

ARTICLE **OPEN**

Check for updates

Genetic etiology of truncus arteriosus excluding 22q1[1](http://crossmark.crossref.org/dialog/?doi=10.1038/s10038-024-01223-y&domain=pdf).2 deletion syndrome and identification of c.1617del, a prevalent variant in TMEM260, in the Japanese population

Hisao Yaoita^{1,9⊠}, Eiichiro Kawai^{2,9}, Jun Takayama^{3,4,5,6,9}, Shinya Iwasawa¹, Naoya Saijo¹, Masayuki Abiko⁷, Kouta Suzuki⁷, $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ masato Kimura², Akira Ozawa², Gen Tamiya^{3,4,5,6}, Shigeo Kure^{1,6,8} and Atsuo Kikuchi $\mathbf{D}^{1,6}$

© The Author(s) 2024, corrected publication 2024

Truncus Arteriosus (TA) is a congenital heart disease characterized by a single common blood vessel emerging from the right and left ventricles instead of the main pulmonary artery and aorta. TA accounts for 4% of all critical congenital heart diseases. The most common cause of TA is 22q11.2 deletion syndrome, accounting for 12–35% of all TA cases. However, no major causes of TA other than 22q11.2 deletion have been reported. We performed whole-genome sequencing of 11 Japanese patients having TA without 22q11.2 deletion. Among five patients, we identified pathogenic variants in TMEM260; the biallelic loss-of-function variants of which have recently been associated with structural heart defects and renal anomalies syndrome (SHDRA). In one patient, we identified a de novo pathogenic variant in GATA6, and in another patient, we identified a de novo probably pathogenic variant in NOTCH1. Notably, we identified a prevalent variant in TMEM260 (ENST00000261556.6), c.1617del (p.Trp539Cysfs*9), in 8/22 alleles among the 11 patients. The c.1617del variant was estimated to occur approximately 23 kiloyears ago. Based on the allele frequency of the c.1617del variant in the Japanese population (0.36%), approximately 26% of Japanese patients afflicted with TA could harbor homozygous c.1617del variants. This study highlights TMEM260, especially c.1617del, as a major genetic cause of TA in the Japanese population.

Journal of Human Genetics (2024) 69:177–183;<https://doi.org/10.1038/s10038-024-01223-y>

INTRODUCTION

Congenital heart disease (CHD) is one of the most common birth defects, affecting approximately 1% of newborns [[1](#page-5-0)]. Its etiology is still unknown for the majority of cases. To date, genetic causes account for only 34% of all CHD cases, with aneuploidy, copy number variants (CNVs), and single-nucleotide variants (SNVs) accounting for 13%, 10%, and 10% of the cases, respectively [[1\]](#page-5-0).

Truncus Arteriosus (TA) is a CHD characterized by a single common blood vessel that emerges from the right and left ventricles instead of the main pulmonary artery and aorta. TA has an incidence of 3–10 per 100,000 live births and accounts for 4% of all critical CHDs [[2](#page-5-0)]. The 22q11.2 deletion syndrome is a major cause of TA and is associated with 12% to 35% of TA cases [\[3\]](#page-5-0). TBX1 is a gene responsible for CHDs of 22q11.2 deletion syndrome [\[4\]](#page-5-0). Apart from 22q11.2 deletion syndrome, several gene abnormalities, including NKX2–5 [[5\]](#page-5-0), NKX2–6 [\[6\]](#page-5-0), and GATA6 [\[7\]](#page-5-0), have also been associated with TA. Recently, biallelic pathogenic variants of TMEM260, which encodes a multipass transmembrane protein, O-mannosyltransferase [\[8\]](#page-5-0), have been associated with structural heart defects and renal anomalies syndrome (SHDRA syndrome; MIM #617478), including TA [\[8](#page-5-0)–[12\]](#page-5-0). However, no major genetic causes of TA, other than 22q11.2 deletion, have been reported. The aim of this study was to explore the genetic causes of TA other than the 22q11.2 deletion in the Japanese population.

MATERIAL AND METHODS **Patients**

Eleven patients with unexplained TA were studied (Table [1](#page-1-0)). Patients with chromosome 22 abnormalities were excluded. This study was approved by the Ethics Committee of the Tohoku University School of Medicine and Tohoku Medical Megabank organization. Written informed consent was obtained from patients or their parents.

Genomic analysis

Genomic DNA was extracted from blood or saliva samples of patients and their families. We performed whole-genome sequencing (WGS) with DNBSEQ T7 (MGI Tech) in 150 bp paired-end mode using PCR-free libraries, according to previously described methods with minor modifications [\[13](#page-5-0), [14\]](#page-5-0). The sequenced reads were mapped to the hg19 human reference genome using BWA MEM (ver 0.7.17-r1188). SNVs, short indels, and CNVs were called using the Genome Analysis Toolkit software v4.2.6.1, and structural variants were called using Smoove (v0.2.8). The variants were

¹Department of Pediatrics, Tohoku University Graduate School of Medicine, Sendai, Japan. ²Department of Pediatric Cardiology, Miyagi Children's Hospital, Sendai, Japan. ³Department of AI and Innovative Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan. ⁴Tohoku Medical Megabank organization, Tohoku University, Sendai, Japan. ⁵Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Tokyo, Japan. ⁶Department of Rare Disease Genomics, Tohoku University Graduate School of Medicine, Sendai, Japan. ⁷Department of Pediatrics, Yamagata University Graduate School of Medicine, Yamagata, Japan. ⁸Miyagi Children's Hospital, Sendai, Japan. ⁹These authors contributed equally: Hisao Yaoita, Eiichiro Kawai, Jun Takayama. [⊠]email: [yaoita.hisao.r4@dc.tohoku.ac.jp;](mailto:yaoita.hisao.r4@dc.tohoku.ac.jp) atsuo.kikuchi.d4@tohoku.ac.jp

ξ. \geq Į. $\frac{2}{\pi}$ $\bar{\vec{P}}$ ļ. š. ⋋ Ž.

H. Yaoita et al.

178

then annotated using SnpEff (v.5.1) after quality filtering. We considered introns ±100 bp from exons, CNVs, and structural variants. All candidate

To test for NMD, Epstein-Barr virus transformed lymphoblastoid cell lines (EBV-LCLs) were established from Patient 1 and her parents using a previously described method [[15](#page-5-0)]. The cells were treated with cycloheximide at a final concentration of 100 μg/ml. Cell lysates were harvested four hours after treatment. Total RNA was extracted from both treated and untreated cell lysates using the RNeasy Kit (Qiagen Inc., Valencia, CA, USA). RT-PCR was performed using a PrimeScript™ RT Reagent Kit with gDNA eraser (Perfect Real Time) (TaKaRa, Shiga, Japan), followed by Sanger sequencing.

The minor allele frequency of c.1617del in TMEM260 in various populations, including Japanese, East Asian, and Southeast Asian populations, were obtained from public genome databases as follows: gnomAD v2.1.1 [\(https://gnomad.broadinstitute.org](https://gnomad.broadinstitute.org)) for various populations [\[16](#page-5-0)], ToMMo 54KJPN (https://jmorp.megabank.tohoku.ac.jp) for the Japanese population [\[17](#page-5-0)], the ChinaMAP [\(http://www.mbiobank.com\)](http://www.mbiobank.com) for the Chinese population [\[18](#page-6-0)], Taiwan Biobank [\(https://taiwanview.twbiobank.org.tw](https://taiwanview.twbiobank.org.tw)) for the Taiwanese population [\[19](#page-6-0)], Korean Variant Archive 2 (KOVA 2, [https://](https://www.kobic.re.kr/kova/) [www.kobic.re.kr/kova\)](https://www.kobic.re.kr/kova/) for the Korean population [\[20\]](#page-6-0), Thai Reference Exome Database (T-REx, <https://trex.nbt.or.th>) for the Thai population [\[21](#page-6-0)], and the 1000 Vietnamese Genome Project (1KVG) ([https://](https://genome.vinbigdata.org/about)

variants were confirmed by Sanger sequencing.

Nonsense-mediated mRNA decay (NMD) inhibition

Variant database surveys in various populations

genome.vinbigdata.org/about) for the Vietnamese population.

Haplotype dataset surveys to estimate variant age

site per generation [[26,](#page-6-0) [27](#page-6-0)].

The age at which the c.1617del variant arose was estimated using GEVA software (Github commit tag edeafb7) [\[22](#page-6-0)] and the haplotype dataset of 3552 Japanese individuals, which consisted of 7104 phased haplotypes and 1,442,809 variants on chromosome 14 [\[23](#page-6-0)]. Because GEVA is designed to treat SNPs, the variant chr14:57099780TG>T was modified to chr14:57099780T->G. Using this dataset, we estimated the variant age by GEVA under a recombination model [[22](#page-6-0)] with a generation time of 25 years
[\[24](#page-6-0)], effective population size of 10,000 [\[25\]](#page-6-0), and mutation rate of 10⁻¹⁰ per

Analysis of CHD and renal function in c.1617del heterozygous population

Considering the possibility that not only c.1617del homozygotes, but also heterozygotes may exhibit minor renal phenotypes, a population with the c.1617del variant was selected from ToMMo 54KJPN and compared with the wild-type population for blood data associated with renal function (serum creatinine, serum blood urea nitrogen [BUN], serum uric acid, and serum cystatin C) and for the presence of CHD. Linear regression analysis was performed for creatinine, BUN, uric acid, and cystatin C levels, using sex and age as covariates.

RESULTS

The clinical characteristics of each patient with TA from each of the 11 families included in this study are displayed in Table [1.](#page-1-0) There were five males and six females, with ages ranging from 8 days to 29 years. None of the patients were related to each other, and none were born to a consanguineous couple. Intracardiac repair was completed in all but two infant patients (Patients 1 and 7). We identified putative biallelic pathogenic variants in TMEM260 in five patients, a de novo heterozygous pathogenic variant in GATA6 in one patient, and a de novo heterozygous probably pathogenic variant in NOTCH1 in one patient (Fig. [1,](#page-1-0) Table 1, Supplementary Table 1). Patients 1−3 harbored the biallelic c.1617del (p.Trp539Cysfs*9) variant in TMEM260. A compound heterozygous patient containing a c.1617del variant reportedly exhibited a SHDRA-like phenotype [[10\]](#page-5-0). Patients 4 and 5 had other heterozygous pathogenic variants, c.332dup, p.(Thr112Hisfs*36) and c.1960C>T, p.(Gln654*), respectively, in addition to the heterozygous c.1617del variant. The p.(Gln654*) variant identified in Patient 5 was located in the last exon. This variant is expected to escape NMD but is expected to cause a loss of the end of the TPR domain, which is likely involved in substrate recognition and selective recruitment of IPT domains for O-linked mannose glycosylation (Supplementary Fig. 1) [[8](#page-5-0)]. Additionally, the substantial loss of protein length can affect the stability of the protein. Therefore, we considered p.(Gln654*) to be a likely pathogenic

(p.Trp539Cysfs*9) p.Thr112Hisfs*36 p.GIn654 p.Leu447Valfs*9
p.Cys453Arg c.193-2A>G p.Tvr567Thrfs*27 $p.GIn620$ n Gln465* Pro549L ufs*46 p.Cvs98T $6.$ Glu582Gln p.Glu55Phefs*20 p.Arg115Lys p.Tyr470 Long isoform of *TMEM260*
(ENST00000261556.6) Short isoform of *TMEM260*
(ENST00000538838.1)

c.1617del

Fig. 1 Schematic of the gene structure of long isoform of TMEM260 (ENST00000261556.6) and short isoform of TMEM260 (ENST000005338838.1). The variants reported in this study (shown in bold and red) and the pathogenic variants have been published previously [\[8](#page-5-0)–[12\]](#page-5-0)

Fig. 2 The frameshift variant c.1617del in TMEM260 leads to NMD. EBV-LCLs from the parents of Patient 1 were treated with cycloheximide to suppress NMD. Sanger sequencing of the c.1617del variant was also performed. The red arrowheads indicate the location of c.1617del. The upper and lower panels show the results before and after CHX loading, respectively

variant. A de novo missense variant, c.1367G>A, p.(Arg456His), in GATA6 was identi fied in Patient 6. There have been reports of CHD patients with the same variant [[28\]](#page-6-0), and patients with TA or tetralogy of Fallot harboring p.(Arg456Cys) [\[28](#page-6-0)]. A de novo missense variant, c.545G>A, p.(Cys182Tyr), in NOTCH1 was identified in Patient 7. This variant has been reported in a large nonsyndromic tetralogy of Fallot cohort [\[29\]](#page-6-0) although its pathogenicity has not been functionally evaluated. No additional potential pathogenic variants were identi fied in this study.

Among the five patients who had TMEM260 variants, regarding the types of TA, three were type I (Collett and Edwards classi fication) and type A1 (Vanpraagh classification), and two were type II and type A2 (Table [1\)](#page-1-0). Other cardiac malformations included cor triatriatum and coronary artery fistula (CAF) in Patient 1, right aortic arch (RAA) in Patient 2, persistent left superior vena cava (PLSVC) in Patients 1 and 5, and partial anomalous pulmonary venous return (PAPVR) in Patients 1, 3, and 5. With regard to renal function, no chronic elevation of creatinine levels was observed in Patients 1 through 5. However, Patient 1 developed heart and renal failure after the palliative Rastelli procedure, requiring the induction of dialysis, was unable to withdraw from dialysis, and died. Patients 2–5 have not had renal dysfunction to date, although Patient 5 had hydronephrosis. Neurodevelopmental delay was present in Patients 2 and 4, although in Patient 2 it was unclear whether the delay was congenital or due to posthemorrhagic hydrocephalus. Hearing impairment was observed in Patient 2.

Among the six patients without pathogenic variants in TMEM260, two had TA type I and type A1, one had TA type II and type A2, two had TA type A3, and one had TA type III. There were no cases of renal dysfunction. Neurodevelopmental disorders were observed in three patients: Patient 6 had dysplasia of the cortical gyrus and language developmental delay, Patient 8 had post hemorrhagic hydrocephalus, and Patient 9 had hyperactive tendencies. Hearing impairment was observed in Patient 11. Patient 6 had no symptoms suggestive of pancreatic abnormalities, a common extracardiac feature of GATA6-related disorders [\[30](#page-6-0)].

We inhibited NMD using cycloheximide to test whether the c.1617del variant leads to NMD. The c.1617del (p.Trp539Cysfs*9) transcript was detected at much lower levels than the wild-type alleles in EBV-LCLs from the parents of Patient 1. Suppression of NMD by cycloheximide treatment increased the levels of frameshift transcripts. These results indicate that frameshift alleles led to unstable transcripts that were subject to NMD (Fig. [2\)](#page-2-0).

In the Japanese population, the allele frequency of c.1617del was much higher (0.36%) than that in other populations (Table 2). Although less frequent than in the Japanese population, the allele frequency in the South Korean population was also elevated (0.14%; Table 2). In contrast, the variant was absent in the databases we analyzed for the Chinese, Taiwanese, Thai, and Vietnamese populations.

We estimated the age of the c.1617del variant. The allele frequency in the haplotype dataset of 3,552 Japanese was 0.32% (23/7104 haploids). The genotype frequencies were 99.4% for the reference homozygous genotype and 0.6% for the heterozygous genotypes. No alternative homozygous individuals were identi fied in the dataset. The c.1617del variant occurred approximately 23 kiloyears ago (kya). This estimate did not change when the mutation rate was changed by a factor of 100 or 0.01 (22 kya for both 10^{-8} and 10^{-12}). The estimate was also not significantly affected when the generation time was changed to 20 or 30 years (18 kya and 27 kya, respectively). These results suggest that the variant occurred at approximately 23 kya.

Among the 54,302 Japanese individuals participating in ToMMo 54KJPN, the data on serum creatinine, BUN, and uric acid were available for 38,207 individuals, and serum cystatin C was available for 38,206 individuals. Of these, 274 had the c.1617del heterozygous variants. There were no significant differences between individuals with heterozygous c.1617del variants and those with

the wild-type allele in any of the above indices (Supplementary Fig. 2). Data on the presence of CHD were available for 13,508 individuals, of whom 88 were heterozygous for the c.1617del variant. None of the heterozygous c.1617del individuals had CHD. None of the individuals in the ToMMo 54KJPN cohort were homozygous for the c.1617del variant.

DISCUSSION

In this study, we identified a prevalent loss-of-function variant of TMEM260, c.1617del, in 8/22 alleles from the 11 Japanese TA patients analyzed who did not have a 22q11.2 deletion: three patients were homozygous for c.1617del variants and two patients harbored compound heterozygous variants of TMEM260. Among our patient cohort, we also identified one de novo GATA6 variant and one de novo NOTCH1 variant. Based on the allele frequency of the c.1617del variant (0.36%) in the Japanese population, the incidence of biallelic c.1617del variants in the Japanese population was calculated as 1 in approximately 76,700 births (Table [2](#page-3-0)). Collectively, approximately 26% of Japanese patients with TA could be explained by the homozygous c.1617del variants in TMEM260, as the incidence of TA in Japan is estimated to be approximately 1/20,000 births, according to a survey of cardiac disease in childhood conducted by the Japanese Society of Pediatric Cardiology and Cardiac Surgery ([https://jspccs.jp/wp](https://jspccs.jp/wp-content/uploads/rare_disease_surveillance_2020.pdf)[content/uploads/rare_disease_surveillance_2020.pdf](https://jspccs.jp/wp-content/uploads/rare_disease_surveillance_2020.pdf)). Considering the other variants in TMEM260, TMEM260 mutations would account for the largest- or the second largest- portion of TA cases after 22q.11.2 deletion syndrome, in the Japanese population. Our study suggests that a genetic cause, including 22q11.2 deletion and mutations in TMEM260, GATA6, or NOTCH1, can be identified in more than half of Japanese patients with TA, although the exact proportion is unclear due to our exclusion of patients with 22q11.2 deletion from this study.

Previous reports suggested that CHD, particularly TA, is the most consistent phenotype of SHDRA [9, 11]. In terms of the types of TA, patients with TMEM260 variants in our cohort had type I (Collett and Edwards classification), type A1 (Vanpraagh classification), type II, and type A2, as previously reported. The presence of cor triatriatum and CAF and the higher rate of complicated cases of PAPVR (3/5 patients) and PLSVC (2/5 patients) than previously reported might be new features of patients carrying the c.1617del variant (Table [3](#page-4-0)).

This study suggests that the phenotype and severity of patients with TMEM260 variants may be more variable than previously reported [9, 11]. Most previous reports have shown a high rate of renal dysfunction and death at a young age in patients with SHDRA [9, 11], although recent reports have described several SHDRA patients with relatively mild symptoms [8, 10, 12]. In contrast, among the patients harboring TMEM260 variants in this study, four of the five survived from 6 to 29 years of age without any signs of renal dysfunction. In particular, Patient 3, at 29 years of age, is the oldest patient with TMEM260 variants reported to date (Table [3\)](#page-4-0) and has not only had no renal dysfunction but also has no noted developmental delays, graduated from university, and works as a caregiver. Therefore, these results suggest that some patients with TMEM260-related CHD can survive long-term without obvious developmental or renal dysfunction.

The c.1617del variant is estimated to have occurred at 23 kya. The divergence between European and Asian populations is estimated to be 20–40 kya [[27\]](#page-6-0), whereas the divergence between Han Chinese and Japanese populations is estimated to be 8–9 kya [[27\]](#page-6-0). Thus, the estimate of 23 kya is broadly consistent with the notion that the variant occurred around the time the Asian population arose and was enriched in the South Korean and Japanese populations by chance, especially in the Japanese population. One limitation of our estimation method was the utilization of the GEVA software for indel variants. However, the fact that changing the mutation rate in the 10,000-fold range resulted in nearly identical estimates suggests that the estimates did not deviate significantly from the true variant age.

Our analysis of the c.1617del heterogeneous population showed no significant differences in blood test values for renal function compared with the wild-type population, and none of the individuals had CHD. To the best of our knowledge, no studies have analyzed large TMEM260 variant carrier populations. Further comprehensive analyses are required to evaluate the effects of heterozygous TMEM260 variants.

In conclusion, we identified a prevalent loss-of-function variant, c.1617del in TMEM260, among Japanese TA patients. Genetic testing for TMEM260, including the c.1617del variant, should be considered in Japanese TA patients lacking 22q11.2 deletion. Simultaneously, researchers at Keio University independently reported the same c.1617del variant of TMEM260 as a major genetic cause of TA in Japan, resulting in the same conclusion as that of our study (Prof. H. Yamagishi at Keio University and A. Kikuchi, personal communication). We propose to name the c.1617del variant as "Keio-Tohoku variant of TMEM260" considering its importance for the Japanese population.

REFERENCES

- 1. Zaidi S, Brueckner M. Genetics and genomics of congenital heart disease. Circ Res. 2017;120:923–40.
- 2. Naimo PS, Konstantinov IE. Surgery for truncus arteriosus: contemporary practice. Ann Thorac Surg. 2021;111:1442–50.
- 3. Pierpont ME, Brueckner M, Chung WK, Garg V, Lacro RV, McGuire AL, et al. Genetic basis for congenital heart disease: revisited: a scientific statement from the american heart association. Circulation. 2018;138:e653–711.
- 4. Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S, et al. Role of TBX1 in human del22q11.2 syndrome. Lancet. 2003;362:1366–73.
- 5. Su W, Zhu P, Wang R, Wu Q, Wang M, Zhang X, et al. Congenital heart diseases and their association with the variant distribution features on susceptibility genes. Clin Genet. 2017;91:349–54.
- 6. Ta-Shma A, El-lahham N, Edvardson S, Stepensky P, Nir A, Perles Z, et al. Conotruncal malformations and absent thymus due to a deleterious NKX2-6 mutation. J Med Genet. 2014;51:268–70.
- 7. Kodo K, Nishizawa T, Furutani M, Arai S, Yamamura E, Joo K, et al. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorinplexin signaling. Proc Natl Acad Sci USA. 2009;106:13933–8.
- 8. Larsen ISB, Povolo L, Zhou L, Tian W, Mygind KJ, Hintze J, et al. The SHDRA syndrome-associated gene TMEM260 encodes a protein-specific O-mannosyltransferase. Proc Natl Acad Sci USA. 2023;120:e2302584120.
- 9. Ta-Shma A, Khan TN, Vivante A, Willer JR, Matak P, Jalas C, et al. Mutations in TMEM260 cause a pediatric neurodevelopmental, cardiac, and renal syndrome. Am J Hum Genet. 2017;100:666–75.
- 10. Kuroda Y, Saito Y, Enomoto Y, Naruto T, Mitsui J, Kurosawa K. PHACES-like syndrome with TMEM260 compound heterozygous variants. Am J Med Genet A. 2023. <https://doi.org/10.1002/ajmg.a.63245>.
- 11. Pagnamenta AT, Jackson A, Perveen R, Beaman G, Petts G, Gupta A, et al. Biallelic TMEM260 variants cause truncus arteriosus, with or without renal defects. Clin Genet. 2022;101:127–33.
- 12. Peng M, Jing S, Duan S, Lu G, Zhou K, Hua Y, et al. A novel homozygous variant of TMEM260 induced cardiac malformation and neurodevelopmental abnormality: case report and literature review. Front Med. 2023;10:1157042.
- 13. Saito-Hakoda A, Kikuchi A, Takahashi T, Yokoyama Y, Himori N, Adachi M, et al. Familial Paget's disease of bone with ocular manifestations and a novel TNFRSF11A duplication variant (72dup27). J Bone Min Metab. 2022. [https://](https://doi.org/10.1007/s00774-022-01392-w) [doi.org/10.1007/s00774-022-01392-w.](https://doi.org/10.1007/s00774-022-01392-w) Available from
- 14. Takayama J, Tadaka S, Yano K, Katsuoka F, Gocho C, Funayama T, et al. Construction and integration of three de novo Japanese human genome assemblies toward a population-specific reference. Nat Commun. 2021;12:1–14.
- 15. Hui-Yuen J, McAllister S, Koganti S, Hill E, Bhaduri-McIntosh S. Establishment of Epstein-Barr virus growth-transformed lymphoblastoid cell lines. J Vis Exp. 2011. [https://doi.org/10.3791/3321.](https://doi.org/10.3791/3321)
- 16. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
- 17. Tadaka S, Hishinuma E, Komaki S, Motoike IN, Kawashima J, Saigusa D, et al. jMorp updates in 2020: large enhancement of multi-omics data resources on the general Japanese population. Nucleic Acids Res. 2020;49:D536–44.
- 18. Cao Y, Li L, Xu M, Feng Z, Sun X, Lu J, et al. The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals. Cell Res. 2020;30:717–31.
- 19. Feng Y-CA, Chen C-Y, Chen T-T, Kuo P-H, Hsu Y-H, Yang H-I, et al. Taiwan Biobank: a rich biomedical research database of the Taiwanese population. Cell Genom. 2022;2:100197.
- 20. Lee J, Lee J, Jeon S, Lee J, Jang I, Yang JO, et al. A database of 5305 healthy Korean individuals reveals genetic and clinical implications for an East Asian population. Exp Mol Med. 2022;54:1862–71.
- 21. Shotelersuk V, Wichadakul D, Ngamphiw C, Srichomthong C, Phokaew C, Wilantho A, et al. The Thai reference exome (T-REx) variant database. Clin Genet. 2021;100:703–12.
- 22. Albers PK, McVean G. Dating genomic variants and shared ancestry in population-scale sequencing data. PLoS Biol. 2020;18:e3000586.
- 23. Tadaka S, Katsuoka F, Ueki M, Makino S, Saito S, Otsuki A, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. Hum Genome Var. 2019;6:1–9.
- 24. Eyre-Walker A, Keightley PD. High genomic deleterious mutation rates in hominids. Nature. 1999;397:344–7.
- 25. Yu N, Zhao Z, Fu YX, Sambuughin N, Ramsay M, Jenkins T, et al. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol Biol Evol. 2001;18:214–22.
- 26. Ophir R, Graur D. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. Gene. 1997;205:191–202.
- 27. Sjödin P, Bataillon T, Schierup MH. Insertion and deletion processes in recent human history. PLoS One. 2010;5:e8650.
- 28. Allen HL, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, et al. GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet. 2011;44:20–2.
- 29. Page DJ, Miossec MJ, Williams SG, Monaghan RM, Fotiou E, Cordell HJ, et al. Whole exome sequencing reveals the major genetic contributors to nonsyndromic tetralogy of fallot. Circ Res. 2019;124:553–63.
- 30. Škorić-Milosavljević D, Tjong FVY, Barc J, Backx APCM, Clur S-AB, van Spaendonck-Zwarts K, et al. GATA6 mutations: Characterization of two novel patients and a comprehensive overview of the GATA6 genotypic and phenotypic spectrum. Am J Med Genet A. 2019;179:1836–45.

ACKNOWLEDGEMENTS

We thank the patients and their families for participating in this study. We appreciate all the volunteers who participated in the Tohoku Medical Megabank project. We thank Yoko Chiba and Kumi Ito for technical assistance. We also thank the Biomedical Research Core of the Tohoku University Graduate School of Medicine and the Biomedical Research Unit of Tohoku University Hospital. We would like to thank Editage [\(www.editage.jp\)](http://www.editage.jp) for English language editing.

FUNDING

This research was supported in part by the Japan Agency for Medical Research and Development, AMED under Grant Number JP21tm0124005 and JP23tm0424402. Drs. Takayama, Tamiya, Kure, and Kikuchi have received a research grant from Astellas Pharma.

COMPETING INTERESTS

Some of the authors declared Financial and Non-Financial Relationships and Activities, and Conflicts of Interest regarding this manuscript as indicated in the supplementary materials. The sponsor had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at [https://doi.org/10.1038/s10038-024-01223-y.](https://doi.org/10.1038/s10038-024-01223-y)

Correspondence and requests for materials should be addressed to Hisao Yaoita or Atsuo Kikuchi.

Reprints and permission information is available at [http://www.nature.com/](http://www.nature.com/reprints) [reprints](http://www.nature.com/reprints)

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

Open Access This article is licensed under a Creative Commons \bigcirc $| (cc)$ Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit [http://](http://creativecommons.org/licenses/by/4.0/) [creativecommons.org/licenses/by/4.0/.](http://creativecommons.org/licenses/by/4.0/)

© The Author(s) 2024, corrected publication 2024