.10			0.06				0.10							1.08
ASL					0.15	0.47		0.02	0.23	0.06		0.23	0.23							0.14				1.53
ARG1	0.04					0.32	0.02		0.02										0.16					0.56
SLC25A15	0.03			0.30	0.05	0.30																		0.68
SLC25A13				0.17	0.15	0.68	0.10	0.06	0.03		0.08					0.17			0.45	0.23				2.10
8 genes	0.19	0.03	0.22	1.04	1.07	4.11	0.39	0.28	0.34	0.42	0.18	0.18	0.36	0.37	0.28	0.42	0.13	0.18	0.78	0.95	0.13	0.25	0.25	12.55
% to all 23 patterns	1.5%	0.2%	1.8%	8.3%	8.6%	32.8%	3.1%	2.2%	2.7%	3.4%	1.4%	1.4%	2.9%	3.0%	2.2%	3.3%	1.0%	1.4%	6.2%	7.6%	1.0%	2.0%	2.0%	100%

S4-4. Occurrence of mutation (S4-2) adjusted by codon usage [d)]

wild type amino acid	AAA K	AAG K	AGA R	CAA Q	CAG Q	CGA R	GAA E	GAG E	GGA G	TAC Y	TAC Y	TAT Y	TAT Y	TCA S	TCA S	TCG S	TGC C	TGT C	TGG W	TGG W	TTA L	TTA L	TTG L	Total
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS					0.05			0.05											0.19	0.05				0.34
CPS1			0.09	0.12	0.38	5.25	0.06	0.07		0.20	0.07			0.06	0.13	0.75			0.20	0.10			0.16	7.63
OTC	0.07	0.03	0.07	0.15	0.32	3.55	0.15	0.13	0.18	0.09	0.09	0.05	0.05	0.21	0.21			0.18	0.28	0.50	0.07	0.14		6.53
ASS1				0.33	0.10	1.24			0.14	0.14			0.06				0.10							2.11
ASL					0.82	3.72		0.02	0.23	0.06		0.23	0.23							0.14				5.46
ARG1	0.04					3.55	0.02		0.04										0.65					4.30
SLC25A15	0.03			0.30	0.05	1.36																		1.74
SLC25A13				0.23	0.19	2.87	0.30	0.17	0.05		0.08					5.08			0.45	0.23				9.64
8 genes	0.19	0.03	0.22	1.35	2.39	23.78	0.71	0.34	0.47	0.54	0.18	0.18	0.36	0.37	0.47	6.44	0.13	0.18	1.38	1.12	0.13	0.25	0.20	41.41
% to all 23 patterns	0.4%	0.1%	0.5%	3.3%	5.8%	57.4%	1.7%	0.8%	1.1%	1.3%	0.4%	0.4%	0.9%	0.9%	1.1%	15.5%	0.3%	0.4%	3.3%	2.7%	0.3%	0.6%	0.5%	100%

[c)] Values in Table S4-1 were divided by the corresponding values shown in Table 2-2.

[d)] Values in Table S4-2 were adjusted as described in the footnote c).

Table S6 Number of nonsense variations in general populations in the CDS of the MANE Select transcripts (MANE Select CDS) in the eight genes causing primary hyperammonemia [a)]

S6-1. Variation sites

wild type	AAA	AAG	AGA	CAA	CAG	CGA	GAA	GAG	GGA	TAC	TAC	TAT	TAT	TCA	TCA	TCG	TGC	TGT	TGG	TGG	TTA	TTA	TTG	Total
amino acid	K	K	R	Q	Q	R	E	E	G	Y	Y	Y	Y	S	S	S	C	C	W	W	L	L	L	
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS		1			2						1					2				2				8
CPS1				2	3	3	1	1	1	1	1			1				1		1				16
OTC																								0
ASS1					1	2		1		1														5
ASL [b)]		1			1	2													1	1				6 [b)]
ARG1 [c)]						2 [c)]	1							1					1					5 [c)]
SLC25A15						2								1					1					4
SLC25A13				1	2 [d)]	8	1	0 [e)]					1		1	1			1	0 [f)]				16 [d-f)]
8 genes	0	2	0	3	9 [d)]	19 [c)]	3	2 [e)]	1	2	2	0	1	3	1	3	0	1	4	4 [f)]	0	0	0	60 [b-f)]
percentage	0.0%	3.3%	0.0%	5.0%	15.0%	31.7%	5.0%	3.3%	1.7%	3.3%	3.3%	0.0%	1.7%	5.0%	1.7%	5.0%	0.0%	1.7%	6.7%	6.7%	0.0%	0.0%	0.0%	100%

S6-2. Number of nonsense alleles

wild type	AAA	AAG	AGA	CAA	CAG	CGA	GAA	GAG	GGA	TAC	TAC	TAT	TAT	TCA	TCA	TCG	TGC	TGT	TGG	TGG	TTA	TTA	TTG	Total
amino acid	K	K	R	Q	Q	R	E	E	G	Y	Y	Y	Y	S	S	S	C	C	W	W	L	L	L	
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS		1			4						1					3				2				11
CPS1				2	3	19	1	1	1	2	1			1				1		1				33
OTC																								0
ASS1					4	3		1		5														13
ASL [b)]		1			1	18													1	1				22 [b)]
ARG1 [c)]						8 [c]	1							2					5					16 [c)]
SLC25A15						36								1					1					38
SLC25A13				1	12 [d)]	24	5	0 [e)]					1		1	33			7	0 [f)]				84 [d-f)]
8 genes	0	2	0	3	24 [d)]	108 [c]	7	2 [e)]	1	7	2	0	1	4	1	36	0	1	14	4 [f)]	0	0	0	217 [b-]
percentage	0.0%	0.9%	0.0%	1.4%	11.1%	49.8%	3.2%	0.9%	0.5%	3.2%	0.9%	0.0%	0.5%	1.8%	0.5%	16.6%	0.0%	0.5%	6.5%	1.8%	0.0%	0.0%	0.0%	100%

[a)] Numbers of nonsense variations within the Mane Select CDS identified in either jMorp 38KJPN or gnomAD v3.1.2.
[b)] Not including one variation at the termination codon from TAG to TAA found in the 38KJPN. Reported from 11 alleles.
[c)] Not including one CGA>TGA variation at codon 20 in transcript ENST00000672233.1. Reported from four alleles in the NFE population.
[d)] Not including one CAG>TAG variation at codon 312 in the longest protein isoform. Reported from 15 alleles in the AFR and NEF populations.
[e)] Not including one GAG>TAG variation at codon 26 in transcript XM_017011663. Reported from one allele in the 38KJPN population.
[f)] Not including one TGG>TGA variation at codon 10 in transcript XM_017011663. Reported from one allele in the NFE population.

S6-3. Frequency of variation sites (S6-1) adjusted by codon usage [g)]

wild type	AAA	AAG	AGA	CAA	CAG	CGA	GAA	GAG	GGA	TAC	TAC	TAT	TAT	TCA	TCA	TCG	TGC	TGT	TGG	TGG	TTA	TTA	TTG	Total
amino acid	K	K	R	Q	Q	R	E	E	G	Y	Y	Y	Y	S	S	S	C	C	W	W	L	L	L	
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS		0.06			0.10						0.14					0.18				0.18				0.67
CPS1				0.15	0.08	0.50	0.02	0.02	0.03	0.05	0.05			0.04				0.08		0.07				1.09
OTC																								0.00
ASS1					0.08	1.00		0.04		0.08														1.21
ASL		0.07			0.05	0.67													0.10	0.10				0.98
ARG1						1.00	0.07							0.33					0.50					1.90
SLC25A15						1.00								0.20					0.50					1.70
SLC25A13				0.08	0.11	1.00	0.04	0.00					0.10		0.13	0.25			0.33	0.00				2.04
8 genes	0.00	0.01	0.00	0.07	0.07	0.63	0.02	0.01	0.01	0.03	0.03	0.00	0.02	0.06	0.02	0.14	0.00	0.04	0.08	0.08	0.00	0.00	0.00	1.31
% to all 23 patterns	0.0%	0.9%	0.0%	1.4%	11.1%	49.8%	3.2%	0.9%	0.5%	3.2%	0.9%	0.0%	0.5%	1.8%	0.5%	16.6%	0.0%	0.5%	6.5%	1.8%	0.0%	0.0%	0.0%	100%

S6-4. Number of nonsense alleles (S6-2) adjusted by codon usage [h)]

wild type	AAA	AAG	AGA	CAA	CAG	CGA	GAA	GAG	GGA	TAC	TAC	TAT	TAT	TCA	TCA	TCG	TGC	TGT	TGG	TGG	TTA	TTA	TTG	Total
amino acid	K	K	R	Q	Q	R	E	E	G	Y	Y	Y	Y	S	S	S	C	C	W	W	L	L	L	
mutation	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TGA	TAA	TAG	TAA	TAG	TAA	TGA	TAG	TGA	TGA	TAG	TGA	TAA	TGA	TAG	
NAGS		0.06			0.20						0.14					0.27				0.18				0.86
CPS1				0.15	0.08	3.17	0.02	0.02	0.03	0.09	0.05			0.04				0.08		0.07				3.80
OTC																								0.00
ASS1					0.33	1.50		0.04		0.42														2.29
ASL		0.07			0.05	6.00													0.10	0.10				6.32
ARG1						4.00	0.07							0.67					2.50					7.23
SLC25A15						18.00								0.20					0.50					18.70
SLC25A13				0.08	0.67	3.00	0.19	0.00					0.10		0.13	8.25			2.33	0.00				14.74
8 genes	0.00	0.01	0.00	0.07	0.18	3.60	0.05	0.01	0.01	0.09	0.03	0.00	0.02	0.08	0.02	1.64	0.00	0.04	0.26	0.08	0.00	0.00	0.00	6.19
% to all 23 patterns	0.0%	0.2%	0.0%	1.1%	2.9%	58.1%	0.9%	0.2%	0.2%	1.5%	0.4%	0.0%	0.3%	1.3%	0.3%	26.4%	0.0%	0.6%	4.3%	1.2%	0.0%	0.0%	0.0%	100%

[g)] Values in Table S6-1 were divided by corresponding values shown in Table 2-2.

[h)] Values in Table S6-2 were adjusted as described in the footnote g).