Journal Pre-proof

Genome Sequencing Reveals the Prevalence of Fabry Disease–Causing Variants in Japan: Evidence from the Tohoku Medical Megabank

Kai Susukita, MD, Kotaro Nochioka, MD PhD MPH, Masato Takase, PhD, Saori Yamamoto, MD PhD, Kazuki Kumada, PhD, Hideaki Suzuki, MD PhD, Kinuko Ohneda, MD PhD, Kenichi Noguchi, PhD, Yoko Izumi, MD PhD, Soichi Ogishima, PhD, Naoki Nakaya, PhD, Yoshikatsu Eto, MD PhD, Atushi Hozawa, MD PhD, Satoshi Yasuda, MD PhD

PII: S2949-7744(26)01000-9

DOI: https://doi.org/10.1016/j.gimo.2026.104490

Reference: GIMO 104490

To appear in: *Genetics in Medicine Open*

Received Date: 4 April 2026

Revised Date: 22 July 2026

Accepted Date: 29 July 2026

Please cite this article as: Susukita K, Nochioka K, Takase M, Yamamoto S, Kumada K, Suzuki H, Ohneda K, Noguchi K, Izumi Y, Ogishima S, Nakaya N, Eto Y, Hozawa A, Yasuda S, Genome Sequencing Reveals the Prevalence of Fabry Disease–Causing Variants in Japan: Evidence from the Tohoku Medical Megabank, *Genetics in Medicine Open* (2026), doi: https://doi.org/10.1016/j.gimo.2026.104490.

Author detail

Genome Sequencing Reveals the Prevalence of Fabry Disease–Causing Variants in Japan: Evidence from the Tohoku Medical Megabank

Kai Susukita MD[a], Kotaro Nochioka MD PhD MPH[a], Masato Takase PhD[b], Saori Yamamoto MD PhD[a], Kazuki Kumada PhD[b], Hideaki Suzuki MD PhD[a], Kinuko Ohneda MD PhD[b], Kenichi Noguchi PhD[b], Yoko Izumi MD PhD[b], Soichi Ogishima PhD[b], Naoki Nakaya PhD[b], Yoshikatsu Eto MD PhD[c], Atushi Hozawa MD PhD[b], Satoshi Yasuda MD PhD[a]

Affiliations:

[a] Department of Cardiovascular Medicine, Tohoku University Graduate School of Medicine, Japan

[b] Tohoku Medical Megabank Organization, Tohoku University.

[c] Advanced Clinical Research Center and Asian LSD Center, Institute of Neurological disorders, Fukushima, Japan.

Address for Correspondence: Kotaro Nochioka, MD MPH, PhD.

Division of Cardiovascular Medicine, Tohoku University Hospital, Seiryo-machi 1-1, Aobaku, Sendai, Miyagi 9808574 Japan. Fax: +8122-717-7156, Tel: +8122-717-7153,

email: kotaro.nochioka.d5@tohoku.ac.jp

Conflict of Interest Statement: None of the co-authors has a conflict of interest with regard to the content of this manuscript.

Author Contributions:

Conceptualization: KN, KN, YI**; Data curation:** KS, KN, MT, KK, SO, AH**; Formal analysis:** KS, KN, MT**; Funding acquisition:** KN, SY; **Investigation:** KS, KN, MT, KK, KO ,SO, NN, AH**; Methodology:** KS, KN, MT, KK, KN, YI, AH**; Project administration:** KS, KN**; Resources:** KN, MT, KK, KO, KN, YI, SO, NN, AH**; Software:** KS, KN**; Supervision:** KO, KN, YI, SO, NN, YE, AH, SY**; Validation:** KS, KN**; Visualization:** KS, KN **;Writing-original draft:** KS, KN, MT, SY**; Writing-review & editing:** KS, KN, MT, SY, KK, HS, KO, KN, YI, SO, NN, YE, AH, SY.

Data Availability Statement: The authors will make available relevant survey data for the purpose of verifying or contextualizing the conclusions. Disclosure will be limited by the need to protect the privacy of the respondents. In addition, the authors will consider requests to supply relevant, unpublished data for the purpose of verifying or contextualizing the conclusions.

Ethics Declaration: This study was approved by the Institutional Review Board of the Tohoku Medical Megabank Organization (approval number: 2024-4-087). All participants provided written informed consent.

Funding: This work was supported as part of the authors' academic endeavors. No funding source was used to support this work.

Acknowledgement: This study used whole-genome sequencing data obtained through the Consortium for integrated analysis of genome, medical and health information. We thank Dr Satoshi Ishii (Glyco Pharma Corporation, Oita, Japan) for measuring serum GLA activity and lyso-Gb3 levels. During the preparation of this work, the author (KN) used ChatGPT (OpenAI) for English language editing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

GIMO-D-26-00069R1

Genome Sequencing Reveals the Prevalence of Fabry Disease–Causing Variants in Japan: Evidence from the Tohoku Medical Megabank

Introduction

Fabry disease (FD) is a rare X-linked lysosomal storage disorder caused by pathogenic variants in the GLA gene, resulting in impaired activity of α-galactosidase A.(1) This enzymatic deficiency leads to progressive accumulation of globotriaosylceramide and related metabolites in vascular endothelial cells and multiple organs, including the heart, kidneys, and nervous system.(1, 2) The prevalence of Fabry disease in the general population has traditionally been estimated at approximately 1 in 40,000–140,000 individuals.(1, 3) Although newborn screening programs, particularly in Taiwan, have provided well-defined estimates of Fabry disease prevalence and have revealed a substantially higher frequency of late-onset variants, population-based genomic data remain limited in other Asian populations. (1, 4) This study aimed to determine the prevalence of Fabry disease–causing variants in genome sequencing data from a community-based cohort in the Tohoku Medical Megabank, Japan.

Methods

The Tohoku Medical Megabank (TMM) Project aims to integrate population genomics, medical genetics, and prospective cohort studies to establish a critical infrastructure for the advancement of precision medicine. A key strength of the TMM Project is its linkage of genomic data with detailed health questionnaires, biospecimens, and

longitudinal clinical information in a large community-based population. The TMM Community Cohort Study is a general population–based prospective cohort comprising individuals aged ≥20 years from Miyagi Prefecture in northeastern Japan. The design, methods, and other details of this study have been described elsewhere.(5, 6) Briefly, initiated in May 2013, the study enrolled more than 50,000 participants through March 2016. The Type 1 survey was conducted mainly at municipal specific health check-up sites (n = 40,433), with additional survey sessions held on separate dates to accommodate overflow at venue sites (n = 664). In contrast, the Type 2 survey was conducted at designated Community Support Centers (n = 13,855), where participants voluntarily underwent more detailed physiological and biochemical examinations and biological specimen collection.(5, 6) Participants provided lifestyle and other health-related information through self-administered questionnaires and submitted blood and urine samples.

We used the quality-controlled, jointly genotyped genome sequencing multisample VCF call set provided by the Tohoku Medical Megabank (ToMMo) Project (69KJPN), aligned to the GRCh38 reference genome build.(7) This publicly available resource comprises large-scale population-based genomic data generated under a standardized sequencing and quality-control framework, enabling genotype-first assessment of rare disease–associated variants in the general population. Variant calling and quality control, including read alignment, joint genotyping, and variant-level filtering, were performed by ToMMo using a GATK Best Practices–based pipeline as described previously.(7-9) We did not perform re-alignment or variant re-calling from raw sequencing reads (FASTQ/BAM); all downstream analyses were conducted using

the released VCF call set. Variants in the Fabry disease gene (GLA, located on the X chromosome) were extracted from the ToMMo GRCh38 call set and annotated using ClinVar (release 2023-02-26).(10) Variant pathogenicity was classified according to ClinVar clinical significance, with variants labeled as Pathogenic or Likely pathogenic considered pathogenic and those labeled as Benign or Likely benign considered non-pathogenic. Variants classified as Uncertain significance or Conflicting interpretations were excluded from the primary pathogenic definition and evaluated separately where relevant. Because GLA is located on the X chromosome, genotype interpretation accounted for sex, with males treated as hemizygous and females as heterozygous or homozygous as applicable.

Results

Among 68,615 individuals (mean age, 47.7 ± 19.5 years; 66% female), four participants with pathogenic GLA variants (mean age, 49.3 ± 15.9 years) and one participant aged 70–79 years with a likely pathogenic variant were identified, corresponding to a prevalence of 1 in 13,723; all were female (**Table 1**). All pathogenic variants were c.902G>A, whereas the likely pathogenic variant was c.1076T>C. Among the five heterozygotes, renal function data were available for four, with a mean estimated glomerular filtration rate (eGFR) of 102.3 ± 21.2 mL/min/1.73 m². Microalbuminuria was observed in one participant (Case 3). Electrocardiographic data were available for three heterozygotes, and no left ventricular hypertrophy was identified based on standard voltage criteria (RV5 ≥ 26 mm or RV5 + SV1 ≥ 35 mm). Heterozygotes of

pathogenic GLA variants had a mean plasma GLA activity of 5.64 ± 1.73 mmol/hr/mL, whereas the heterozygote of the likely pathogenic variant had a plasma GLA activity of 8.36 mmol/hr/mL. Plasma lyso-Gb3 levels were 3.7 ± 0.79 ng/mL in pathogenic variants and 0.96 ng/mL in the likely pathogenic variant. (**Table 1**) The identified variants of uncertain significance (n = 17; 1 in 4,036) included p.Ala15Thr (n = 8), p.Asp153Asn (n = 6), p.Gln212Arg (n = 2), and p.Asn419Asp (n = 1). No variants with conflicting interpretations were identified in this analysis.

Discussion

In this study, genome sequencing in a large Japanese community cohort from the TMM Project identified pathogenic or likely pathogenic GLA variants in 1 of every 13,723 individuals, indicating a higher prevalence of Fabry disease–associated variants than historically estimated from clinically ascertained populations.(1) All identified heterozygotes were female, and all pathogenic variants were c.902G>A,(11) suggesting that late-onset Fabry disease, particularly among women, may remain underrecognized in the general population. Historically, the prevalence of Fabry disease has been estimated at 1 in 40,000 to 1 in 170,000 based on clinical diagnosis,(1) whereas population-based genomic data from the UK Biobank showed a substantially higher prevalence of 1 in 5,573.(12) Although the prevalence observed in our cohort was lower than that reported in the UK Biobank, this difference may reflect ethnic variation in GLA variant distribution and differences in cohort composition between European and Japanese population-based studies. Nevertheless, both studies consistently suggest that

Fabry disease–causing variants are more common than previously recognized.(3) Newborn screening studies similarly report prevalences ranging from approximately 1 in 1,600 to 1 in 11,854, largely driven by late-onset variants.(4, 13-16) Together, these findings support the presence of a substantial reservoir of undiagnosed Fabry disease in the general population.

Plasma lyso-Gb3, a disease-related metabolite and potential biomarker of therapeutic response, exceeded the upper limit of normal (≤2.0 ng/mL) (17) in all four participants with c.902G>A, supporting biochemical relevance despite the absence of overt clinical manifestations. We previously reported marked clinical heterogeneity among patients with c.902G>A in the same region, (2) and its repeated detection in the present cohort may reflect regional enrichment of this late-onset phenotype.

No hemizygous males with pathogenic GLA variants were identified. One possible explanation is that males with clinically overt disease may have already been diagnosed and therefore were less likely to participate in this community-based cohort. We also identified one likely pathogenic variant (Case 5, c.1076T>C), previously reported in Fabry disease but with conflicting evidence regarding pathogenicity.(18, 19) In our cohort, the heterozygote showed normal lyso-Gb3 levels, no left ventricular hypertrophy, and preserved renal function without microalbuminuria, suggesting limited phenotypic impact of this variant. These findings support the value of longitudinal follow-up in population-based biobanks to clarify the clinical penetrance of Fabry disease–associated variants, particularly in women with late-onset genotypes. Information on family history and cascade screening of relatives was not available in the present biobank-based analysis. Therefore, we could not determine how many

family members were found to be at risk or newly diagnosed through the identification of these probands. Future linkage with clinical follow-up and family-based evaluations may help clarify the broader diagnostic impact of genotype-first screening.

In conclusion, population-based genome sequencing in the TMM revealed that pathogenic or likely pathogenic Fabry disease variants are more prevalent than previously recognized, indicating that late-onset Fabry disease in women may remain substantially underdiagnosed.

Reference

1. Germain DP. Fabry disease. Orphanet J Rare Dis. 2010;5:30.

2. Yamamoto S, Nagasawa T, Sugimura K, Kanno A, Tatebe S, Aoki T, et al. Clinical Diversity in Patients with Anderson-fabry Disease with the R301Q Mutation. Intern Med. 2019;58(4):603-7.

3. Biegstraaten M, Arngrimsson R, Barbey F, Boks L, Cecchi F, Deegan PB, et al. Recommendations for initiation and cessation of enzyme replacement therapy in patients with Fabry disease: the European Fabry Working Group consensus document. Orphanet J Rare Dis. 2015;10:36.

4. Hwu WL, Chien YH, Lee NC, Chiang SC, Dobrovolny R, Huang AC, et al. Newborn screening for Fabry disease in Taiwan reveals a high incidence of the later-onset GLA mutation c.936+919G>A (IVS4+919G>A). Hum Mutat. 2009;30(10):1397-405.

5. Yasuda J, Katsuoka F, Danjoh I, Kawai Y, Kojima K, Nagasaki M, et al. Regional genetic differences among Japanese populations and performance of genotype imputation using whole-genome reference panel of the Tohoku Medical Megabank Project. BMC Genomics. 2018;19(1):551.

6. Hozawa A, Tanno K, Nakaya N, Nakamura T, Tsuchiya N, Hirata T, et al. Study Profile of the Tohoku Medical Megabank Community-Based Cohort Study. J Epidemiol. 2021;31(1):65-76.

7. Tadaka S, Katsuoka F, Ueki M, Kojima K, Makino S, Saito S, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. Hum Genome Var. 2019;6:28.

8. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.

9. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.

10. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.

11. Palhais B, Dembic M, Sabaratnam R, Nielsen KS, Doktor TK, Bruun GH, et al. The prevalent deep intronic c. 639+919 G>A GLA mutation causes pseudoexon activation and Fabry disease by abolishing the binding of hnRNPA1 and hnRNP A2/B1 to a splicing silencer. Mol Genet Metab. 2016;119(3):258-69.

12. Gilchrist M, Casanova F, Tyrrell JS, Cannon S, Wood AR, Fife N, et al. Prevalence of Fabry disease-causing variants in the UK Biobank. J Med Genet. 2023;60(4):391-6.

13. Burton BK, Charrow J, Hoganson GE, Waggoner D, Tinkle B, Braddock SR, et al. Newborn Screening for Lysosomal Storage Disorders in Illinois: The Initial 15-Month Experience. J Pediatr. 2017;190:130-5.

14. Sawada T, Kido J, Yoshida S, Sugawara K, Momosaki K, Inoue T, et al. Newborn screening for Fabry disease in the western region of Japan. Mol Genet Metab Rep. 2020;22:100562.

15. Spada M, Pagliardini S, Yasuda M, Tukel T, Thiagarajan G, Sakuraba H, et al. High incidence of later-onset fabry disease revealed by newborn screening. Am J Hum Genet. 2006;79(1):31-40.

16. Sawada T, Kido J, Tsukimura T, Sugawara K, Shiga T, Saito S, et al. Evaluation of GLA variants detected in newborn screening for Fabry disease using biomarker analysis. Mol Genet Metab Rep. 2025;44:101245.

17. Maruyama H, Miyata K, Mikame M, Taguchi A, Guili C, Shimura M, et al. Effectiveness of plasma lyso-Gb3 as a biomarker for selecting high-risk patients with Fabry disease from multispecialty clinics for genetic analysis. Genet Med. 2019;21(1):44-52.

18. Kobayashi M, Ohashi T, Kaneshiro E, Higuchi T, Ida H. Mutation spectrum of alpha-Galactosidase gene in Japanese patients with Fabry disease. J Hum Genet. 2019;64(7):695-9.

19. Stiles AR, Zhang H, Dai J, McCaw P, Beasley J, Rehder C, et al. A comprehensive testing algorithm for the diagnosis of Fabry disease in males and females. Mol Genet Metab. 2020;130(3):209-14.

Table 1. Clinical characteristics of 5 participants with Fabry disease-causing variants

Case	Age (years)	Sex	dbSNP ID	Variant	Protein change	Disease type	Serum GLA activity (mmol/h/mL)	Serum Lyso GB3 (ng/mL)	LVH	eGFR (mL/min/1.73m^2)	Urinary albumin (mg/gCr)
1	30s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	5.64	2.67[a]	-	NA	NA
2	40s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	6.92	4.60[a]	-	132.9	22.2
3	50s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	3.20	3.82[a]	NA	114.2	48.6
4	70s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	6.80	3.71[a]	NA	88.1	15.1
5	70s	Female	rs1928132914	c.1076T>C	p.Ile359Thr	Likely-pathogenic	8.36	0.96	-	74.1	6.8

LVH was defined according to electrocardiographic voltage criteria as RV5 + SV1 ≥ 35 mm or RV5 ≥ 26 mm.

[a] Exceeded the upper limit of normal (normal upper limit≤2.0ng/mL).

Table 1. Clinical characteristics of 5 participants with Fabry disease-causing variants

Case	Age (years)	Sex	dbSNP ID	Variant	Protein change	Disease type	Serum GLA activity (mmol/h/mL)	Serum Lyso GB3 (ng/mL)	LVH	eGFR (mL/min/1.73m^2)	Urinary albumin (mg/gCr)
1	30s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	5.64	2.67 *a*	-	NA	NA
2	40s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	6.92	4.60 *a*	-	132.9	22.2
3	50s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	3.20	3.82 *a*	NA	114.2	48.6
4	70s	Female	rs104894828	c.902G>A	p.Arg301Gln	Late-onset	6.80	3.71 *a*	NA	88.1	15.1
5	70s	Female	rs1928132914	c.1076T>C	p.Ile359Thr	Likely-pathogenic	8.36	0.96	-	74.1	6.8

LVH was defined according to electrocardiographic voltage criteria as RV5 + SV1 ≥ 35 mm or RV5 ≥ 26 mm.

a Exceeded the upper limit of normal (normal upper limit≤2.0ng/mL).