Human Genetics (2019) 138:389–409
https://doi.org/10.1007/s00439-019-01998-7

ORIGINAL INVESTIGATION

Estimating carrier frequencies of newborn screening disorders using a whole-genome reference panel of 3552 Japanese individuals

Yumi Yamaguchi-Kabata[1,2] · Jun Yasuda[1,2,3] · Akira Uruno[1,2] · Kazuro Shimokawa[1] · Seizo Koshiba[1] · Yoichi Suzuki[1,2,4] · Nobuo Fuse[1,2] · Hiroshi Kawame[1,2] · Shu Tadaka[1] · Masao Nagasaki[1,2,5] · Kaname Kojima[1,2,5] · Fumiki Katsuoka[1,2] · Kazuki Kumada[1] · Osamu Tanabe[1,2,6] · Gen Tamiya[1,2,7] · Nobuo Yaegashi[1,2] · Kengo Kinoshita[1,5,8,9] · Masayuki Yamamoto[1,2,9] · Shigeo Kure[1,2] · The Tohoku Medical Megabank Project Study Group

Received: 7 January 2019 / Accepted: 6 March 2019 / Published online: 18 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Incidence rates of Mendelian diseases vary among ethnic groups, and frequencies of variant types of causative genes also vary among human populations. In this study, we examined to what extent we can predict population frequencies of recessive disorders from genomic data, and explored better strategies for variant interpretation and classification. We used a whole-genome reference panel from 3552 general Japanese individuals constructed by the Tohoku Medical Megabank Organization (ToMMo). Focusing on 32 genes for 17 congenital metabolic disorders included in newborn screening (NBS) in Japan, we identified reported and predicted pathogenic variants through variant annotation, interpretation, and multiple ways of classifications. The estimated carrier frequencies were compared with those from the Japanese NBS data based on 1,949,987 newborns from a previous study. The estimated carrier frequency based on genomic data with a recent guideline of variant interpretation for the *PAH* gene, in which defects cause hyperphenylalaninemia (HPA) and phenylketonuria (PKU), provided a closer estimate to that by the observed incidence than the other methods. In contrast, the estimated carrier frequencies for *SLC25A13*, which causes citrin deficiency, were much higher compared with the incidence rate. The results varied greatly among the 11 NBS diseases with single responsible genes; the possible reasons for departures from the carrier frequencies by reported incidence rates were discussed. Of note, (1) the number of pathogenic variants increases by including additional lines of evidence, (2) common variants with mild effects also contribute to the actual frequency of patients, and (3) penetrance of each variant remains unclear.

Introduction

Mendelian diseases, which show distinct inheritance patterns due to genetic changes in a single gene, and many of them manifest serious conditions at early ages. Although the prevalence is generally very low for recessive diseases, carrier frequency is relatively high. Causative variants exist at low frequencies in human populations; therefore, clarifying the type and frequency of causative variants in a population is not simple. Rates of incidence and the types of genetic variant causing Mendelian diseases vary among ethnic groups. Information on variant frequencies for Mendelian diseases is not well known for the Japanese population.

In the 1950s and 1960s, amount of recessive mutations in the human genome was estimated by analyzing the rate of early infant deaths caused by consanguineous marriages in target populations (Morton et al. 1956). Since the 1960s, genetic polymorphisms of enzyme loci have been examined with electrophoresis, and the rates of null mutations of the enzymes, which were critical for genetic diseases, have been reported (Hamilton et al. 1961; Satoh et al. 1983). Since the 1980s, many causative genes for Mendelian diseases have been identified, and genetic variations in those genes have been examined. However, until recently,

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00439-019-01998-7) contains supplementary material, which is available to authorized users.

✉ Yumi Yamaguchi-Kabata
yamaguchi@megabank.tohoku.ac.jp

✉ Kengo Kinoshita
kengo@ecei.tohoku.ac.jp

Extended author information available on the last page of the article

patient genomes were examined mainly to identify causative variants for a genetic disease. Identified causative variants have been reported in scientific articles, and were curated in mutation databases, such as the Human Gene Mutation Database (HGMD) (Cooper et al. 1998; Stenson et al. 2003) or ClinVar database (Landrum et al. 2016). However, frequencies of these causative variants in the general population are largely unknown. One reason is that each causative variant exists at a low frequency in the general population, and it usually requires the genomes of a large number of individuals to be sequenced, which is costly. Therefore, previous reports on the frequency of recessive diseases focused on limited numbers of diseases and genomic sites (Chong et al. 2012; Song et al. 2012). Owing to technical progress and decreasing costs (Fujimoto et al. 2010; Levy et al. 2007; Roach et al. 2010), sequencing an individual's entire genome or exome has become feasible, and is a powerful approach in medical genomics(Bell et al. 2011; Yang et al. 2013).

These days, genome cohort studies are ongoing in many countries to expedite genomic research for personalized medicine, prevention, and healthcare. These studies utilizing whole-genome sequencing (WGS) or whole-exome sequencing (WES) provide huge genomic datasets for a given population, providing the opportunity to detect low-frequency variants and allow for estimating frequencies of those variants (UK 10K Consortium et al. 2015; Gudbjartsson et al. 2015; Lek et al. 2016), including pathogenic variants (Amendola et al. 2015; Dorschner et al. 2013; Tabor et al. 2014; Xue et al. 2012).

However, it is not easy to identify real pathogenic variants, even with the use of a good panel of WGS or WES and the latest databases of pathogenic variants (Rehm et al. 2015; Stenson et al. 2003). For example, the use of reported pathogenic variants from databases may include false positives, and therefore, the relationship between variant and disease needs to be evaluated (Yamaguchi-Kabata et al. 2018). Furthermore, genomic variants among non-European populations are underrepresented in the databases of reported pathogenic variants (Kessler et al. 2016). In addition, real causative variants may exist, which have not been reported yet. A recent guideline of variant interpretation by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) (Richards et al. 2015) proposed criteria to classify variants into five classes (pathogenic, likely pathogenic, variants of uncertain significance, likely benign, and benign) in a gradient from benign to pathogenic. This guideline enables us to detect possible pathogenic variants, including unreported variants as pathogenic.

Studies on recessive genetic effects in the Japanese population started as estimations of mutational damages as a result of consanguineous marriages, focusing on Nagasaki and Hiroshima through a collaboration with the Atomic Bomb Causality Commission (ABCC) after World War II (Neel and Schull 1962; Schull et al. 1968, 1970; Yamaguchi et al. 1970). Since 1980s, genetic variations in the Japanese population have been investigated at the DNA level from the standpoint of human history (Hammer and Horai 1995; Horai et al. 1996; Japanese Archipelago Human Population Genetics Consortium et al. 2012) and a high-quality SNP catalogue based on gene-based SNP discovery lead early success of genome-wide association studies (Haga et al. 2002; Hirakawa et al. 2002). The Japanese mainland (Hondo) has a relatively homogeneous population, as genetic differentiations among local regions are very small (Yamaguchi-Kabata et al. 2008). It has been reported that incidence rates of congenital metabolic disorders are lower in the Japanese than in Caucasians (Yamaguchi 2008). On the other hand, there are Japanese-specific diseases, such as Nakajo-Nishimura syndrome (MIM: 256040) (Nakajo 1939; Nishimura 1950), and diseases showing higher incidence rates in the Japanese population, such as APRT deficiency (MIM: 614723) (Mimori et al. 1991; Kamatani et al. 1990) and moyamoya disease (MIM: 617151) (Kamada et al. 2011). Therefore, clarifying pathogenic variant allele frequencies that account for incidence rates of genetic diseases are desired.

The Tohoku University Medical Megabank Organization (ToMMo) along with Iwate Medical University started genome cohort studies with a biobank that integrated medical and genomic information of cohort participants with the general population (Kuriyama et al. 2016). We used whole-genome sequences of Japanese individuals to construct a whole-genome reference panel to expedite medical genomics research (Kawai et al. 2015; Minari et al. 2018; Nagasaki et al. 2015; Yamaguchi-Kabata et al. 2015; Yasuda et al. 2018). Recently, we constructed the whole-genome reference panel (3.5KJPNv2), which contains approximately 47 million autosomal genomic variants (di-allelic SNVs and short indels, filtered by VQSR), including abundant low-frequency variants.

In this study, focusing on congenital metabolic disorders, as targets for newborn screening (NBS) in Japan (Kitagawa 2012; Yamaguchi 2012), we (1) examined to what extent we can predict the frequency of recessive disorders from genomic data, and (2) explored better strategies of variant interpretation and classification, using the 3.5 KJPNv2 whole-genome reference panel. We identified pathogenic variants for congenital metabolic disorders through NBS disease genes by annotation, interpretation, and classification by multiple methods. Then, we estimated carrier frequencies for specific NBS genes, and compared them to the observed incidence rates reported NBS study. This is a precious opportunity to examine the relationship between genomic variations and disease incidence rates.

Materials and methods

This project was performed as a part of prospective cohort studies at ToMMo with the approval of the Ethical Committee at the Tohoku University School of Medicine and ToMMo. Samples were obtained from the cohort participants, all of whom gave their written consent. Under the terms of the informed consent provided by the participants in our cohort project, whole-genome data, including sequenced data, variant calls, and inferred genotypes were securely controlled under the Materials and Information Distribution Review Committee of Tohoku Medical Megabank Project. Sharing of data with other researchers was discussed in each research proposal by the review committee.

Data source

We used the 3.5KJPNv2 whole-genome reference panel (Tadaka et al. 2019), which was constructed based on sequencing the whole genome of 3552 healthy Japanese individuals from the Tohoku Medical Megabank Project. Genomic DNA extracted from peripheral blood samples of the 3552 individuals were subjected to paired-end sequencing using the Illumina HiSeq 2500 platform as previously described (Nagasaki et al. 2015). This 3.5KJPNv2 panel was constructed through an approach based on GATK Best Practices. Data on allele frequencies for 3.5KJPNv2 variants are publicly available through the portal site, Japanese Multi Omics Reference Panel (https://jmorp.megabank.tohoku.ac.jp/) (Tadaka et al. 2018) or NBDC Human Database with accession number hum0015 (https://humandbs.biosciencedbc.jp/en/hum0015-v3). Di-allelic variants (after VQSR filtering) were used for further analysis.

Variant annotation and primary interpretation

An initial step of variant interpretation was conducted using InterVar (Li and Wang 2017) (version 0.1.7), which is based on ACMG-AMP guidelines of variant interpretation (Richards et al. 2015) and variant annotation by Annovar (Wang et al. 2010). Annovar annotation output included gene-based functional annotation, predicted pathogenicity by several methods, ClinVar (Landrum et al. 2016) (2016.3), and allele frequencies of other populations from external population databases. Among the 28 criteria distinguishing pathogenic from benign variants in the ACMG-AMP guidelines, 18 criteria (PVS1, PS1, PS4, PM1, PM2, PM4, PM5, PP2, PP3, PP5, BA1, BS1, BS2, BP1, BP3, BP4, BP6, and BP7) were implemented in InterVar for automatic interpretation using variant annotation with open data. Using InterVar as an initial step of variant interpretation, approximately 47 M variants in 3.5KJPNv2 (after VQSR filtering) were classified into five classes: pathogenic (P), likely pathogenic (LP), variants of uncertain significance (VUS), likely benign (LB), and benign (B). In addition, overlaps between variants in the Human Gene Mutation Database (HGMD) professional (2017.4) (Cooper et al. 1998) and 3.5KJPNv2 were examined based on genomic coordinates and on the consistency of the allele bases.

Genes for newborn screening (NBS)

Information on NBS diseases and their responsible genes was obtained from the Genetic Home Reference website (https://ghr.nlm.nih.gov/), and obtained a list of 79 autosomal genes, after excluding X-linked genes. In this study, we focused on 32 genes for 17 congenital metabolic disorders examined at a higher priority due to medical actionability in NBS tests in Japan (Kitagawa 2012; Yamaguchi 2012) (Table 1). We analyzed the allele frequencies of the reported pathogenic variants in the genes responsible for the congenital metabolic disorders. Incidence rate data (x) for these diseases were obtained from the domestic report (Yamaguchi 2012), based on 1,949,987 newborns. Confidence intervals (95%) of the incidences were calculated based on a binomial distribution (Table 1). If there were no affected individuals detected through the NBS, the upper limit was calculated based on Poisson distribution. For diseases with a single responsible gene, carrier frequency was estimated. Assuming a recessive disorder, the frequency of a chromosome with any risk allele in a population (Q) was estimated by the square root of the incidence rate (Q = square root of (x)). The frequency of carriers was estimated by $2*Q*(1-Q)$.

Classification of pathological variants

Using the primary annotation output for 3.5KJPNv2, variants in 32 NBS genes were selected for analysis by including 1 kb upstream and downstream regions. Starting with the output of variant interpretation by InterVar (five classes: P, LP, VUS, LB, and B), we classified variants based on (1) filtering by allele frequency in 3.5KJPNv2 (allele frequency < 0.03) (Tabor et al. 2014), (2) if the variant was previously reported in ClinVar (classes P and LP), and (3) seeking for additional reported pathogenic variants in class VUS (Fig. 1). For variants categorized into class P or LP, they were divided into "previously reported" or "interpreted (predicted)" based on if the variant corresponded to ClinVar variants (class P or LP). Among class VUS variants, we sought additional candidates for pathogenic variants [ClinVar variants and HGMD-DM (disease-causing mutations) variants].

An estimation of the frequencies of risk alleles and carriers in the population was conducted four ways of including pathogenic variants (Fig. 1). First, conservatively selecting

Table 1 Congenital metabolic disorders for Japanese newborn screening and reported incidence rates

Disease	MIM#	Incidence rate in a previous NBS study					Expected carrier frequency**	Gene
		No. of case*	Incidence (x)*	Lower limit (x)	Upper limit (x)	1/x		
Phenylketonuria/hyperphenylalaninemia	261600	37	1.9E−05	1.3E−05	2.5E−05	39,859.0	0.00867	*PAH*
Argininosuccinic aciduria	207900	2	1.0E−06	− 4.0E−07	2.4E−06	408,642.6	0.00202	*ASL*
Citrullinemia type 1	215700	6	3.1E−06	6.1E−07	5.5E−06	180,537.9	0.00350	*ASS1*
Citrin deficiency***	605814	23	1.2E−05	7.0E−06	2.0E−05	84,782.0	0.00685	*SLC25A13*
Maple syrup urine disease	248600	1	5.1E−07	− 4.9E−07	1.5E−06	658,779.4		*BCKDHA* *BCKDHB* *DBT* *DLD*
Homocystinuria	236200; 236270; 250940; 236250	3	1.5E−06	− 2.0E−07	3.3E−06	304,932.7		*CBS* *MMADHC* *MTHFR* *MTR* *MTRR*
Glutaric acidemia type I	231670	7	3.6E−06	9.3E−07	6.2E−06	160,023.0	0.00378	*GCDH*
Multiple carboxylase synthetase deficiency	253270; 253260	3	1.5E−06	− 2.0E−07	3.3E−06	304,932.7		*BTD* *HLCS*
Lsovaleric acidemia	243500	3	1.5E−06	− 2.0E−07	3.3E−06	304,932.7	0.00248	*IVD*
3-Hydroxy-3-methylglutaryl-CoA lyase deficiency****	246450	0	0.0E+00	NA	2.6E−08	NA		*HMGCL*
3-Methylcrotonyl-CoA carboxylase deficiency	210200; 210210	13	6.7E−06	3.0E−06	1.0E−05	97,174.5		*MCCC1* *MCCC2*
Propionic acidemia	606054	43	2.2E−05	1.5E−05	2.9E−05	74,100.2		*PCCA* *PCCB*
Methylmalonic acidemia	251000; 277400; 251100; 251110	18	9.2E−06	5.0E−06	1.3E−05	74,100.2		*MCEE* *MMAA* *MMAB* *MUT*
Medium-chain acyl-CoA dehydrogenase deficiency	201450	18	9.2E−06	5.0E−06	1.3E−05	74,100.2	0.00606	*ACADM*
Very long-chain acyl-CoA dehydrogenase deficiency	201475	12	6.2E−06	2.7E−06	9.6E−06	103,780.0	0.00495	*ACADVL*
Carnitine palmitoyltransferase I deficiency	255120	5	2.6E−06	3.2E−07	4.8E−06	207,828.2	0.00320	*CPT1A*
Carnitine palmitoyltransferase II deficiency***	600649	7	3.6E−06	9.3E−07	6.2E−06	160,023.0	0.00378	*CPT2*

Table 1 (continued)

Disease	MIM#	Incidence rate in a previous NBS study					Expected carrier frequency**	Gene
		No. of case*	Incidence (x)*	Lower limit (x)	Upper limit (x)	1/x		
Mitochondrial trifunctional protein deficiency	609015	2	1.0E−06	− 4.0E−07	2.4E−06	408,642.6		*HADHA HADHB*

*These statistics were from a domestic NBS report (Yamaguchi 2012) based on 1,949,987 newborns. Confidence intervals (95% CI) of incidence rates were calculated

**Calculated based on the reported incidence rates and recessive mode for diseases with single responsible gene

***Disease as the second target in Japanese NBS. *SLC25A13* is also responsible for citrullinemia type 2. *CPT2* is included in our analysis because we suspected that the frequency of carriers in Japan is higher than that of *CPT1A*

****Only upper limit was calculated based on Poisson distribution

pathogenic variants, we used variants interpreted as class P or LP. In the most conservative set of pathogenic variants, we used class P and LP variants corresponding to ClinVar variants as P or LP (Set 1). The second set included all class P and LP variants (Set 2). To utilize additional pathogenic variants reported among class VUS, the third set consisted of Set 2 variants and ClinVar variants (P or LP) in class VUS (Set 3). The fourth group included disease-causing mutation (DM) variants in HGMD as well as the variants in Set 3 (Set 4). For each set of pathogenic variants, the sum of risk allele frequencies (as population frequency of risk alleles), and the expected frequency of heterozygous individuals were estimated for each class (see below).

Estimation of population frequency of risk alleles and expected carriers

Suppose there are n genomic sites of pathogenic variants for a disease gene. First, we calculated the sum of frequencies of risk alleles for n sites, as an estimated population frequency of pathogenic alleles of the gene (Q):

$$Q = \sum_{i=1}^{n} q_i.$$

We assumed that a haploid genome in the population has any risk allele for the disease with the probability Q, and does not have any risk allele with the probability, $1-Q$. Then, we calculated the expected carrier frequency by 2*(1 − Q)*Q based on the Hardy–Weinberg equilibrium.

Results

Genomic variants in 3.5KJPNv2 (46,873,740 di-allelic variants after variant quality score recalibration filtering) were tentatively classified into pathogenic ($n = 1624$), likely pathogenic ($n = 3388$), variants of uncertain significance ($n = 40{,}163{,}877$), likely benign ($n = 162{,}587$), and benign ($n = 6{,}542{,}264$) based on variant annotation and variant interpretation using InterVar (Li and Wang 2017) based on 18 of the 28 criteria of the ACMG-AMP guidelines of variant interpretation (Richards et al. 2015) (See methods and Table S1). Then 73,787 variants in 79 NBS genes (see Materials and Methods), including 5′ and 3′ 1000-bp flanking regions, were selected for further analysis. Starting with the output of primary interpretation (mentioned above), we selected possible pathogenic variants in these genes by four methods, from conservative criteria to broad selection including candidates (Fig. 1), based on filtering by minor allele frequency (MAF) of < 0.03, the output of InterVar, and matching with variants in ClinVar/HGMD (Table S2). Here, we focus on 32 genes for 17 diseases, as primary targets for NBS in Japan (Kitagawa 2012; Yamaguchi 2012), and variants for the 32 genes were used for further analysis (Table 2, see Table S3 for an integrated form of Tables 1, 2). Except for the *MMAB* gene, at least one selected pathogenic variant was found for 31 NBS genes in Set 4.

The total number of variants classified as P or LP by InterVar was 88 for the 32 NBS genes. In Set 1, the most conservative selection, the only reported variants in ClinVar were included, and 41 of the 88 variants were detected. In Set 2, 23 of the 32 genes had additional variants interpreted. The number of selected pathogenic variants increased in 11 genes in Set 3 compared to Set 2. This increase was due to the inclusion of ClinVar variants that did not have sufficient score in the 18 criteria for being interpreted as class P or LP. Comparing the results of Set 4, which included the use of HGMD to Set 3, 19 genes had additional pathogenic variant candidates, while 12 genes did not.

The number of selected pathogenic variants varied among the 32 genes. In Set 2, the genes showing the highest allele counts were *SLC25A13* ($n = 71$), *ACADVL* ($n = 21$), *PAH* ($n = 20$), and *GCDH* ($n = 18$). However, in Set 4, the genes

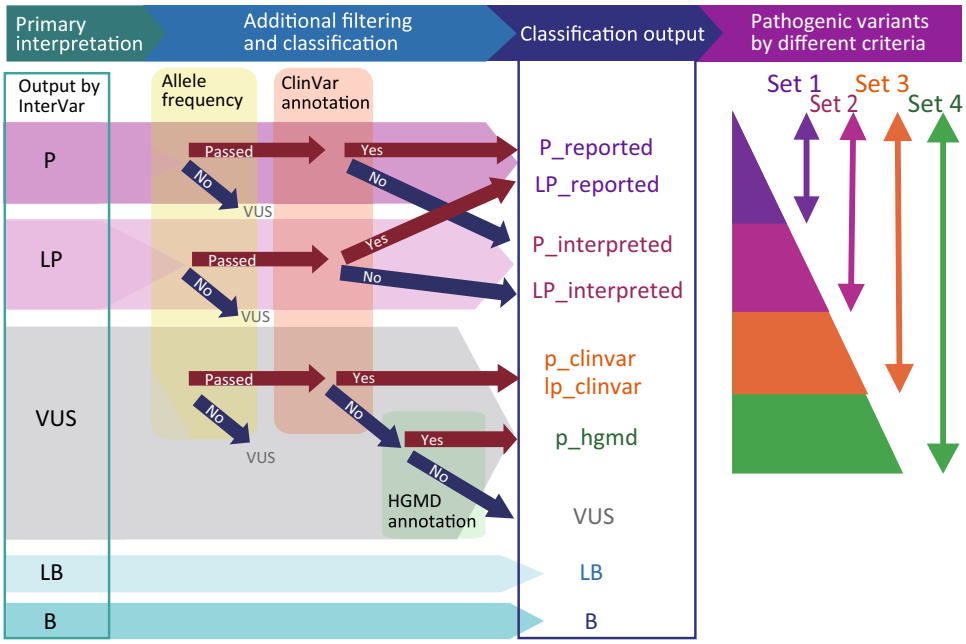

Fig. 1 Classification of variants from the 3.5KJPNv2 panel and selection of pathogenic variants for newborn screening genes. 3.5KJPNv2 (Tadaka et al. 2019) di-allelic genomic variants (after VQSR filtering) were annotated for primary variant interpretation using InterVar based on Annovar annotation. Variant interpretation was based on 18 out of 28 criteria in ACMG-AMP guidelines (2015) and the variants were divided into five classes; pathogenic (P), likely pathogenic (LP), variants of uncertain significance (VUS), likely benign (LB) and benign (B). Output of annotation and interpretation from InterVar for classes P, LP, and VUS was filtered by allele frequency (< 3%) and variants that passed were used for further classification. For P and LP groups, we checked whether a variant was already reported in ClinVar (clinical significance = P or LP). The variants were divided into either "reported" or "interpreted". The variants in the VUS group were used to seek additional pathogenic variants by matching to ClinVar or HGMD disease-causing mutation (DM) variants. We obtained four sets (1–4) of pathogenic variants from the conservative selection (Set 1) and the broad selection, including candidates of pathogenic variants suggested in ClinVar and HGMD (Sets 3 and 4)

showing the largest allele counts were *PCCA* ($n = 169$) and *SLC25A13* ($n = 109$), while these genes had fewer numbers in Set 2. Although *GCDH* had the largest number of pathogenic variants ($n = 13$) in Sets 2–4, the total frequency of pathogenic variants was not the largest because this gene had only low-frequency pathogenic variants. The number of pathogenic variants in *GCDH* did not increase in Set 3 and Set 4 compared to Set 2. All the 13 variants were detected by InterVar.

The number of individuals having at least 1 of the selected pathogenic variants were 153 (4.3%), 211 (5.9%), 319 (9.0%), and 605 (17.0%) for Sets 1–4, respectively. A small fraction of these individuals had more than one pathogenic allele. The number of individuals which had 2 pathogenic variants in different NBS genes were 2, 5, and 14 for Sets 1–3, respectively. In Set 4, there was 1 individual having 3 pathogenic variants in different genes and 40 individuals had 2 variants. There was one possible compound heterozygous individual for *SLC25A13*, who had an intronic variant, c.1311 + 1G > A, and a synonymous variant, p.Lys5Lys, reported in ClinVar and HGMD. However, after manual inspection and literature survey, p.Lys5Lys was not found in the original article, and we believe, this was not a real pathogenic variant. There were two 1-bp insertions at adjacent genomic positions in *BTD*, which were identified in one chromosome (*cis*) in only one individual. Therefore, these two insertions may be due to a single mutational event.

Using the allele frequencies for the selected variants, the population frequencies of risk alleles were estimated (Table 2. Table S3), and carrier frequencies were estimated based on the Hardy–Weinberg equilibrium. Then, focusing on diseases with only one responsible gene, the estimated carrier frequencies were compared with the observed incidence rates from the NBS data in Japan based on 1,949,987 individuals (Fig. 2).

PAH

Phenylalanine hydroxylase (PAH) is involved in the conversion of phenylalanine to tyrosine, and functional defects cause hyperphenylalaninemia (HPA) (MIM: 261,600) and phenylketonuria (PKU), which are recessively inherited congenital metabolic diseases (Knox and Messinger 1958; Woo 1989; Woo et al. 1983). The incidence rates of HPA (including PKU) in Japan were estimated to be

Table 2 Detection of pathogenic variants and estimated carrier frequencies of 32 genes for newborn screening diseases

Gene	MIM#	Pathogenic variants in 3.5KJPNv2 (3552 individuals)											
		Set 1*			Set 2*			Set 3*			Set 4*		
		No. Var	Allele Cnt	Carrier Frq.	No. Var	Allele Cnt	Carrier Frq.	No. Var	Allele Cnt	Carrier Frq.	No. Var	Allele Cnt	Carrier Frq.
PAH	612,349	5	16	0.00450	7	20	0.00562	8	27	0.00758	19	71	0.01979
ASL	608,310	0	0	0.00000	0	0	0.00000	1	2	0.00056	1	2	0.00056
ASS1	603,470	2	8	0.00225	2	8	0.00225	2	8	0.00225	5	11	0.00309
SLC25A13	603,859	8	70	0.01951	9	71	0.01979	10	72	0.02007	13	109	0.03022
BCKDHA	608,348	0	0	0.00000	2	2	0.00056	2	2	0.00056	3	3	0.00085
BCKDHB	248,611	0	0	0.00000	0	0	0.00000	2	3	0.00085	2	3	0.00085
DBT	248,610	0	0	0.00000	2	2	0.00056	2	2	0.00056	2	2	0.00056
DLD	238,331	1	1	0.00028	1	1	0.00028	1	1	0.00028	1	1	0.00028
CBS	613,381	0	0	0.00000	2	3	0.00085	5	14	0.00393	6	15	0.00422
MMADHC	611,935	0	0	0.00000	1	1	0.00028	1	1	0.00028	1	1	0.00028
MTHFR	607,093	1	1	0.00028	2	2	0.00056	3	22	0.00618	4	23	0.00646
MTR	156,570	0	0	0.00000	1	1	0.00028	1	1	0.00028	2	2	0.00056
MTRR	602,568	0	0	0.00000	0	0	0.00000	0	0	0.00000	1	1	0.00028
GCDH	608,801	3	4	0.00113	13	18	0.00506	13	18	0.00506	13	18	0.00506
BTD	609,019	2	2	0.00056	4	4	0.00113	7	9	0.00253	8	14	0.00394
HLCS	609,018	1	6	0.00169	1	6	0.00169	2	9	0.00253	2	9	0.00253
IVD	607,036	1	1	0.00028	2	2	0.00056	2	2	0.00056	2	2	0.00056
HMGCL	613,898	1	1	0.00028	2	2	0.00056	2	2	0.00056	2	2	0.00056
MCCC1	609,010	0	0	0.00000	1	1	0.00028	1	1	0.00028	2	14	0.00393
MCCC2	609,014	2	7	0.00197	3	8	0.00225	3	8	0.00225	4	9	0.00253
PCCA	232,000	0	0	0.00000	1	1	0.00028	1	1	0.00028	4	169	0.04645
PCCB	232,050	3	3	0.00085	3	3	0.00085	5	54	0.01509	7	58	0.01620
MCEE	608,419	0	0	0.00000	1	1	0.00028	1	1	0.00028	1	1	0.00028
MMAA	607,481	1	1	0.00028	2	4	0.00113	2	4	0.00113	3	5	0.00141
MMAB	607,568	0	0	0.00000	0	0	0.00000	0	0	0.00000	0	0	0.00000
MUT	609,058	2	7	0.00197	7	15	0.00422	7	15	0.00422	11	22	0.00618
ACADM	607,008	4	12	0.00337	5	13	0.00366	5	13	0.00366	8	16	0.00450
ACADVL	609,575	4	16	0.00450	8	21	0.00590	8	21	0.00590	13	40	0.01121
CPT1A	600,528	0	0	0.00000	2	2	0.00056	3	3	0.00085	3	3	0.00085
CPT2	600,650	0	0	0.00000	0	0	0.00000	4	14	0.00394	6	17	0.00478
HADHA	600,890	0	0	0.00000	2	2	0.00056	2	2	0.00056	2	2	0.00056
HADHB	143,450	0	0	0.00000	2	2	0.00056	2	2	0.00056	5	5	0.00141

*See "Materials and methods" and Fig. 1

approximately 1 in 70,000, which is lower than other countries. In a recent report on NBS in Japan, 37 cases were detected among 1,949,987 newborns, making the incidence rate 1 in 52,702 (confidence interval = 39,859–77,757) individuals and the expected carrier frequency is approximately 1/115 (0.87%) (Table 1). We identified pathogenic variants in the *PAH* gene in 3.5KJPNv2. Seven variants (p.Arg241Cys, p.Arg243Gln, p.Arg252Trp, p.Tyr356*, p.Thr380Met, c.536delA:p.Lys179fs, and c.344_347del:p.Lys115fs) were interpreted as pathogenic or likely pathogenic (Table 3), and they were included in Set 2. Five of the seven variants were reported missense or nonsense variants in ClinVar, the other two were frameshifting indels and predicted to likely be pathogenic due to their effects on translation. A pathogenic nonsense variant (p.Tyr356*), a reported variant in ClinVar, was detected in one heterozygote. Four missense variants, p.Arg252Trp, p.Arg241Cys, p.Arg243Gln, and p.Thr380Met, interpreted as likely pathogenic were detected in two, six, three, and four individuals, respectively, as heterozygous. In addition, two short deletions, c.344_347del:p.Lys115fs and c.551delA:p.Lys184fs, were detected in two individuals as heterozygous for each type. Furthermore, among VUS variants by initial interpretation, there was one ClinVar variant, p.Arg413Pro, detected in seven individuals as heterozygotes (AF = 0.00099). This variant was reported as frequently found among PKU/HPA patients in Japan (approximately 30%) (Okano et al. 1998, 2011). The amino acid residue p.Arg413 is located between the catalytic domain and tetramerization domain, and suspected to affect tetramer formation (Erlandsen and Stevens 1999). Recently, the missense variant, p.Arg413Pro, was reported to increase the levels of phenylalanine in heterozygous individuals according to metabolomics analysis of our cohort participants (Koshiba et al. 2016). Using HGMD, 11 additional HGMD-DM variants were found in the VUS group (Table 2, see Table S3 for an integrated form of Tables 1, 2). Among them, a missense variant, p.His170Gln, was most frequent in 3.5KJPNv2 (found in ten individuals as heterozygous). Another missense variant, p.Val379Ala (Okano et al. 2011), was found in four individuals as heterozygous. This variant was located close to the catalytic site and was reported to increase the levels of phenylalanine in heterozygous individuals according to the metabolomics analysis of our cohort participants (Koshiba et al. 2016). The carrier frequency was estimated for the four sets of pathogenic variants to be 0.00450, 0.00562, 0.00758, and 0.0198 for Sets 1–4, respectively (Table 2; Fig. 2). The carrier frequency estimated by Set 3 (0.00758) was the closest to the estimated carrier frequency of the disease incidence (0.0087) than the methods of including pathogenic variants (Fig. 2). On the other hand, using only reported pathogenic variants in ClinVar resulted in a lower estimate, and including all HGMD disease-causing mutation (DM) variants resulted in a much higher estimate. In addition to low-frequency variants with strong effects, there exists a common missense variant, p.Arg53His (rs118092776) in 3.5KJPNv2, at the frequency of 0.052 (detected in 348 as heterozygous and 10 individuals as homozygous). In our previous study, this variant increased phenylalanine levels by 19% in the blood of heterozygous individuals (Koshiba et al. 2016). Although this variant was discarded by filtering based on allele frequency, and was not interpreted as pathogenic, it was found in PKU/HPA patients as compound heterozygotes with other pathogenic variants (Lee et al. 2004). Allele frequencies of this variant were very low in other ethnic groups, such as European Americans in the Exome Sequencing Project (AF = 0.009).

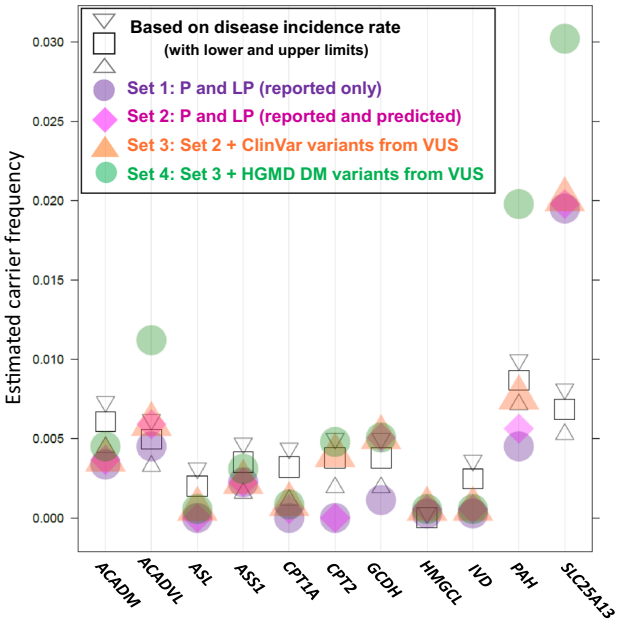

Fig. 2 Comparison of carrier frequencies estimated from genomic variants for newborn screening diseases. Estimated carrier frequencies are shown for 11 NBS diseases with single responsible gene (*ACADM* for MACD deficiency; *ACADVL* for VLCAD deficiency; *ASL* for argininosuccinic aciduria; *ASS1* for citrullinemia type 1; *CPT1A* for CPT1 deficiency; *CPT2* for CPT2 deficiency; *GCDH* for glutaric acidemia type 1; *HMGCL* for HMG-CoA lyase deficiency; IVD for isovaleric academia; *PAH* for PKU/PAH; *SLC25A13* for citrin deficiency). Carrier frequencies were estimated from the selected pathogenic variants by four methods, and are shown by the colored symbols (violet circle, magenta diamond, orange triangle, and green circle) for the four Sets (1–4) of pathogenic variants. Open squares show the estimated carrier frequency with lower and upper limits (open triangles) based on the observed incidence rates by a domestic NBS report (Yamaguchi 2012)

ASS1 and SLC25A13

Citrullinemia, an inherited disorder, causes an increased concentration of ammonia and citrulline in the blood. Two

Table 3 Possible pathogenic variants in the selected NBS genes

Genomic position (GRCh37/hg19)	dbSNP (build144)	Variation	Frequency in 3.5KJPNv2				Output by InterVar*	Mutation database		Classification**
			Genotype			Allele		ClinVar CS	HGMD	
			Ref/Ref	Het	Alt/Alt	Alt				
PAH (phenylketonuria/ hyperphenylalaninemia)										
chr12:103306579	rs118092776	c.G158A:p.Arg53His	3194	348	10	0.0518	VUS	VUS	DM	VUS
chr12:103288518	rs199475648	c.344_347del:p.Lys115fs	3550	2	0	0.0003	LP		DM	LP_interpreted
chr12:103249110	rs199475652	c.T510A:p.His170Gln	3542	10	0	0.0014	VUS		DM	p_hgmd
chr12:103249069		c.551delA:p.Lys184fs	3550	2	0	0.0003	LP		DM	LP_interpreted
chr12:103249009	rs62514927	c.A611G:p.Tyr204Cys	3551	1	0	0.0001	VUS	Other	DM	p_hgmd
chr12:103246714	rs76687508	c.C721T:p.Arg241Cys	3546	6	0	0.0008	LP	P	DM	LP_reported
chr12:103246707	rs62508588	c.G728A:p.Arg243Gln	3549	3	0	0.0004	LP	P	DM	LP_reported
chr12:103246681	rs5030847	c.C754T:p.Arg252Trp	3550	2	0	0.0003	LP	P | other	DM	LP_reported
chr12:103246641	rs62507335	c.G794A:p.Cys265Tyr	3551	1	0	0.0001	VUS		DM	p_hgmd
chr12:103246602	rs62507262	c.C833T:Thr278Ile	3550	2	0	0.0003	VUS		DM	p_hgmd
chr12:103240678	rs62514957	c.G964A:p.Ala322Thr	3551	1	0	0.0001	VUS		DM	p_hgmd
chr12:103238075	rs62510582	c.1065 + 39G > T	3551	1	0	0.0001	VUS		DM	p_hgmd
chr12:103237555		c.C1068A:p.Tyr356*	3551	1	0	0.0001	P	P	DM	P_reported
chr12:103237506	rs62508717	c.G1117A:p.Ala373Thr	3551	1	0	0.0001	VUS		DM	p_hgmd
chr12:103237500	rs184148104	c.C1123G:p.Gln375Glu	3549	3	0	0.0004	VUS		DM	p_hgmd
chr12:103237487	rs746203167	c.T1136C:p.Val379Ala	3548	4	0	0.0006	VUS		DM	p_hgmd
chr12:103237484	rs62642937	c.C1139T:p.Thr380Met	3548	4	0	0.0006	LP	P | other	DM	LP_reported
chr12:103234274	rs62644465	c.C1219T:p.Pro407Ser	3549	3	0	0.0004	VUS		DM	p_hgmd
chr12:103234255	rs79931499	c.G1238C:p.Arg413Pro	3545	7	0	0.0010	VUS	P	DM	p_clinvar
chr12:103232809	rs375319584	c.*144A > G	3535	17	0	0.0024	VUS		DM	p_hgmd
ASS1 (Citrullinemia type 1)										
chr9:133333870	rs575001023	c.G257A:p.Arg86His	3551	1	0	0.0001	VUS		DM	p_hgmd
chr9:133342110	rs751930594	c.421-2A > G	3547	5	0	0.0007	P	P	DM	P_reported
chr9:133346880		c.C575T:p.Ala192Val	3551	1	0	0.0001	VUS		DM	p_hgmd
chr9:133355236	rs763389916	c.773 + 49C > T	3551	1	0	0.0001	VUS		DM	p_hgmd
chr9:133364791	rs121908642	c.C910T:p.Arg304Trp	3549	3	0	0.0004	LP	P	DM	LP_reported
SLC25A13 (Citrullinemia type 2 and Citrin deficiency)										
chr7:95951267	rs541276426	c.T2C:p.Met1Thr	3521	31	0	0.0044	VUS	VUS	DM	p_hgmd
chr7:95906507		exon3:c.212 + 1G > A	3551	1	0	0.0001	P			P_interpreted
chr7:95838175		c.A443G:p.Tyr148Cys	3550	2	0	0.0003	VUS		DM	p_hgmd
chr7:95820501	rs80338719	c.C674A:p.Ser225*	3548	4	0	0.0006	P	P	DM	P_reported
chr7:95818951	rs531991442	c.G790A:p.Val264Ile	3548	4	0	0.0006	VUS		DM	p_hgmd
chr7:95818684	rs569808959	c.852_855del:p.Arg284fs	3536	16	0	0.0023	P	P	DM	P_reported
chr7:95813588	rs80338722	c.1177 + 1G > A	3522	30	0	0.0042	P	P	DM	P_reported

Table 3 (continued)

Genomic position (GRCh37/hg19)	dbSNP (build144)	Variation	Frequency in 3.5KJPNv2				Output by InterVar*	Mutation database		Classification**
			Genotype			Allele		ClinVar CS	HGMD	
			Ref/Ref	Het	Alt/Alt	Alt				
chr7:95799356	rs80338723	c.1311+1G>A	3540	12	0	0.0017	P	P	DM	P_reported
chr7:95751309	rs80338724	c.G1592A:p. ly531Asp	3549	3	0	0.0004	LP	P	DM	LP_reported
chr7:95751240	rs80338725	c.1660_1661insGAGATTACAGGT GGCTGCCCGGG:p.A554fs	3550	2	0	0.0003	P	P		P_reported
chr7:95751007	rs80338727	c.G1801T:p.Glu601*	3550	2	0	0.0003	P	P	DM	P_reported
MTHFR (Homocystinuria)										
chr1:11863038	rs138189536	c.C136T:p.Arg46Trp	3532	20	0	0.0028	VUS	P	DM	p_clinvar
chr1:11861449		c.C244T:p.Arg82Trp	3551	1	0	0.0001	LP	P	DM	LP_reported
chr1:11854919	rs759031330	c.C1033T:p.Arg345Cys	3551	1	0	0.0001	VUS		DM	p_hgmd
chr1:11854500		c.G1262A:p.Trp421*	3551	1	0	0.0001	P		DM	P_interpreted
BTD (multiple carboxylase synthetase deficiency)										
chr3:15677102		c.216_217insT:p.Asn72fs	3551	1	0	0.0001	LP			LP_interpreted
chr3:15677103		c.217_218insT:p.Pro73fs	3551	1	0	0.0001	LP			LP_interpreted
chr3:15677122	rs397514343	c.G236A:p.Arg79His	3551	1	0	0.0001	LP	P	DM	LP_reported
chr3:15685994	rs372844636	c.C631T:p.Arg211Cys	3551	1	0	0.0001	LP	P	DM	LP_reported
chr3:15686469	rs397514400	c.C1106T:p.Pro369Leu	3551	1	0	0.0001	VUS	P	DM	p_clinvar
chr3:15686647	rs35145938	c.C1284A:p.Tyr428*	3551	1	0	0.0001	VUS	P	DM	p_clinvar
chr3:15686802	rs558477960	c.G1439A:p.Gly480Glu	3547	5	0	0.0007	VUS		DM	p_hgmd
chr3:15686829	rs104893692	c.A1466C:p.Asn489Thr	3549	3	0	0.0004	VUS	P	DM	p_clinvar
HLCS (multiple carboxylase synthetase deficiency)										
chr21:38309035	rs119103227	c.T710C:p.Leu237Pro	3549	3	0	0.0004	VUS	P	DM	p_clinvar
chr21:38308963	rs771944310	c.782delG:p.Gly261fs	3546	6	0	0.0008	P	P	DM	P_reported
CPT1A (carnitine palmitoyltransferase I deficiency)										
chr11:68566685		exon6:c.693+1G>A	3551	1	0	0.0001	P			P_interpreted
chr11:68548130	rs80356779	c.C1436T:p.Pro479Leu	3551	1	0	0.0001	VUS	P	DM	p_clinvar
chr11:68527699		c.2136delT:p.Phe712fs	3551	1	0	0.0001	LP			LP_interpreted
CPT2 (carnitine palmitoyltransferase I deficiency)										
chr1:53668099	rs74315294	c.C338T:p.Ser113Leu	3546	6	0	0.0008	VUS	P	DM	p_clinvar
chr1:53675866	rs28936674	c.G520A:p.Glu174Lys	3550	2	0	0.0003	VUS	P	DM	p_clinvar
chr1:53676494	rs74315295	c.T1148A:p.Phe383Tyr	3547	5	0	0.0007	VUS	P	DM	p_clinvar
chr1:53676835		c.G1489A:p.Gly497Ser	3551	1	0	0.0001	VUS		DM	p_hgmd
chr1:53679103	rs751557097	c.G1813C:p.Val605Leu	3550	2	0	0.0003	VUS		DM	p_hgmd
chr1:53679181	rs74315293	c.C1891T:p.Arg631Cys	3551	1	0	0.0001	VUS	P	DM	p_clinvar

Table 3 (continued)

*Output of 5-tier classification of pathogenicity of variant, using InterVar based on 18 of 28 criteria of ACMG-AMP guideline (2015) of variant interpretation

**See "Materials and methods" and Fig. 1

types of citrullinemia, type 1 and 2 (CTLN1 and CTLN2), with different responsible genes, have been described. Clinically, CTLN1 and CTLN2 are associated with a wide spectrum of phenotypes, ranging from life-threatening neonatal hyperammonemia to late-onset forms (Diez-Fernandez et al. 2017). CTNL1 presents an acute neonatal form, which typically appears normal at birth. Shortly thereafter, hyperammonemia develops and the neonate becomes progressively lethargic, feeds poorly, often vomits, and may develop signs of increased intracranial pressure. The late-onset form may be milder than the acute neonatal form. However, women with the onset of severe symptoms during pregnancy or in the postpartum period have been reported (Gao et al. 2003). CTLN2 is characterized by recurring episodes of hyperammonemia and neurologic and psychotic symptoms that closely resemble those of hepatic encephalopathy or genetic urea cycle disorders. Misdiagnosed or mistreated CTLN2 patients usually have poor outcomes, often resulting in death due to hyperammonemic encephalopathy.

Argininosuccinate synthetase (ASS1) is involved in the urea cycle, and its gene product synthesizes argininosuccinic acid from citrulline and aspartic acid. Impairment of ASS1 function leads to citrullinemia type 1 (CTLN1), an autosomal recessive inborn error of the urea cycle (Beaudet et al. 1986). In Japan, the incidence of CTNL1 is estimated to be approximately 1 in 530,000 (Nagata et al. 1991), while it is estimated to be 1 in 250,000, in the United States, and is the third most frequent urea cycle disorder (Summar et al. 2013). There were five pathogenic variants in the *ASS1* gene in 3.5KJPNv2 (p.Arg86His, c.421-2A>G, p.Ala192Val, c.773 + 49C> T, and p.Arg304Trp) interpreted as pathogenic or likely pathogenic (Table 3). Two of the five variants, c.421-2A>G (AF = 0.00070) and p.Arg304Trp (AF = 0.00042), were annotated to be pathogenic variants in ClinVar. Variant c.421-2A > G was observed in over 50% of Japanese CTLN1 patients (Kobayashi et al. 1995), a high frequency in Japan and also in Korea, but not in Caucasian populations (Gao et al. 2003). A missense variant, p.Arg304Trp, found in patients from the Indian subcontinent, Turkey, Germany, and Japan, was present in 40% of Japanese patients (Diez-Fernandez et al. 2017). The other three variants, p.Arg86His (AF = 0.00014), p.Ala192Val (AF = 0.00014), and c.773 + 49C > T (AF = 0.00014) were annotated to be disease-causing mutations in the HGMD. The estimated carrier frequency of *ASS1* was estimated to be 0.00225 (Set 1–3) with two variants (c.421-2A > G and p.Arg304Trp) that were reported in ClinVar. The estimated carrier frequency was 0.00309 in Set 4, including additional three HGMD variants; p.Arg86His, p.Ala192Val, and c.773 + 49C > T (Table 2, Table S3), which was closest to that based on the incidence rate (0.0035).

The *SLC25A13* gene encodes a protein called citrin, a member of the mitochondrial carrier family. This protein

is localized in the mitochondrial inner membrane, active chiefly in the liver, kidneys, and heart. Within the cells of these organs, citrin is involved in transporting molecules into and out of energy-producing structures. Citrin deficiency can manifest in newborns or infants as neonatal intrahepatic cholestasis, and in adults as recurrent hyperammonemia with neuropsychiatric symptoms in citrullinemia type II (CTLN2). The incidence of CTNL2 is estimated to be 1 in 100,000 in Japan (Kobayashi et al. 1993), whereas estimated frequency of homozygotes from the carrier frequency (1/69 in Japan) in a previous study was 1/19,000 (Tabata et al. 2008).

We examined pathogenic variants in *SLC25A13* in 3.5KJPNv2, and there were 13 variants (p.Met1Thr, p.Lys5Lys, c.212 + 1G > A, p.Tyr148Cys, c.615 + 1G > C, p.Ser225*, p.Val264Ile, p.Arg284fs, c.1177 + 1G > A, c.1311 + 1G > A, p.Gly531Asp, p.Ala554fs, p.Glu601*) interpreted as pathogenic, likely pathogenic, or possible candidates of pathogenic variants (Table 3). Nine of these variants, p.Glu601* (AF = 0.00028), frameshifting insertion p.A554fs (AF = 0.00028), p.Gly531Asp (AF = 0.00042), c.1311 + 1G > A (AF = 0.00169), c.1177 + 1G > A (AF = 0.00422), frameshifting deletion p.Arg284fs (AF = 0.00225), p.Ser225* (AF = 0.00056), c.615 + 1G > C (AF = 0.00014), and synonymous variant p.Lys5Lys (AF = 0.00014) were annotated to be pathogenic variants in ClinVar. A variant affecting splicing, c.1177 + 1G > A, was detected in 12 individuals in 3.5KJPNv2, causing a G > A substitution at the 5′ end of intron 13, resulting in abnormal splicing and deletion of 27 aa (codons 411–437) between predicted transmembrane 2 and 3 (Kobayashi et al. 1999). Another variant affecting splicing, c.1177 + 1G > A, was detected in 30 individuals in 3.5KJPNv2, causing a G > A substitution at the 5′ end of intron 11, resulting in abnormal splicing and deletion of exon 11 in mRNA (Kobayashi et al. 1999). This variant was also found in one individual as heterozygous among school children, who showed specific food preferences in our questionnaire for detecting citrin deficiency (Miyashita et al. 2017). Further investigation in epidemiology and omics concerning these careers would clarify how this variant affects lifestyle and physiologic function. Five of the 13 variants, p.Arg553Gln (AF = 0.00310), p.Val264Ile (AF = 0.00056), p.Tyr148Cys (AF = 0.00028), splicing variant c.212 + 1G > A (AF = 0.00014), and p.Met1Thr (AF = 0.00436) were annotated to be disease-causing mutations in the HGMD. p.Met1Thr was detected in 31 individuals in 3.5KJPNv2. This variant was found to have a carrier frequency of 1 in 18 in a Thai population (Wongkittichote et al. 2013a), and functional studies in yeast found the variant expresses a truncated nonfunctional protein (Wongkittichote et al. 2013b). Among 3552 individuals, we found 1 individual who had 2 pathogenic variants (c.1311 + 1G > A and p.Lys5Lys) as a compound heterozygote. However, we did not find any description for p.Lys5Lys in the original article (Tabata et al. 2008), and this variant might have been mistakenly registered in the database. The estimated carrier frequency of *SLC25A13* was estimated to be 0.01951, 0.01979, 0.02007, and 0.03022 for Sets 1–4 (Table 2, Table S3). All of the four estimations were much higher than the estimated carrier frequency based on the observed incidence rate (0.00685) (Fig. 2).

HLCS and BTD

Multiple carboxylase deficiency (MCD) is caused by either a deficiency of holocarboxylase synthetase (HCS) (MIM: 253270) or biotinidase (MIM: 253260). In Japan, the incidence rates of both diseases were estimated to be very low. In a domestic NBS study, the incidence rate of MCD was estimated to be 1 in 650,000, based on the observation of 3 cases among 1,949,987 newborns (Table 1). An estimation of the incidence rate of HCS deficiency was lower than 1 in 1,000,000, based on the fact that there were a total of 28 cases of HCS deficiency over 30 years (1982–2011). Biotinidase deficiency is even rarer among Japanese people, having only several case reports in Japan, while the incidence rate was estimated to be 1 in 60,000 in Caucasians. In Japan, most MCD cases are due to HCS deficiency.

In *HLCS*, missense variant p.Leu237Pro and a 1-bp deletion, c.782delG, were found in three and six heterozygotes, respectively, in 3.5KJPNv2 (Table 3). These two variants, p.Leu237Pro and c.782delG (described as 780delG in the original article due to a repeat of G nucleotide), are frequently found variants in Japanese patients (Suzuki et al. 1994, 2005). They are used in the diagnosis of this disease. Based on allele frequencies of these 2 variants in 3.5KJPNv2, frequency of risk alleles in the population was estimated to be 0.00127, and the proportion of individuals who have risk alleles on 2 chromosomes (homozygous or compound heterozygous) was estimated to be approximately 1 in 620,000.

In *BTD*, two missense variants, p.Arg79His and p.Arg211Cys, were identified in one heterozygote, were reported in ClinVar, and were classified as likely pathogenic (Table 3). In addition, two frameshifting 1-bp insertions, c.216_217insT and c.217_218insT, located at adjacent nucleotides, were found in one chromosome in a single individual, and classified as pathogenic due to their frameshifting effects. These variants in *cis* could be treated as one variant, and we expect that its translational effect is frameshifting. Furthermore, additional candidates of pathogenic variants found from ClinVar and HGMD were p.Pro369Leu (singleton), p.Tyr428* (singleton), p.Gly480Glu (five heterozygotes), and p.Asn489Thr (three heterozygotes).

The number of carriers of pathogenic variants in *BTD* was smaller than for *HLCS* for Set 2 (interpreted as pathogenic or likely pathogenic); however, this number became higher for *BTD* in Sets 3–4 when additional HGMD-DM and ClinVar variants were included. Further evaluation will be needed to clarify the clinical significance for reported pathogenic variants in HGMD and ClinVar which do not satisfy criteria for pathogenic or likely pathogenic in variant interpretation.

CPT1A and CPT2

When long-chain fatty acids are oxidized in the mitochondrial matrix to be utilized as fuel, they are conjugated to carnitine to be translocated from the cytoplasm into the mitochondrial matrix. The acyl group is transferred from coenzyme A (CoA) to carnitine to form acyl carnitine. This reaction is catalyzed by carnitine palmitoyltransferase 1 (CPT1) in the outer mitochondrial membrane. Thereafter, acyl carnitine translocase carries acyl carnitine across the inner mitochondrial membrane, and the acyl group is transferred to CoA on the matrix side of the membrane. This reaction is catalyzed by carnitine palmitoyltransferase 2 (CPT2). As these carnitine palmitoyltransferases play critical roles in long-chain fatty acid oxidation, deficiency of CPTs impair long-chain fatty acid oxidation and causes metabolic abnormalities (Bonnefont et al. 2004).

Deficiency of CPT1A, a liver isoform of CPT1, influences fatty-acid metabolic abnormalities. In 3.5KJPNv2, missense variant p.Pro479Leu was found in only one individual as heterozygous (AF = 0.00014) (Table 3). It has been reported that this variant causes mild decreases in the catalytic activity of CPT and resistance to inhibition by malonyl-CoA derived from glucose products (Brown et al. 2001). In fact, this variant is rarely found in East Asian and other human populations (Song et al. 2015). However, missense variant p.Pro479Leu exists at very high frequencies in indigenous populations in the northern territories of Canada (e.g., AF = 0.77 in Nuvavut) (Collins et al. 2010) where the traditional diet contains high fat, moderate protein, and low carbohydrates. Although the association of this variant and sudden unexpected death in infancy (SUDI) was suspected, a recent review (Fohner et al. 2017) concluded that the data were not sufficient and further research is needed to clarify the causal association. In addition to this reported variant, 3.5KJPNv2 had two variants, exon6:c.693 + 1G > A (one heterozygote) and exon17:c.2136delT:p.Phe712fs (one heterozygote), which were predicted to be pathogenic or likely pathogenic (Table 3). The sum of the allele frequencies for these three variants was 0.00042, and the carrier frequency was estimated to be 0.00085 (Table 2, Table S3). This was lower than the expected carrier frequency observed in the NBS study in Japan (0.0032) (Fig. 2).

CPT2 deficiency presents three clinical phenotypes, including adult form, infantile form, and neonatal form (Bonnefont et al. 2004). In an adult form patient, the clinical symptoms are mild. In contrast, the symptoms are severe in the infantile and neonatal forms. Although there was no variant in 3.5KJPNv2 described as pathogenic that satisfied the criteria of variant guidelines, we observed the following variants reported in ClinVar or HGMD: p.Ser113Leu (AF = 0.0017) (Taroni et al. 1992), p.Glu174Lys (AF = 0.0017)(Yamamoto et al. 1996), p.Phe383Tyr (AF = 0.0014) (Yamamoto et al. 1996), p.Gly497Ser (AF = 0.00014), p.Val605Leu (AF = 0.00028), and p.Arg631Cys (AF = 0.00028) (Taroni et al. 1992). Among them, two missense variants, p.Ser113Leu and p.Glu174Lys, were previously reported to be found in the adult form of CPT2 deficiency and p.Arg631Cys was found in a severe infantile form of CPT2 deficiency as homozygous. The estimate carrier frequency of *CPT2* was estimated to be 0.00394 (set 3) and this was closest to that by the incidence rates than other sets (1, 2, and 4) (Table 2, Table S3, and Fig. 2). In addition, it is reported that CPT2 deficiency is related to SUDI as p.Phe383Tyr was found in a case of SUDI (Yamamoto et al. 2015). In addition to low-frequency variants with possible pathogenic effects, we found common missense variants of *CPT2* p.Phe352Cys (AF = 0.18) and p.Val368Ile (AF = 0.69) in 3.5KJPNv2 (Table S4). The combination of heterozygous p.Phe352Cys and homozygous p.Val368Ile was found in an additional case of SUDI (Takahashi et al. 2016). Therefore, sequencing of *CTP2* may be useful for screening of CPT2 deficiency to decrease the risk of SUDI.

MTHFR

Methylenetetrahydrofolate reductase (MTHFR) is involved in folate-mediated one-carbon metabolism. Deficiency of MTHFR increases homocysteine levels and causes hyperhomocysteinemia with homocystinuria (MIM: 236,250), mild hyperhomocysteinemia, and is also involved in other diseases, such as vascular disease, neural tube defects, cancer, and developmental delay (Frosst et al. 1995; Lange et al. 2010; Larsson et al. 2006). Additional causative genes of hyperhomocysteinemia with homocystinuria are *CBS, MTR, MTRR*, and *MMADHC*. Human MTHFR protein acts as a homodimer, consisting of an N-terminal catalytic domain (residue numbers 1-356), a linker region (357–362), and a C-terminal regulatory region (363–656)(Froese et al. 2016; Guenther et al. 1999).

In 3.5KJPNv2, there are four possible pathogenic variants (Table 3). First, missense variant p.Arg82Trp, which was classified as likely pathogenic and has been reported in ClinVar and HGMD, was found in one individual as heterozygous (AF = 0.00014). On the other hand, nonsense variant

p. Trp421* in the regulatory domain, was unreported, interpreted as pathogenic and was found in one heterozygous individual. In addition to the above two variants satisfying the variant guidelines, there were two missense variants, p.Arg46Trp and p.Arg345Cys, that did not satisfy criteria for classes P or LP. However, they were reported in ClinVar or HGMD. Missense variants p.Arg46Trp, p.Arg82Trp, and p.Arg345Cys were originally found in families with severe MTHFR deficiency (Burda et al. 2015; Froese et al. 2016), and are located in the N-terminal catalytic domain. The estimated carrier frequency of *MTHFR* was estimated to be 0.00618 (Set3) (Table 2, Table S3), although this gene is one of five causative genes for homocystinuria. In addition to rare variants affecting Mendelian diseases, there are reports on other variants that mildly or severely affect MTHFR activities and/or metabolic profiles in the blood (Frosst et al. 1995; Koshiba et al. 2016; Weisberg et al. 2001). For example, missense variant p.Ala222Val (rs1801133, c.665C > T, also described as 677C > T in previous studies) is found in 1674 individuals as heterozygous and 522 individuals as homozygous (AF = 0.383) in 3.5KJPNv2 (Table S4). This variant exists at high frequencies in other populations, and is associated with various types of diseases, such as neural tube defects, age-related hearing loss, anencephaly, heart diseases, stroke, and hypertension. Individuals having this variant as homozygous have decreased enzyme activity (approximately 30% of the normal enzyme activity) (Frosst et al. 1995). In fact, our previous study showed association of this variant with the concentration of plasma formate by analyzing blood metabolites in the cohort samples (Koshiba et al. 2016).

Discussion

Understanding the relationship between genomic variants in a population and disease incidence rates is very important. Herein, we have conducted the first comprehensive study on detecting genomic pathogenic variants and estimation of carrier frequencies for 17 congenital metabolic disorders included in Japanese NBS using data based on whole-genome sequencing of individual genomes. The estimated carrier frequency was compared to the reported incidence rates from a domestic NBS report (Yamaguchi 2012). Since an automatic and efficient method for variant interpretation has not yet been established, we looked at multiple criteria for selecting pathogenic variants and detected reported and predicted pathogenic variants. An estimated carrier frequency, based on genomic data with recent variant interpretation guidelines, for the *PAH* gene, which causes PKU/HPA, gave a closer estimate to the observed incidence than the other methods (Fig. 2). However, results varied greatly among diseases with a single responsible gene. For example, in citrin deficiency, the estimated carrier frequencies for *SLC25A13* by all four methods of detecting pathogenic variants were higher than the observed incidence rate (Fig. 2). On the other hand, for *CPT1A*, the cause of CPT1 deficiency, the selected pathogenic variants were not enough to explain the observed incidence rate (Fig. 2).

Departures in the estimated carrier frequency by the selected pathogenic variants compared with that based on the observed incidence rate may be due to several factors (Table 4). Overestimation of carrier frequency by genomic data may be due to (1) assuming complete penetrance of the selected pathogenic variants, or (2) selected pathogenic variants include false positives or variants with mild effects. On the other hand, underestimation of carrier frequency may occur when other responsible gene(s) exist (Wada et al. 2018) or other causative variants are present in the focused gene (Table 4). First, the number of pathogenic variants may increase by including additional lines of evidence in the variant interpretation. Second, common variants, as well as rare variants, may also contribute to pathogenesis in heterozygous with another pathogenic variant. In addition, there may exist causative larger insertions or deletions not included in the WGS panel or not detected in variant call in whole-genome sequencing. Another reason for underestimation may be that the observed disease incidence rate was slightly higher than expected by the Hardy–Weinberg equilibrium, possibly due to genomic imbalance, such as uniparental disomy (UPD) (Table 4). It was reported that the incidence of UPD of any chromosome is estimated to be approximately 1 in 3500 live births (Robinson 2000) and that recessive conditions for enzyme deficiency were manifested by UPD. Further study is needed to clarify how frequently UPD affects congenital metabolic disorders.

Another aim of this study was to explore better strategies of identifying pathogenic variants. Although the results of comparing estimated carrier frequencies were very different among the 11 diseases and there are unsolved issues, let us discuss this to some extent with the 10 diseases, excluding *SLC25A13* showing much higher carrier frequencies with genomic data compared with that by the incidence rate. In *PAH* and *CPT2*, the estimated carrier frequencies based on Set 3 were closer to those by the incidence rates than the other sets (Fig. 2). With the selected variants in Set 3, the difference in estimated carrier frequency between the genomic data and the incidence rate was not larger than other sets in *ASL, CPT1A, GCDH, HMGCL*, and *IVD*. If we assume that Set 3 is better than the others in selecting pathogenic variants accounting for the reported incidence rate, exceptions are *ACADM, ACADVL* and *ASS1* (in which Set 1 or 4 gave the closest estimations of carrier frequency to that by the incidence rate). However, we should be careful to the possibility that one factor (Table 4) elevated the estimated carrier frequency, and another factor (Table 4) lowered the

Table 4 Possible reasons for the departure of the estimated carrier frequency by genomic variants and the observed incidence rates

Reason	Direction of departure	Example
Incomplete penetrance, variants with mild effects, and false positives	Overestimation	Assuming complete penetrance of pathogenic variants, while not all the selected pathogenic variants have 100% penetrance
		Selected pathogenic variants include false positives or variants with mild effects
Existence of other variants or loci that affect pathogenesis	Underestimation	The VUS group contains additional pathogenic variants, which do not have enough evidence yet or their evidences are not yet utilized for automatic analysis
		Common variants (> 3%) also contribute to pathogenesis, possibly in heterozygotes with another pathogenic variant
		Some causative variants (e.g., indels or other types of variants) were not included in the WGS panel, possibly due to a no call in NGS or exclusion in the quality-control process
Disease incidence rate is higher than expected by H–W principle	Underestimation	Frequency of homozygote risk alleles in somatic cells might be higher than expected by Hardy–Weinberg principle due to some mechanism, such as genomic imbalance (e.g., uniparental disomy)
Diagnosis	Both	Possible false positives or true negatives

estimation in a single disease, and that Set 3 gave closer estimations in appearance in several diseases. Furthermore, increase of reported pathogenic variants in ClinVar in the near future will elevate the estimated carrier frequency in Set 3. Therefore, we cannot draw clear conclusion at present, however, an appropriate strategy of variant interpretation based on multiple lines of evidence, which are currently increasing, would enable us to identify pathogenic variants.

Currently there are unsolved reasons for departures in carrier frequency, further examination of the relationship between variant frequencies of risk alleles and incidence rates of diseases with updated data and revised analysis would clarify the relationship between genomic variants and disease incidence rates. Variant annotation and interpretation presented in this study can be improved by including experimental or medical evidence and curation by medical experts. Furthermore, there could be false positives and false negatives in diagnoses (Table 4) (Holtzman et al. 1986; McCabe and McCabe 2002; Zaffanello et al. 2003), and thereby affecting the observed incidence rates. If there is any difference in the estimated carrier frequency based on genomes and the reported incidence for a disease, the direction and degree of the differences between the genomic data and the incidence should be utilized to suspect possible reasons for the departures (e.g., additional responsible genomic loci or variant, and environmental factors).

It was previously reported that the Japanese population has a lower incidence rate of congenital metabolic disorders compared with Caucasians, and PKU/PHA is one example. It was also reported that the variant types in patients with PKU/PHA were different among East Asians and Europeans. We found the previous frequently reported pathogenic variants in the *PAH* gene, p.Arg413Pro and p.His170Gln, in Japanese patients. On the contrary, missense variant p.Arg408Trp was a frequently found variant among Caucasian patients (Eisensmith et al. 1995; Konecki and Lichter-Konecki 1991), existing at the frequency of 0.0017 in 4300 European American individuals in Exome Sequencing Project (ESP) data (Fu et al. 2013) (allele count = 15 out of 8600), while it was not detected in 3.5KJPNv2. A previous report suggested that there was a founder effect in transmitting responsible variants for PKU/HPA to East Asia from other areas of the world (Wang et al. 1991) because variant types detected in the patients were limited to East Asia. It remains unclear whether limited types of variants explain lower incidence rates in Japan for other congenital metabolic disorders, and further study is required to examine the relationship between inter-ethnic incidence rate differences and inter-ethnic genomic differences [e.g., genetic basis of a very low incidence rate of cystic fibrosis in Japan (Yamashiro et al. 1997)]. Our results also showed genes with relatively higher carrier frequencies, such as *SLC25A13* for citrin deficiency and *PCCA* for propionic academia.

The first report estimating the carrier frequency of a recessive disorder for the Japanese population at the molecular level (protein polymorphisms) was seen in early 1960s for acatalasemia (MIM: 614,097) (Hamilton et al. 1961; Takahara et al. 1960), for which the responsible gene is *CAT* (MIM: 115,500). They showed that the carrier frequency varied among local populations in Japan at that time. Regional differences in the incidence rates for recessive disorders may still exist in the current Japanese population,

as seen in a recent report on higher PKU rates in Nagasaki (Dateki et al. 2016), located at the west end of Japan. In fact, higher rates of PKU are found in China than Japan, and it would be interesting to infer how genomic variants for PKU have been transmitted to Japan. As we expand the whole reference panel by including individuals from other parts of Japan, the estimated allele frequency becomes closer to the average frequency in the Japanese population. However, possible regional differences in frequency of pathogenic variants may need clarification by further studies using genomic data from larger numbers of individuals in each geographic area.

Our results showed that the proportion of individuals having at least 1 risk allele for the 17 diseases for NBS was 5.9% as a conservative estimate and could be more than 10% if all HGMD-DM variants are included. Identifying variant carriers for congenital metabolic disorders would be important for not only for estimating disease prevalence for public health, but also individual precision medicine (Punj et al. 2018). Reports on the effects of heterozygous states at the omics level (e.g., metabolomics) or based on population genetic analysis are increasing (Cassa et al. 2017; Koshiba et al. 2018; Long et al. 2017). Taking an example of the *MTHFR* gene, there existed a common variant (p.Ala222Val) in 3.5KJPNv2, which was previously shown to be associated with plasma formate concentration (Koshiba et al. 2016), as well as low frequency reported pathogenic variants. Further studies on genomic variations using omics data and cohort data would clarify genomic variants affecting the biochemical and physiological status of the human body.

It has been difficult to directly estimate disease prevalence rates in a general population for many genetic diseases. Although the number of patients is usually available from the hospitals, it was not easy to estimate the denominator, how many general residents were examined. In rare diseases, such as MCD, it is difficult to appropriately estimate the frequency of patients in a country without a nationwide system for epidemiological statistics. It would be quite useful to estimate the frequency of potential patients by analyzing genomic variations for congenital metabolic disorders. This study on estimating carrier frequencies from WGS data with variant annotation and interpretation would (1) improve the frequency estimations of potential patients with genetic diseases and (2) expedite further understanding of the relationships among genomic variants, individual biochemical status, and diseases. Although this study has an impact on genetic epidemiology and public health in Japan, it would be highly useful for precision medicine for infants. A recent study showed that precision medicine based on rapid WGS (Farnaes et al. 2018) decreased cost and the period of hospitalization. Such a strategy is feasible in many facilities in the near future. It would be quite important to prevent sudden death or serious conditions of infants through genome-based precision medicine. This type of study will provide information about how frequently each genetic disease may be observed in newborns in Japan, through estimated probabilities of genetic defects in each disease gene in the Japanese population and a better bioinformatics strategy assisting efficient variant interpretation.

Acknowledgements We are indebted to all volunteers who participated in this Tohoku Medical Megabank project. We would like to acknowledge all members associated with this project; the member list is available at the following web site: http://www.megabank.tohoku.ac.jp/english/a171201/. This work was supported in part by the Tohoku Medical Megabank Project from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Reconstruction Agency; the Japan Agency for Medical Research and Development (AMED; Grant Numbers JP18km0105001 and JP18km0105002) for Tohoku University. All computational resources were provided by the Tohoku University Tohoku Medical Megabank Organization (ToMMo) supercomputer system (http://sc.megabank.tohoku.ac.jp/en), which is supported by Facilitation of R&D Platform for AMED Genome Medicine Support conducted by AMED (Grant Number JP18km0405001). This research is also supported by the Center of Innovation Program from Japan Science and Technology Agency, JST.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, Murray ML, Tokita MJ, Gallego CJ, Kim DS, Bennett JT, Crosslin DR, Ranchalis J, Jones KL, Rosenthal EA, Jarvik ER, Itsara A, Turner EH, Herman DS, Schleit J, Burt A, Jamal SM, Abrudan JL, Johnson AD, Conlin LK, Dulik MC, Santani A, Metterville DR, Kelly M, Foreman AK, Lee K, Taylor KD, Guo X, Crooks K, Kiedrowski LA, Raffel LJ, Gordon O, Machini K, Desnick RJ, Biesecker LG, Lubitz SA, Mulchandani S, Cooper GM, Joffe S, Richards CS, Yang Y, Rotter JI, Rich SS, O'Donnell CJ, Berg JS, Spinner NB, Evans JP, Fullerton SM, Leppig KA, Bennett RL, Bird T, Sybert VP, Grady WM, Tabor HK, Kim JH, Bamshad MJ, Wilfond B, Motulsky AG, Scott CR, Pritchard CC, Walsh TD, Burke W, Raskind WH, Byers P, Hisama FM, Rehm H, Nickerson DA, Jarvik GP (2015) Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res 25:305–315. https://doi.org/10.1101/gr.183483.114

Beaudet AL, O'Brien WE, Bock HG, Freytag SO, Su TS (1986) The human argininosuccinate synthetase locus and citrullinemia. Adv Hum Genet 15:161–196, 291–2

Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, Sheth V, Woodward JE, Peckham HE, Schroth GP, Kim RW, Kingsmore SF (2011) Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci Transl Med 3:65ra4. https://doi.org/10.1126/scitranslmed.3001756

Bonnefont JP, Djouadi F, Prip-Buus C, Gobin S, Munnich A, Bastin J (2004) Carnitine palmitoyltransferases 1 and 2: biochemical,

molecular and medical aspects. Mol Aspects Med 25:495–520. https://doi.org/10.1016/j.mam.2004.06.004

Brown NF, Mullur RS, Subramanian I, Esser V, Bennett MJ, Saudubray JM, Feigenbaum AS, Kobari JA, Macleod PM, McGarry JD, Cohen JC (2001) Molecular characterization of L-CPT I deficiency in six patients: insights into function of the native enzyme. J Lipid Res 42:1134–1142

Burda P, Schafer A, Suormala T, Rummel T, Burer C, Heuberger D, Frapolli M, Giunta C, Sokolova J, Vlaskova H, Kozich V, Koch HG, Fowler B, Froese DS, Baumgartner MR (2015) Insights into severe 5,10-methylenetetrahydrofolate reductase deficiency: molecular genetic and enzymatic characterization of 76 patients. Hum Mutat 36:611–621. https://doi.org/10.1002/humu.22779

Cassa CA, Weghorn D, Balick DJ, Jordan DM, Nusinow D, Samocha KE, O'Donnell-Luria A, MacArthur DG, Daly MJ, Beier DR, Sunyaev SR (2017) Estimating the selective effects of heterozygous protein-truncating variants from human exome data. Nat Genet 49:806–810. https://doi.org/10.1038/ng.3831

Chong JX, Ouwenga R, Anderson RL, Waggoner DJ, Ober C (2012) A population-based study of autosomal-recessive disease-causing mutations in a founder population. Am J Hum Genet 91:608–620. https://doi.org/10.1016/j.ajhg.2012.08.007

Collins SA, Sinclair G, McIntosh S, Bamforth F, Thompson R, Sobol I, Osborne G, Corriveau A, Santos M, Hanley B, Greenberg CR, Vallance H, Arbour L (2010) Carnitine palmitoyltransferase 1A (CPT1A) P479L prevalence in live newborns in Yukon, Northwest Territories, and Nunavut. Mol Genet Metab 101:200–204. https://doi.org/10.1016/j.ymgme.2010.07.013

Cooper DN, Ball EV, Krawczak M (1998) The human gene mutation database. Nucleic Acids Res 26:285–287

Dateki S, Watanabe S, Nakatomi A, Kinoshita E, Matsumoto T, Yoshiura K, Moriuchi H (2016) Genetic background of hyperphenylalaninemia in Nagasaki, Japan. Pediatr Int 58:431–433. https://doi.org/10.1111/ped.12924

Diez-Fernandez C, Rufenacht V, Haberle J (2017) Mutations in the Human Argininosuccinate Synthetase (ASS1) Gene, Impact on Patients, Common Changes, and Structural Considerations. Hum Mutat 38:471–484. https://doi.org/10.1002/humu.23184

Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, Kim JH, Rosenthal EA, Kim DS, National Heart L, Blood Institute Grand Opportunity Exome Sequencing P, Tabor HK, Bamshad MJ, Motulsky AG, Scott CR, Pritchard CC, Walsh T, Burke W, Raskind WH, Byers P, Hisama FM, Nickerson DA, Jarvik GP (2013) Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet 93:631–640. https://doi.org/10.1016/j.ajhg.2013.08.006

Eisensmith RC, Goltsov AA, O'Neill C, Tyfield LA, Schwartz EI, Kuzmin AI, Baranovskaya SS, Tsukerman GL, Treacy E, Scriver CR et al (1995) Recurrence of the R408W mutation in the phenylalanine hydroxylase locus in Europeans. Am J Hum Genet 56:278–286

Erlandsen H, Stevens RC (1999) The structural basis of phenylketonuria. Mol Genet Metab 68:103–125. https://doi.org/10.1006/mgme.1999.2922

Farnaes L, Hildreth A, Sweeney NM, Clark MM, Chowdhury S, Nahas S, Cakici JA, Benson W, Kaplan RH, Kronick R, Bainbridge MN, Friedman J, Gold JJ, Ding Y, Veeraraghavan N, Dimmock D, Kingsmore SF (2018) Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. NPJ Genom Med 3:10. https://doi.org/10.1038/s41525-018-0049-4

Fohner AE, Garrison NA, Austin MA, Burke W (2017) Carnitine palmitoyltransferase 1A P479L and infant death: policy implications of emerging data. Genet Med 19:851–857. https://doi.org/10.1038/gim.2016.202

Froese DS, Huemer M, Suormala T, Burda P, Coelho D, Gueant JL, Landolt MA, Kozich V, Fowler B, Baumgartner MR (2016) Mutation Update and Review of Severe Methylenetetrahydrofolate Reductase Deficiency. Hum Mutat 37:427–438. https://doi.org/10.1002/humu.22970

Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, Boers GJ, den Heijer M, Kluijtmans LA, van den Heuvel LP et al (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. Nat Genet 10:111–113. https://doi.org/10.1038/ng0595-111

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, Nickerson DA, Bamshad MJ, Project NES, Akey JM (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493:216–220. https://doi.org/10.1038/nature11690

Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T, Boroevich KA, Nagasaki M, Yamaguchi R, Shibuya T, Kubo M, Miyano S, Nakamura Y, Tsunoda T (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. Nat Genet 42:931–936. https://doi.org/10.1038/ng.691

Gao HZ, Kobayashi K, Tabata A, Tsuge H, Iijima M, Yasuda T, Kalkanoglu HS, Dursun A, Tokatli A, Coskun T, Trefz FK, Skladal D, Mandel H, Seidel J, Kodama S, Shirane S, Ichida T, Makino S, Yoshino M, Kang JH, Mizuguchi M, Barshop BA, Fuchinoue S, Seneca S, Zeesman S, Knerr I, Rodes M, Wasant P, Yoshida I, De Meirleir L, Abdul Jalil M, Begum L, Horiuchi M, Katunuma N, Nakagawa S, Saheki T (2003) Identification of 16 novel mutations in the argininosuccinate synthetase gene and genotype-phenotype correlation in 38 classical citrullinemia patients. Hum Mutat 22:24–34. https://doi.org/10.1002/humu.10230

Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A, Besenbacher S, Magnusson G, Halldorsson BV, Hjartarson E, Sigurdsson GT, Stacey SN, Frigge ML, Holm H, Saemundsdottir J, Helgadottir HT, Johannsdottir H, Sigfusson G, Thorgeirsson G, Sverrisson JT, Gretarsdottir S, Walters GB, Rafnar T, Thjodleifsson B, Bjornsson ES, Olafsson S, Thorarinsdottir H, Steingrimsdottir T, Gudmundsdottir TS, Theodors A, Jonasson JG, Sigurdsson A, Bjornsdottir G, Jonsson JJ, Thorarensen O, Ludvigsson P, Gudbjartsson H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Arnar DO, Magnusson OT, Kong A, Masson G, Thorsteinsdottir U, Helgason A, Sulem P, Stefansson K (2015) Large-scale whole-genome sequencing of the Icelandic population. Nat Genet 47:435–444. https://doi.org/10.1038/ng.3247

Guenther BD, Sheppard CA, Tran P, Rozen R, Matthews RG, Ludwig ML (1999) The structure and properties of methylenetetrahydrofolate reductase from Escherichia coli suggest how folate ameliorates human hyperhomocysteinemia. Nat Struct Biol 6:359–365. https://doi.org/10.1038/7594

Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. J Hum Genet 47:605–610. https://doi.org/10.1007/s100380200092

Hamilton HB, Neel JV, Kobara TY, Ozaki K (1961) The frequency in Japan of carriers of the rare "recessive" gene causing acatalasemia. J Clin Invest 40:2199–2208. https://doi.org/10.1172/JCI104446

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y (2002) JSNP: a database of common gene variations in the Japanese population. Nucleic Acids Res 30:158–162

Holtzman C, Slazyk WE, Cordero JF, Hannon WH (1986) Descriptive epidemiology of missed cases of phenylketonuria and congenital hypothyroidism. Pediatrics 78:553–558

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Japanese Archipelago Human Population Genetics Consortium, Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, Kimura R, Ohashi J, Tajima A, Yamamoto T, Tanabe H, Mano S, Suto Y, Kaname T, Naritomi K, Yanagi K, Niikawa N, Omoto K, Tokunaga K, Saitou N (2012) The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. J Hum Genet 57: 787–795. https://doi.org/10.1038/jhg.2012.114

Kamada F, Aoki Y, Narisawa A, Abe Y, Komatsuzaki S, Kikuchi A, Kanno J, Niihori T, Ono M, Ishii N, Owada Y, Fujimura M, Mashimo Y, Suzuki Y, Hata A, Tsuchiya S, Tominaga T, Matsubara Y, Kure S (2011) A genome-wide association study identifies RNF213 as the first Moyamoya disease gene. J Hum Genet 56:34–40. https://doi.org/10.1038/jhg.2010.132

Kamatani N, Kuroshima S, Hakoda M, Palella TD, Hidaka Y (1990) Crossovers within a short DNA sequence indicate a long evolutionary history of the APRT*J mutation. Hum Genet 85:600–604

Kawai Y, Mimori T, Kojima K, Nariai N, Danjoh I, Saito R, Yasuda J, Yamamoto M, Nagasaki M (2015) Japonica array: improved genotype imputation by designing a population-specific SNP array with 1070 Japanese individuals. J Hum Genet 60:581–587. https://doi.org/10.1038/jhg.2015.68

Kessler MD, Yerges-Armstrong L, Taub MA, Shetty AC, Maloney K, Jeng LJ, Ruczinski I, Levin AM, Williams LK, Beaty TH, Mathias RA, Barnes KC, Connor TD (2016) Consortium on Asthma among African-ancestry Populations in the A, O'. Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. Nat Commun 7:12521. https://doi.org/10.1038/ncomms12521

Kitagawa T (2012) Newborn screening for inborn errors of metabolism in Japan. A history of the development of newborn screening. Pediatr Endocrinol Rev 10(Suppl 1):8–25

Knox WE, Messinger EC (1958) The detection in the heterozygote of the metabolic effect of the recessive gene for phenylketonuria. Am J Hum Genet 10:53–60

Kobayashi K, Shaheen N, Kumashiro R, Tanikawa K, O'Brien WE, Beaudet AL, Saheki T (1993) A search for the primary abnormality in adult-onset type II citrullinemia. Am J Hum Genet 53:1024–1030

Kobayashi K, Kakinoki H, Fukushige T, Shaheen N, Terazono H, Saheki T (1995) Nature and frequency of mutations in the argininosuccinate synthetase gene that cause classical citrullinemia. Hum Genet 96:454–463

Kobayashi K, Sinasac DS, Iijima M, Boright AP, Begum L, Lee JR, Yasuda T, Ikeda S, Hirano R, Terazono H, Crackower MA, Kondo I, Tsui LC, Scherer SW, Saheki T (1999) The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. Nat Genet 22:159–163. https://doi.org/10.1038/9667

Konecki DS, Lichter-Konecki U (1991) The phenylketonuria locus: current knowledge about alleles and mutations of the phenylalanine hydroxylase gene in various populations. Hum Genet 87:377–388

Koshiba S, Motoike I, Kojima K, Hasegawa T, Shirota M, Saito T, Saigusa D, Danjoh I, Katsuoka F, Ogishima S, Kawai Y, Yamaguchi-Kabata Y, Sakurai M, Hirano S, Nakata J, Motohashi H, Hozawa A, Kuriyama S, Minegishi N, Nagasaki M, Takai-Igarashi T, Fuse N, Kiyomoto H, Sugawara J, Suzuki Y, Kure S, Yaegashi N, Tanabe O, Kinoshita K, Yasuda J, Yamamoto M (2016) The structural origin of metabolic quantitative diversity. Sci Rep 6:31463. https://doi.org/10.1038/srep31463

Koshiba S, Motoike I, Saigusa D, Inoue J, Shirota M, Katoh Y, Katsuoka F, Danjoh I, Hozawa A, Kuriyama S, Minegishi N, Nagasaki M, Takai-Igarashi T, Ogishima S, Fuse N, Kure S, Tamiya G, Tanabe O, Yasuda J, Kinoshita K, Yamamoto M (2018) Omics research project on prospective cohort studies from the Tohoku Medical Megabank Project. Genes Cells 23:406–417. https://doi.org/10.1111/gtc.12588

Kuriyama S, Yaegashi N, Nagami F, Arai T, Kawaguchi Y, Osumi N, Sakaida M, Suzuki Y, Nakayama K, Hashizume H, Tamiya G, Kawame H, Suzuki K, Hozawa A, Nakaya N, Kikuya M, Metoki H, Tsuji I, Fuse N, Kiyomoto H, Sugawara J, Tsuboi A, Egawa S, Ito K, Chida K, Ishii T, Tomita H, Taki Y, Minegishi N, Ishii N, Yasuda J, Igarashi K, Shimizu R, Nagasaki M, Koshiba S, Kinoshita K, Ogishima S, Takai-Igarashi T, Tominaga T, Tanabe O, Ohuchi N, Shimosegawa T, Kure S, Tanaka H, Ito S, Hitomi J, Tanno K, Nakamura M, Ogasawara K, Kobayashi S, Sakata K, Satoh M, Shimizu A, Sasaki M, Endo R, Sobue K,, Yamamoto M (2016) Tohoku Medical Megabank Project Study Group T. The Tohoku Medical Megabank Project: Design and Mission. J Epidemiol 26:493–511. https://doi.org/10.2188/jea.JE20150268

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR (2016) ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44:D862–D868. https://doi.org/10.1093/nar/gkv1222

Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, McDade TW, Wang Y, Li Y, Levy S, Borja JB, Lange EM, Adair LS, Mohlke KL (2010) Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. Hum Mol Genet 19:2050–2058. https://doi.org/10.1093/hmg/ddq062

Larsson SC, Giovannucci E, Wolk A (2006) Folate intake, MTHFR polymorphisms, and risk of esophageal, gastric, and pancreatic cancer: a meta-analysis. Gastroenterology 131:1271–1283. doi: S0016-5085(06)01727-6

Lee DH, Koo SK, Lee KS, Yeon YJ, Oh HJ, Kim SW, Lee SJ, Kim SS, Lee JE, Jo I, Jung SC (2004) The molecular basis of phenylketonuria in Koreans. J Hum Genet 49:617–621. https://doi.org/10.1007/s10038-004-0197-5

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291. https://doi.org/10.1038/nature19057

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC,

Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC (2007) The diploid genome sequence of an individual human. PLoS Biol 5:e254. https://doi.org/10.1371/journal.pbio.0050254

Li Q, Wang K (2017) InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 100:267–280. https://doi.org/10.1016/j.ajhg.2017.01.004

Long T, Hicks M, Yu HC, Biggs WH, Kirkness EF, Menni C, Zierer J, Small KS, Mangino M, Messier H, Brewerton S, Turpaz Y, Perkins BA, Evans AM, Miller LA, Guo L, Caskey CT, Schork NJ, Garner C, Spector TD, Venter JC, Telenti A (2017) Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet 49:568–578. https://doi.org/10.1038/ng.3809

McCabe LL, McCabe ER (2002) Newborn screening as a model for population screening. Mol Genet Metab 75:299–307. https://doi.org/10.1016/S1096-7192(02)00005-7

Mimori A, Hidaka Y, Wu VC, Tarle SA, Kamatani N, Kelley WN, Pallela TD (1991) A mutant allele common to the type I adenine phosphoribosyltransferase deficiency in Japanese subjects. Am J Hum Genet 48:103–107

Minari J, Brothers KB, Morrison M (2018) Tensions in ethics and policy created by National Precision Medicine Programs. Human Genomics 12:22. https://doi.org/10.1186/s40246-018-0151-9

Miyashita M, Ishikuro M, Kikuya M, Yamanaka C, Mizuno S, Nagai M, Sato Y, Obara T, Metoki H, Kikuchi A, Nakaya N, Hozawa A, Tsuji I, Yaegashi N, Yamamoto M, Kure S, Kuriyama S (2017) Development of a Questionnaire Method of Screening for Citrin Deficiency in Schoolchildren. Journal of Pediatrics and Congenital Disorders 4

Morton NE, Crow JF, Muller HJ (1956) An Estimate of the Mutational Damage in Man from Data on Consanguineous Marriages. Proc Natl Acad Sci U S A 42:855–863

Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, Yamaguchi-Kabata Y, Yokozawa J, Danjoh I, Saito S, Sato Y, Mimori T, Tsuda K, Saito R, Pan X, Nishikawa S, Ito S, Kuroki Y, Tanabe O, Fuse N, Kuriyama S, Kiyomoto H, Hozawa A, Minegishi N, Douglas Engel J, Kinoshita K, Kure S, Yaegashi N,, Yamamoto M (2015) ToMMo Japanese Reference Panel Project. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat Commun 6:8018. https://doi.org/10.1038/ncomms9018

Nagata N, Matsuda I, Oyanagi K (1991) Estimated frequency of urea cycle enzymopathies in Japan. Am J Med Genet 39:228–229. https://doi.org/10.1002/ajmg.1320390226

Nakajo A (1939) Secondary hypertrophic osteoperiostosis with pernio (Japanese). J Derm Urol 45:77–86

Neel JV, Schull WJ (1962) The effect of inbreeding on mortality and morbidity in two Japanese cities. Proc Natl Acad Sci U S A 48:573–582

Nishimura N, Deki T, Kato S. (Japanese (1950) Hypertrophic pulmonary osteo-arthropathy with pernio-like eruption in the two families: report of the three cases. Jpn J Derm Venereol 60:136–141

Okano Y, Asada M, Kang Y, Nishi Y, Hase Y, Oura T, Isshiki G (1998) Molecular characterization of phenylketonuria in Japanese patients. Hum Genet 103:613–618

Okano Y, Kudo S, Nishi Y, Sakaguchi T, Aso K (2011) Molecular characterization of phenylketonuria and tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency in Japan. J Hum Genet 56:306–312. https://doi.org/10.1038/jhg.2011.10

Punj S, Akkari Y, Huang J, Yang F, Creason A, Pak C, Potter A, Dorschner MO, Nickerson DA, Robertson PD, Jarvik GP, Amendola LM, Schleit J, Simpson D, Rope AF, Reiss J, Kauffman T, Gilmore MJ, Himes P, Wilfond B, Goddard K, Richards SC (2018) Preconception Carrier Screening by Genome Sequencing: Results from the Clinical Laboratory. The American Journal of Human Genetics. https://doi.org/10.1016/j.ajhg.2018.04.004

Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Sherry ST, Watson MS, ClinGen (2015) ClinGen–the Clinical Genome Resource. N Engl J Med 372: 2235-42. https://doi.org/10.1056/NEJMsr1406261

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424. https://doi.org/10.1038/gim.2015.30

Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639. https://doi.org/10.1126/science.1186802

Robinson WP (2000) Mechanisms leading to uniparental disomy and their clinical consequences. Bioessays 22:452–459

Satoh C, Neel JV, Yamashita A, Goriki K, Fujita M, Hamilton HB (1983) The frequency among Japanese of heterozygotes for deficiency variants of 11 enzymes. Am J Hum Genet 35:656–674

Schull WJ, Komatsu I, Nagano H, Yamamoto M (1968) Hirado: temporal trends in inbreeding and fertility. Proc Natl Acad Sci U S A 59:671–679

Schull WJ, Nagano H, Yamamoto M, Komatsu I (1970) The effects of parental consanguinity and inbreeding in Hirado, Japan. I. Stillbirths and prereproductive mortality. Am J Hum Genet 22:239–262

Song MJ, Lee ST, Lee MK, Ji Y, Kim JW, Ki CS (2012) Estimation of carrier frequencies of six autosomal-recessive Mendelian disorders in the Korean population. J Hum Genet 57:139–144. https://doi.org/10.1038/jhg.2011.144

Song W, Gardner SA, Hovhannisyan H, Natalizio A, Weymouth KS, Chen W, Thibodeau I, Bogdanova E, Letovsky S, Willis A, Nagan N (2015) Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. Genet Med. https://doi.org/10.1038/gim.2015.180

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN (2003) Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 21:577–581. https://doi.org/10.1002/humu.10212

Summar ML, Koelker S, Freedenberg D, Le Mons C, Haberle J, Lee HS, Kirmse B, European R, Network for Intoxication Type Metabolic Diseases. Electronic address hwe-ioeip, Members of the Urea Cycle Disorders Consortium. Electronic address hreueu (2013) The incidence of urea cycle disorders. Mol Genet Metab 110: 179–80. https://doi.org/10.1016/j.ymgme.2013.07.008

Suzuki Y, Aoki Y, Ishida Y, Chiba Y, Iwamatsu A, Kishino T, Niikawa N, Matsubara Y, Narisawa K (1994) Isolation and characterization of mutations in the human holocarboxylase synthetase cDNA. Nat Genet 8:122–128. https://doi.org/10.1038/ng1094-122

Suzuki Y, Yang X, Aoki Y, Kure S, Matsubara Y (2005) Mutations in the holocarboxylase synthetase gene HLCS. Hum Mutat 26:285–290. https://doi.org/10.1002/humu.20204

Tabata A, Sheng JS, Ushikai M, Song YZ, Gao HZ, Lu YB, Okumura F, Iijima M, Mutoh K, Kishida S, Saheki T, Kobayashi K (2008) Identification of 13 novel mutations including a retrotransposal insertion in SLC25A13 gene and frequency of 30 mutations found in patients with citrin deficiency. J Hum Genet 53:534–545. https://doi.org/10.1007/s10038-008-0282-2

Tabor HK, Auer PL, Jamal SM, Chong JX, Yu JH, Gordon AS, Graubert TA, O'Donnell CJ, Rich SS, Nickerson DA, Project NES, Bamshad MJ (2014) Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: implications for the return of incidental results. Am J Hum Genet 95:183–193. https://doi.org/10.1016/j.ajhg.2014.07.006

Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, Koshiba S, Yamamoto M, Kinoshita K (2018) jMorp: Japanese Multi Omics Reference Panel. Nucleic Acids Res 46:D551–D557. https://doi.org/10.1093/nar/gkx978

Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, Otsuki A, Gocho C, Sakurai-Yageta M, Danjo I, Motoike IN, Yamaguchi-Kabata Y, Shirota M, Koshiba S, Nagasaki M, Minegishi N, Hozawa A, Kuriyama S, Shimizu A, Yasuda J, Fuse N, The Tohoku Medical Megabank Project Study Group, Tamiya G, Yamamoto M, Kinoshita K (2019) 3.5KJPNv2, An allele frequency panel of 3,552 Japanese Individuals. bioRxiv. doi: https://doi.org/10.1101/529529

Takahara S, Hamilton HB, Neel JV, Kobara TY, Ogura Y, Nishimura ET (1960) Hypocatalasemia: a new genetic carrier state. J Clin Invest 39:610–619. https://doi.org/10.1172/JCI104075

Takahashi Y, Sano R, Kominato Y, Kubo R, Takahashi K, Nakajima T, Takeshita H, Ishige T (2016) A case of sudden unexpected infant death involving a homozygotic twin with the thermolabile CPT2 variant, accompanied by rotavirus infection and treatment with an antibiotic containing pivalic acid. Leg Med (Tokyo) 22:13–17. https://doi.org/10.1016/j.legalmed.2016.07.005

Taroni F, Verderio E, Fiorucci S, Cavadini P, Finocchiaro G, Uziel G, Lamantea E, Gellera C, DiDonato S (1992) Molecular characterization of inherited carnitine palmitoyltransferase II deficiency. Proc Natl Acad Sci U S A 89:8429–8433

UK 10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JR, Xu C, Futema M, Lawson D, Iotchkova V, Schiffels S, Hendricks AE, Danecek P, Li R, Floyd J, Wain LV, Barroso I, Humphries SE, Hurles ME, Zeggini E, Barrett JC, Plagnol V, Richards JB, Greenwood CM, Timpson NJ, Durbin R, Soranzo N (2015) The UK10K project identifies rare variants in health and disease. Nature 526: 82–90. https://doi.org/10.1038/nature14962

Wada Y, Kikuchi A, Arai-Ichinoi N, Sakamoto O, Takezawa Y, Iwasawa S, Niihori T, Nyuzuki H, Nakajima Y, Ogawa E, Ishige M, Hirai H, Sasai H, Fujiki R, Shirota M, Funayama R, Yamamoto M, Ito T, Ohara O, Nakayama K, Aoki Y, Koshiba S, Fukao T, Kure S (2018) Biallelic GALM pathogenic variants cause a novel type of galactosemia. Genetics in Medicine. https://doi.org/10.1038/s41436-018-0340-x

Wang T, Okano Y, Eisensmith RC, Harvey ML, Lo WH, Huang SZ, Zeng YT, Yuan LF, Furuyama JI, Oura T et al (1991) Founder effect of a prevalent phenylketonuria mutation in the Oriental population. Proc Natl Acad Sci U S A 88:2146–2150

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164. https://doi.org/10.1093/nar/gkq603

Weisberg IS, Jacques PF, Selhub J, Bostom AG, Chen Z, Curtis Ellison R, Eckfeldt JH, Rozen R (2001) The 1298A–>C polymorphism in methylenetetrahydrofolate reductase (MTHFR): in vitro expression and association with homocysteine. Atherosclerosis 156:409–415

Wongkittichote P, Sukasem C, Kikuchi A, Aekplakorn W, Jensen LT, Kure S, Wattanasirichaigoon D (2013a) Screening of SLC25A13 mutation in the Thai population. World J Gastroenterol 19:7735–7742. https://doi.org/10.3748/wjg.v19.i43.7735

Wongkittichote P, Tungpradabkul S, Wattanasirichaigoon D, Jensen LT (2013b) Prediction of the functional effect of novel SLC25A13 variants using a S. cerevisiae model of AGC2 deficiency. J Inherit Metab Dis 36:821–830. https://doi.org/10.1007/s10545-012-9543-5

Woo SL (1989) Molecular basis and population genetics of phenylketonuria. Biochemistry 28:1–7

Woo SL, Lidsky AS, Guttler F, Chandra T, Robson KJ (1983) Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. Nature 306:151–155

Xue Y, Chen Y, Ayub Q, Huang N, Ball EV, Mort M, Phillips AD, Shaw K, Stenson PD, Cooper DN, Tyler-Smith C, Genomes Project C (2012) Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am J Hum Genet 91:1022–1032. https://doi.org/10.1016/j.ajhg.2012.10.015

Yamaguchi S (2008) Newborn screening in Japan: restructuring for the new era. Ann Acad Med Singapore 37:13–15

Yamaguchi S (2012) Newborn mass-screening by tandem mass spectrometry research: Preparation and quality improvement (in Japanese). Research Report of Health, Labor and Welfare Science Research Grant for year 2010–2012, Tokyo, pp 3–14

Yamaguchi M, Yanase T, Nagano H, Nakamoto N (1970) Effects of inbreeding on mortlaity in Fukuoka population. Am J Hum Genet 22:145–159

Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, Nakamura Y, Kamatani N (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83:445–456. https://doi.org/10.1016/j.ajhg.2008.08.019

Yamaguchi-Kabata Y, Nariai N, Kawai Y, Sato Y, Kojima K, Tateno M, Katsuoka F, Yasuda J, Yamamoto M, Nagasaki M (2015) iJGVD: an integrative Japanese genome variation database based on whole-genome sequencing. Hum Genome Var 2:15050. https://doi.org/10.1038/hgv.2015.50

Yamaguchi-Kabata Y, Yasuda J, Tanabe O, Suzuki Y, Kawame H, Fuse N, Nagasaki M, Kawai Y, Kojima K, Katsuoka F, Saito S, Danjoh I, Motoike IN, Yamashita R, Koshiba S, Saigusa D, Tamiya G, Kure S, Yaegashi N, Kawaguchi Y, Nagami F, Kuriyama S, Sugawara J, Minegishi N, Hozawa A, Ogishima S, Kiyomoto H, Takai-Igarashi T, ToMMo Study Group, Kinoshita Yamamoto K M (2018) Evaluation of reported pathogenic variants and their frequencies in a Japanese population based on a whole-genome reference panel of 2049 individuals. J Hum Genet 63:213–230. https://doi.org/10.1038/s10038-017-0347-1

Yamamoto S, Abe H, Kohgo T, Ogawa A, Ohtake A, Hayashibe H, Sakuraba H, Suzuki Y, Aramaki S, Takayanagi M, Hasegawa S, Niimi H (1996) Two novel gene mutations (Glu174–>Lys, Phe383–>Tyr) causing the "hepatic" form of carnitine palmitoyltransferase II deficiency. Hum Genet 98:116–118

Yamamoto T, Mishima H, Mizukami H, Fukahori Y, Umehara T, Murase T, Kobayashi M, Mori S, Nagai T, Fukunaga T, Yamaguchi S, Yoshiura KI, Ikematsu K (2015) Metabolic autopsy with next generation sequencing in sudden unexpected death in infancy: Postmortem diagnosis of fatty acid oxidation disorders. Mol Genet Metab Rep 5:26–32. https://doi.org/10.1016/j.ymgmr.2015.09.005

Yamashiro Y, Shimizu T, Oguchi S, Shioya T, Nagata S, Ohtsuka Y (1997) The estimated incidence of cystic fibrosis in Japan. J Pediatr Gastroenterol Nutr 24:544–547

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M, Person R, Bekheirnia MR, Leduc MS, Kirby A, Pham P, Scull J, Wang M, Ding Y, Plon SE, Lupski JR, Beaudet AL, Gibbs RA, Eng CM (2013) Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 369:1502–1511. https://doi.org/10.1056/NEJMoa1306555

Yasuda J, Kinoshita K, Katsuoka F, Danjoh I, Sakurai-Yageta M, Motoike IN, Kuroki Y, Saito S, Kojima K, Shirota M, Saigusa D, Otsuki A, Kawashima J, Yamaguchi-Kabata Y, Tadaka S, Aoki Y, Mimori T, Kumada K, Inoue J, Makino S, Kuriki M, Fuse N, Koshiba S, Tanabe O, Nagasaki M, Tamiya G, Shimizu R, Takai-Igarashi T, Ogishima S, Hozawa A, Kuriyama S, Sugawara J, Tsuboi A, Kiyomoto H, Ishii T, Tomita H, Minegishi N, Suzuki Y, Suzuki K, Kawame H, Tanaka H, Taki Y, Yaegashi N, Kure S, Nagami F, Tohoku Medical Megabank Project Study Group, Kosaki K, Sutoh Y, Hachiya T, Shimizu A, Sasaki M, Yamamoto M (2018) Genome analyses for the Tohoku Medical Megabank Project toward establishment of personalized healthcare. J Biochem. https://doi.org/10.1093/jb/mvy096

Zaffanello M, Zamboni G, Maffeis C, Tato L (2003) Neonatal birth parameters of positive newborns at PKU screening as predictors of false-positive and positive results at recall-testing. J Med Screen 10:181–183. https://doi.org/10.1258/096914103771773276

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Yumi Yamaguchi-Kabata[1,2] · Jun Yasuda[1,2,3] · Akira Uruno[1,2] · Kazuro Shimokawa[1] · Seizo Koshiba[1] · Yoichi Suzuki[1,2,4] · Nobuo Fuse[1,2] · Hiroshi Kawame[1,2] · Shu Tadaka[1] · Masao Nagasaki[1,2,5] · Kaname Kojima[1,2,5] · Fumiki Katsuoka[1,2] · Kazuki Kumada[1] · Osamu Tanabe[1,2,6] · Gen Tamiya[1,2,7] · Nobuo Yaegashi[1,2] · Kengo Kinoshita[1,5,8,9] · Masayuki Yamamoto[1,2,9] · Shigeo Kure[1,2] · The Tohoku Medical Megabank Project Study Group

1 Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryo-machi, Aoba-ku, Sendai 980-8573, Japan

2 Graduate School of Medicine, Tohoku University, 2-1 Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan

3 Present address: Miygai Cancer Center Research Institute, 47-1, Noda-yama,, Medeshima-shiode, Natori 981-1293, Japan

4 Present address: Ageo Central General Hospital, 1-10-10 Kashiwa-za, Ageo 362-8588, Japan

5 Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki Aza-Aoba, Aoba-ku, Sendai 980-8579, Japan

6 Present address: Radiation Effects Research Foundation Hiroshima Laboratory, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732-0815, Japan

7 Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan

8 Institute of Development, Aging and Cancer, Tohoku University, 4-1 Seiryo-machi, Aoba-ku, Sendai 980-8575, Japan

9 Advanced Research center for Innovations in Next-Generation Medicine, Tohoku University, 2-1 Seiryo-machi, Aoba-ku, Sendai 980-8573, Japan