Received: 21 November 2022 | Accepted: 12 January 2023

DOI: 10.1002/rmb2.12504

Reproductive Medicine and Biology

WILEY

ORIGINAL ARTICLE

Exome-based genome-wide screening of rare variants associated with the risk of polycystic ovary syndrome

Satoshi Tamaoka[1] | Kazuki Saito[1,2] | Tomoko Yoshida[1] | Kazuhiko Nakabayashi[3] | Kenichi Tatsumi[4] | Toshihiro Kawamura[5] | Toshiya Matsuzaki[6] | Keiko Matsubara[1] | Hiroko Ogata-Kawata[3] | Kenichiro Hata[3] | Yuko Kato-Fukui[1] | Maki Fukami[1]

[1]Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan

[2]Department of Perinatal and Maternal Medicine (Ibaraki), Graduate School, Tokyo Medical and Dental University, Tokyo, Japan

[3]Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan

[4]Umegaoka Women's Clinic, Tokyo, Japan

[5]Denentoshi Ladies Clinic, Kanagawa, Japan

[6]Department of Obstetrics and Gynecology, Yoshinogawa Medical Center, Tokushima, Japan

Correspondence
Maki Fukami, Department of Molecular Endocrinology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya, Tokyo 157-8535, Japan.
Email: fukami-m@ncchd.go.jp

Funding information
Japan Agency for Medical Research and Development, Grant/Award Number: 22ek0109464h0003; Japan Society for the Promotion of Science, Grant/Award Number: 20H00539 and 21K19283; National Center for Child Health and Development, Grant/Award Number: 2022A-1; Takeda Science Foundation

Abstract

Purpose: Genetic factors associated with the risk of polycystic ovary syndrome (PCOS) remain largely unknown. Here, we conducted an optimal sequence kernel association test (SKAT-O), an exome-based rare variant association study, to clarify whether rare variants in specific genes contribute to the development of PCOS.

Methods: SKAT-O was performed using exome data of 44 Japanese patients with PCOS and 301 control women. We analyzed frequencies of rare probably damaging variants in the genome.

Results: Rare variants of *GSTO2* were more commonly identified in the patient group than in the control group (6/44 vs. 1/301; Bonferroni-corrected *p*-value, 0.028), while the frequencies of variants in other genes were comparable between the two groups. The identified *GSTO2* variants were predicted to affect the function, structure, stability, hydrophobicity, and/or the formation of intrinsically disordered regions of the protein. *GSTO2* encodes a glutathione transferase that mediates the oxidative stress response and arsenic metabolism. Previously, common variants in *GSTO2* and its paralog *GSTO1* were associated with the risk of PCOS.

Conclusions: The results indicate that there are no genes whose rare variants account for a large fraction of the etiology of PCOS, although rare damaging variants in *GSTO2* may constitute a risk factor in some cases.

KEYWORDS
association test, exome sequencing, PCOS, rare variant, SNP

1 | INTRODUCTION

Polycystic ovary syndrome (PCOS) is a common multifactorial disorder characterized by anovulation, hyperandrogenism, and ovarian cysts.[1] The underlying mechanism of PCOS is not fully understood; although previous studies have suggested that various cellular events, such as chronic inflammation, oxidative stress, and endoplasmic reticulum (ER) stress, contribute to this condition.[2,3] In addition, increased serum arsenic levels were observed in patients with PCOS.[4] To date, efforts were made to identify genetic factors

associated with the risk of PCOS. Candidate gene approaches identified several single-nucleotide polymorphisms (SNPs) and repeat-number polymorphisms in various genes, including *GSTO1*, *CYP19A1*, and *DENND1A*, whose frequencies were significantly higher in patients with PCOS than in unaffected women.[5–12] Furthermore, genome-wide association studies identified 46 relatively common SNPs that are possibly associated with the risk of PCOS.[10,13–21] However, these known risk variants are likely to account for only a small fraction of the genetic heritability of PCOS.[10]

In the present study, we tested a hypothesis that rare damaging variants in specific genes constitute risk factors for PCOS. To this end, we conducted an optimal sequence kernel association test (SKAT-O), a gene-based rare variant association study that uses exome data of cases and controls.[22] SKAT-O is optimized for small-sized groups.[22] To date, SKAT-O has successfully identified risk genes for several multifactorial disorders including idiopathic scoliosis and chronic central serous chorioretinopathy.[23,24] In the present study, we performed SKAT-O for 44 patients with PCOS and 301 control women.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

This study was approved by the Institutional Review Board Committee at the National Center for Child Health and Development and performed after obtaining informed consent from all participants. This study was performed in accordance with the Declaration of Helsinki and relevant guidelines and regulations.

2.2 | Subjects

The patient group comprised 44 Japanese women with PCOS. All patients satisfied the Rotterdam criteria.[1] Clinical data of the patients are summarized in Table S1. The control group consisted of 301 healthy Japanese women aged between 19 and 49 years who had at least one child.

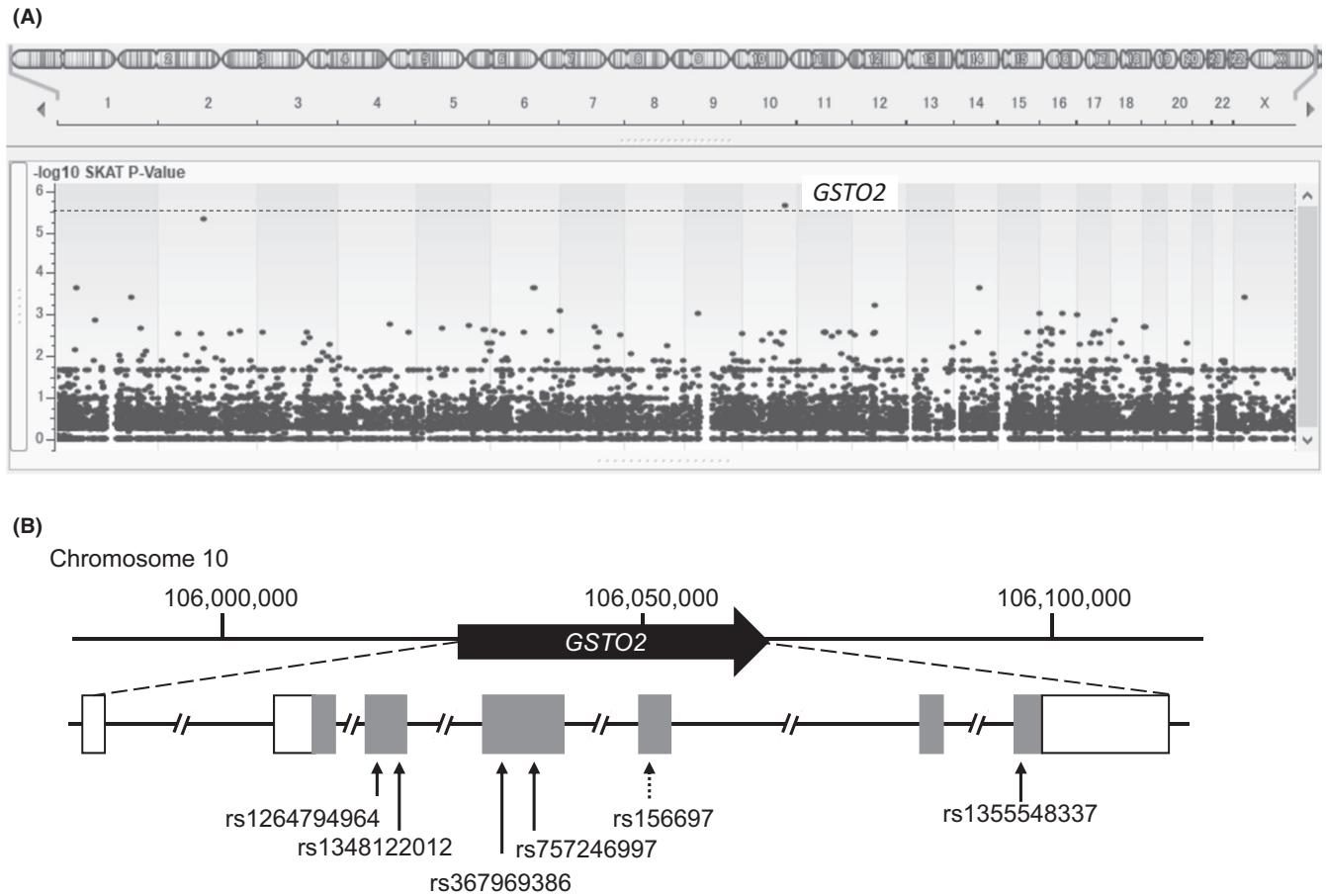

FIGURE 1 Results of molecular analysis. (A) Manhattan plot of the results of SKAT-O. The broken line depicts the Bonferroni-corrected *p*-value of 0.05. Only *GSTO2* on 10q25.1 showed a significant association with POCS. (B) Genomic structure of *GSTO2*. The thick black arrow indicates the genomic position of *GSTO2* on chromosome 10 (GRCh37/hg19). The white and gray boxes indicate the noncoding and coding exons, respectively. The PCOS-associated missense variants identified in the present and previous studies are indicated by thin solid and broken arrows, respectively.

2.3 | Exome sequencing and SKAT-O

Genomic DNA was extracted from the peripheral blood samples of the participants. DNA libraries were constructed using the SureSelect Human All Exon V6 kit (Agilent Technologies) and sequenced on NovaSeq or HiSeq sequencers (Illumina). VCF files were created using DRAGEN (version 3.9.5; Illumina). The sequence reads were mapped to the human reference genome (GRCh37/hg19 with decoy sequences [hs37d5]).

Exome data of the patients and control individuals were subjected to SKAT-O using the SNP and Variation Suite (version 8.4.1; Golden Helix). We focused on protein-altering variants (missense and nonsense variants, indels, and splice-site substitutions), whose allele frequency in the ToMMo database (version 8.3KJPN; https://www.megabank.tohoku.ac.jp/) was less than 5%.[25] All missense variants underwent in silico functional assessment; we selected variants that were assessed as damaging by three or more of the six programs in dbNSFP (version 3.0, http://database.liulab.science/dbNSFP/).

SKAT-O was carried out using the very-small-sample algorithm with the rho = 1 setting. We searched for genes whose rare variants were more commonly present in the patient group than in the control group. In addition, we examined whether rare variants of known PCOS-related genes accumulated in the patient group. Bonferroni-corrected *p*-values of <0.05 were considered statistically significant.

The effects of identified variants on protein function and structure were assessed using the combined annotation-dependent depletion program (CADD; https://cadd.gs.washington.edu/snv) and PyMOL (version 2.5, https://pymol.org/2/), respectively. CADD scores of ≥20 were assessed as probably damaging.[26] The protein IDs were obtained from the protein data bank (https://www.rcsb.org/). The effects of the variants on splice-site recognition were analyzed with Human Splicing Finder (https://hsf.genomnis.com/home), Alternative Splice-Site Predictor (http://wangcomputing.com/assp/), and NNSPLICE (http://fruitfly.org/seq_tools/splice.html), and the effects on the protein stability were predicted using I Mutant Suite (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi). In addition, the hydrophobicity of wildtype and variant GSTO2 proteins was assessed by using ProtScale (https://web.expasy.org/protscale/) with the Kyte and Doolittle model.[27] The formation of intrinsically disordered regions, a region without fixed 3-dimensional structures,[28] was predicted by the Predictor of natural disordered regions (PONDR; http://www.pondr.com/) using the VL-XT method.[29]

3 | RESULTS

Exome sequencing of the patients and control individuals identified 42 168 rare protein-altering variants of 11 961 genes, which were assessed as probably damaging by in silico analyses. SKAT-O for these variants revealed that, with the exception of *GSTO2*, there were no genes whose rare variants were significantly accumulated in the patient group (Figure 1). In particular, the frequencies of rare SNPs of

TABLE 1 Identified variants in *GSTO2*

Variant			Allele frequency		In silico functional assessment		Number of patients with the variant ($n = 44$)		Number of control individuals with the variant ($n = 301$)		Effect on splicing			Protein stability
Nucleotide	Amino acid	dbSNP	ToMMo[a]	gnomAD[b]	dbNSFP[c]	CADD[d]	Homozygote	Heterozygote	Homozygote	Heterozygote	HSF[e]	ASSP[f]	NNSPLICE[g]	I mutant suite[h]
46C>A	Pro16Thr	rs1264794964	4.2×10^{-4}	6.6×10^{-6}	5/6	26.7	0	1	0	0	Potential alteration	No effect	No effect	Decrease
136G>T	Asp46Tyr	rs1348122012	1.9×10^{-3}	6.6×10^{-6}	3/6	23.9	0	3	0	0	Potential alteration	No effect	No effect	Decrease
146A>T	His49Leu	rs367969386	1.6×10^{-3}	2.0×10^{-5}	5/6	25.8	0	1	0	0	Potential alteration	No effect	No effect	Neutral
230C>T	Thr77Ile	rs757246997	4.2×10^{-4}	–	4/6	17.8	0	1	0	0	No effect	No effect	No effect	Neutral
617C>T	Ala206Val	rs1355548337	4.8×10^{-4}	–	3/6	22.3	0	0	0	1	Potential alteration	No effect	No effect	Neutral

[a]ToMMo: Tohoku Medical Megabank Organization Database (https://www.megabank.tohoku.ac.jp/).

[b]gnomAD: Genome Aggregation Database (https://gnomad.broadinstitute.org/).

[c]dbNSFP: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site single-nucleotide variants (http://database.liulab.science/dbNSFP/). The number indicates the ratio of in silico programs, which predicted deleterious effects of the variant.

[d]CADD: Combined Annotation Dependent Depletion (https://cadd.gs.washington.edu/). Scores of ≥20 indicative of deleterious are underlined.

[e]HSF: Human Splicing Finder (http://www.umd.be/HSF3/).

[f]ASSP: Alternative Splice-Site Predictor (http://wangcomputing.com/assp/). The default threshold is 4.50.

[g]http://fruitfly.org/seq_tools/splice.html. The default threshold is 0.40.

[h]http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi.

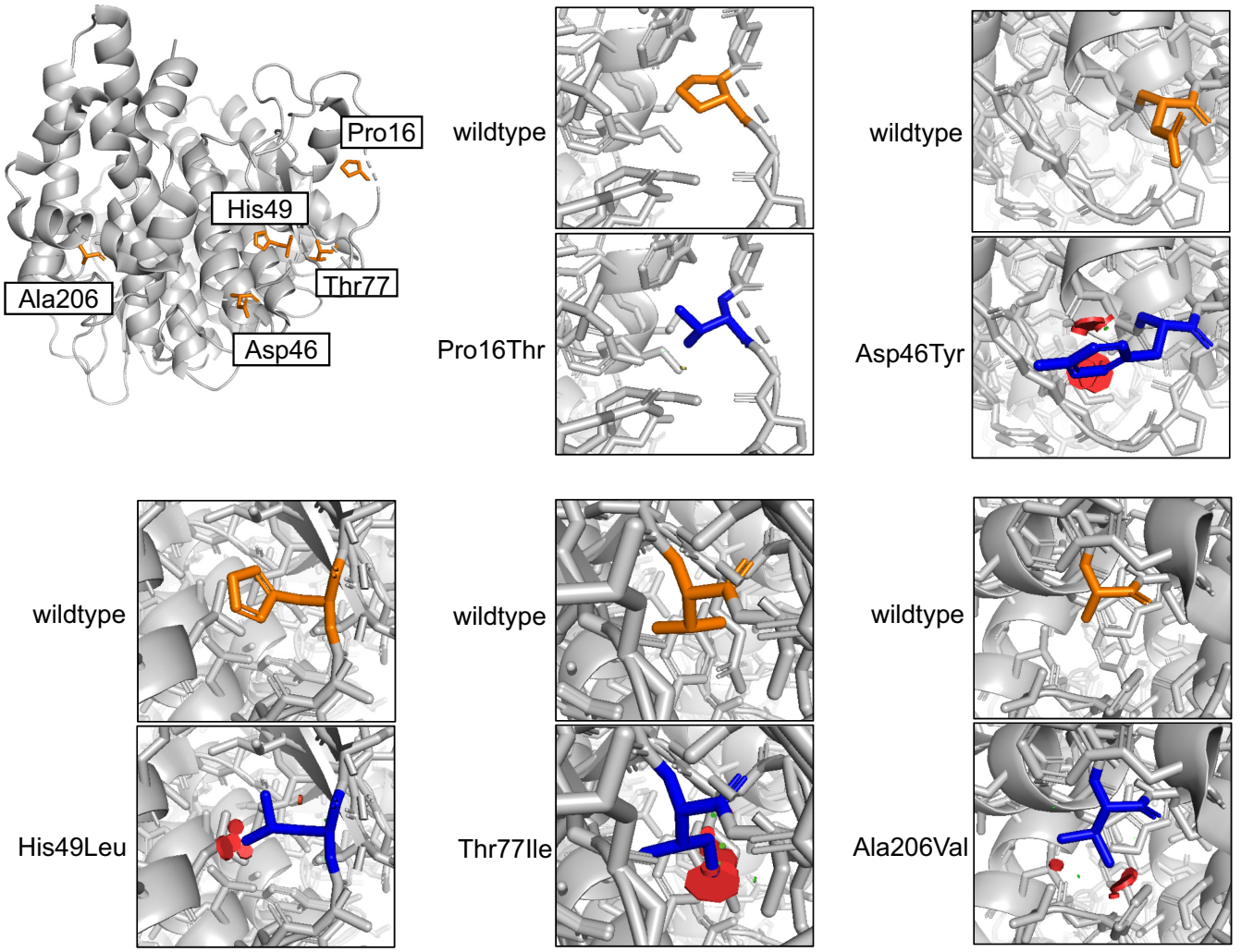

FIGURE 2 Predicted structures of GSTO2 variants. The protein structure of GSTO2 was obtained from the protein data bank (PDB ID: 3Q19). The most likely structure of each variant protein is shown. The wildtype and mutant amino acids are shown as orange and blue sticks, respectively. Red disks in the variant proteins indicate steric clashes.

the previously reported PCOS-related genes were comparable between the patient and control groups (Table S2).

Rare variants in *GSTO2* were more frequently identified in the patient group than in the control group (Bonferroni-corrected *p*-value, 0.028) (Figure 1). Specifically, four rare missense variants in *GSTO2* were shared by six of the 44 patients with PCOS (13.6%), while only one missense variant was present in one of the 301 control individuals (0.3%) (Table 1 and Table S1). The *GSTO2* variants were identified as monoallelic SNPs each in one individual. The five variants are known as very rare SNPs in the general population (Table 1). Four of the five variants were assessed as damaging by CADD. The five variants were predicted to alter the structure, splice recognition, stability, hydrophobicity, and/or the formation of intrinsically disordered regions of the GSTO2 protein (Table 1, Figures 2 and 3).

4 | DISCUSSION

The results of SKAT-O provide no evidence that specific genes except for *GSTO2* are associated with the risk of PCOS. In particular, rare variants in *CYP19A1*, *DENND1A*, and other known PCOS-related genes did not accumulate in the patient group. Our data argue against the notion that rare damaging SNPs of a few genes play a major role in the etiology of PCOS. These results may reflect the etiological heterogeneity of PCOS. Furthermore, environmental factors may significantly contribute to the development of PCOS. Alternatively, our negative results may be due to the small number of subjects. Although we utilized the SKAT-O algorithm optimized for small sample sizes,[22] the power of this study may not be sufficient to detect all rare risk variants. Moreover, because menstrual patterns and blood hormone levels of women in the control group were not examined, it remains possible that some of these women had mild PCOS.

SKAT-O showed a possible association between *GSTO2* and PCOS. Four rare missense variants of *GSTO2* were shared by six of 44 patients, whereas only one missense variant was identified in one of 301 control individuals. The identified variants were rare SNPs in the general population and were assessed as damaging by multiple in silico programs. In addition, these variants were likely to alter the structure, stability, hydrophobicity, and/or the formation

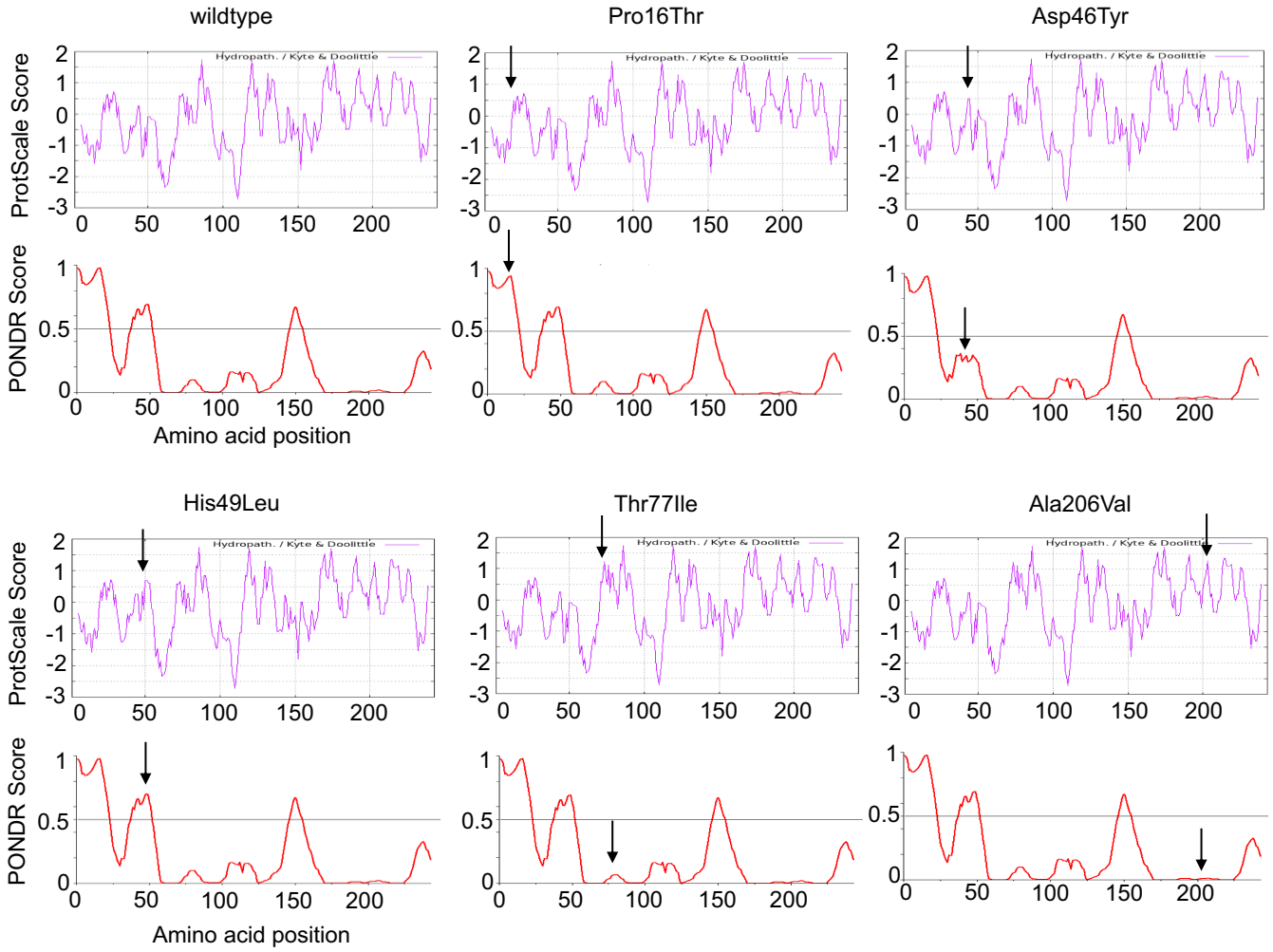

FIGURE 3 Hydrophobicity and intrinsically disordered regions of the wildtype and variant GSTO2 proteins. The upper and lower panels of each GSTO2 protein depict hydrophobicity plots and the Predictor of natural disordered regions (PONDR) scores, respectively. The black arrows indicate the position of the variants.

of intrinsically disordered regions of the GSTO2 protein. *GSTO2* is a broadly expressed gene that encodes omega-class glutathione S-transferase.[30] Miraghaee et al.[5] proposed that common SNPs in *GSTO2* and its paralog *GSTO1* increase the risk of PCOS in Iranian women. The authors found that women carrying the minor alleles of rs156697 in *GSTO2* (c.424A>T, p.Asn142Tyr) and rs4925 in *GSTO1* (c.419C>A, p.Ala140Asp) had a 1.53-fold higher risk of PCOS than women without these alleles. GSTO2 is known to be involved in the oxidative stress response and arsenic metabolism,[30,31] both of which have been linked to the mechanism of PCOS.[2,4] Since the dehydroascorbate reductase activity of GSTO2 is approximately 70–100-fold higher than that of GSTO1, GSTO2 may play a more protective role against oxidative stress than GSTO1.[32] Furthermore, because GSTO2 has been implicated in p38 phosphorylation,[33] which is essential for ER stress,[34] *GSTO2* variants may increase the risk of PCOS through aberrant ER stress. However, the clinical significance of the *GSTO2* variants identified in this study needs to be confirmed in future studies. Moreover, because rare *GSTO2* variants were shared only by 13.6% of our patients, the contribution of these variants to the entire etiology of PCOS appears to be limited. Our findings imply that rare SNPs of *GSTO2* are possibly associated with the risk of PCOS in a small percentage of the cases.

In conclusion, the results of SKAT-O indicate that rare damaging variants in *GSTO2*, but not in other genes in the genome, constitute a risk factor for PCOS. Our data await further validation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Satoshi Tamaoka https://orcid.org/0000-0001-7409-4423
Kazuhiko Nakabayashi https://orcid.org/0000-0003-2927-0963
Yuko Kato-Fukui https://orcid.org/0000-0001-6858-3693
Maki Fukami https://orcid.org/0000-0001-9971-4035

REFERENCES

1. Azziz R. Polycystic ovary syndrome. Obstet Gynecol. 2018;132(2):321–36.

2. Sadeghi HM, Adeli I, Calina D, Docea AO, Mousavi T, Daniali M, et al. Polycystic ovary syndrome: a comprehensive review of pathogenesis, management, and drug repurposing. Int J Mol Sci. 2022;23(2):583.
3. Wang C, Zhang Y. Endoplasmic reticulum stress: a new research direction for polycystic ovary syndrome? DNA Cell Biol. 2022;41(4):356–67.
4. Abudawood M, Tabassum H, Alanazi AH, Almusallam F, Aljaser F, Ali MN, et al. Antioxidant status in relation to heavy metals induced oxidative stress in patients with polycystic ovarian syndrome (PCOS). Sci Rep. 2021;11(1):22935.
5. Miraghaee SS, Sohrabi M, Jalili C, Bahrehmand F. Assessment of GSTO1 (A140D) and GSTO2 (N142D) gene polymorphisms in Iranian women with polycystic ovarian syndrome. Rep Biochem Mol Biol. 2020;9(1):8–13.
6. San Millan JL, Botella-Carretero JI, Alvarez-Blasco F, Luque-Ramirez M, Sancho J, Moghetti P, et al. A study of the hexose-6-phosphate dehydrogenase gene R453Q and 11beta-hydroxysteroid dehydrogenase type 1 gene 83557insA polymorphisms in the polycystic ovary syndrome. J Clin Endocrinol Metab. 2005;90(7):4157–62.
7. Panneerselvam P, Sivakumari K, Jayaprakash P, Srikanth R. SNP analysis of follistatin gene associated with polycystic ovarian syndrome. Adv Appl Bioinform Chem. 2010;3:111–9.
8. Khan MJ, Ullah A, Basit S. Genetic basis of polycystic ovary syndrome (PCOS): current perspectives. Appl Clin Genet. 2019;12:249–60.
9. Azziz R, Carmina E, Chen Z, Dunaif A, Laven JS, Legro RS, et al. Polycystic ovary syndrome. Nat Rev Dis Primers. 2016;2:16057.
10. Hiam D, Moreno-Asso A, Teede HJ, Laven JSE, Stepto NK, Moran LJ, et al. The genetics of polycystic ovary syndrome: an overview of candidate gene systematic reviews and genome-wide association studies. J Clin Med. 2019;8(10):1606.
11. Urbanek M. The genetics of the polycystic ovary syndrome. Nat Clin Pract Endocrinol Metab. 2007;3(2):103–11.
12. Saddick SY. Identifying genes associated with the development of human polycystic ovary syndrome. Saudi J Biol Sci. 2020;27(5):1271–9.
13. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet. 2012;44(9):1020–5.
14. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet. 2011;43(1):55–9.
15. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nat Commun. 2015;6:8464.
16. Day F, Karaderi T, Jones MR, Meun C, He C, Drong A, et al. Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. PLoS Genet. 2018;14(12):e1007813.
17. Zhang Y, Ho K, Keaton JM, Hartzel DN, Day F, Justice AE, et al. A genome-wide association study of polycystic ovary syndrome identified from electronic health records. Am J Obstet Gynecol. 2020;223(4):559.e1–21.
18. Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, et al. Leveraging Northern European population history: novel low-frequency variants for polycystic ovary syndrome. Hum Reprod. 2022;37(2):352–65.
19. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun. 2015;6:7502.
20. Lee H, Oh JY, Sung YA, Chung H, Kim HL, Kim GS, et al. Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. Hum Reprod. 2015;30(3):723–31.
21. Hwang JY, Lee EJ, Jin GM, Sung YA, Lee HJ, Heon KS, et al. Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. J Hum Genet. 2012;57(10):660–4.
22. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91(2):224–37.
23. Oliazadeh N, Gorman KF, Eveleigh R, Bourque G, Moreau A. Identification of elongated primary cilia with impaired mechanotransduction in idiopathic scoliosis patients. Sci Rep. 2017;7:44260.
24. Schellevis RL, Breukink MB, Gilissen C, Boon CJF, Hoyng CB, de Jong EK, et al. Exome sequencing in patients with chronic central serous chorioretinopathy. Sci Rep. 2019;9(1):6598.
25. Tadaka S, Saigusa D, Motoike IN, Inoue J, Aoki Y, Shirota M, et al. jMorp: Japanese multi omics reference panel. Nucleic Acids Res. 2018;46(D1):D551–7.
26. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(1):886–94.
27. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.
28. Van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014;114(13):6589–631.
29. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins. 2001;42(1):38–48.
30. Whitbread AK, Masoumi A, Tetlow N, Schmuck E, Coggan M, Board PG. Characterization of the omega class of glutathione transferases. Methods Enzymol. 2005;401:78–99.
31. Kim Y, Cha SJ, Choi HJ, Kim K. Omega class glutathione S-transferase: antioxidant enzyme in pathogenesis of neurodegenerative diseases. Oxid Med Cell Longev. 2017;2017:5049532.
32. Schmuck EM, Board PG, Whitbread AK, Tetlow N, Cavanaugh JA, Blackburn AC, et al. Characterization of the monomethylarsonate reductase and dehydroascorbate reductase activities of omega class glutathione transferase variants: implications for arsenic metabolism and the age-at-onset of Alzheimer's and Parkinson's diseases. Pharmacogenet Genomics. 2005;15(7):493–501.
33. Sumiya R, Terayama M, Hagiwara T, Nakata K, Sekihara K, Nagasaka S, et al. Loss of GSTO2 contributes to cell growth and mitochondria function via the p38 signaling in lung squamous cell carcinoma. Cancer Sci. 2022;113(1):195–204.
34. Jin J, Ma Y, Tong X, Yang W, Dai Y, Pan Y, et al. Metformin inhibits testosterone-induced endoplasmic reticulum stress in ovarian granulosa cells via inactivation of p38 MAPK. Hum Reprod. 2020;35(5):1145–58.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tamaoka S, Saito K, Yoshida T, Nakabayashi K, Tatsumi K, Kawamura T, et al. Exome-based genome-wide screening of rare variants associated with the risk of polycystic ovary syndrome. Reprod Med Biol. 2023;22:e12504. https://doi.org/10.1002/rmb2.12504