Journal of Assisted Reproduction and Genetics (2024) 41:135–146
https://doi.org/10.1007/s10815-023-02981-y

GENETICS

NOBOX gene variants in premature ovarian insufficiency: ethnicity-dependent insights

Pénélope Jordan[1] · Camille Verebi[1] · Sandrine Perol[2] · Sarah Grotto[2] · Corinne Fouveaut[1] · Sophie Christin-Maitre[3] · Aude Brac de la Perrière[4] · Virginie Grouthier[5] · Sophie Jonard-Catteau[6] · Philippe Touraine[7] · Geneviève PLU-Bureau[2] · Jean Michel Dupont[1] · Laila El Khattabi[1,8] · Thierry Bienvenu[1]

Received: 3 August 2023 / Accepted: 20 October 2023 / Published online: 3 November 2023

Abstract

Purpose Premature ovarian insufficiency (POI) affects approximately 1% of women before the age of 40. Genetic contribution is a significant component of POI. The *NOBOX* gene was considered one of the major genetic causes of POI. However, the pathogenicity and the penetrance of *NOBOX* variants remain unclear.

Methods We studied the whole coding region of the *NOBOX* gene by next generation sequencing in a cohort of 810 patients with POI, and we compared the frequency of each identified *NOBOX* variant to the general population taking into account the ethnicity of each individual.

Results Screening of the whole coding region of the *NOBOX* gene allowed us to identify 35 different variants, including 5 loss-of-function variants. In total, 171 patients with POI (25%) carried out at least one *NOBOX* variant. Regarding missense variants, we observed a significant overrepresentation of the most frequent ones in our 810 POI patients as compared to the general, except for p.(Arg117Trp). However, taking into account the ethnic origin of the individuals, we observed no significant OR difference for p.(Arg44Leu) and p.(Arg117Trp) in African subgroup and for p.(Asp452Asn) in European subgroup.

Conclusion This population study suggests that the p.(Arg44Leu) variant could be considered benign variant and that the p.(Asp452Asn) and p.(Arg117Trp) variants could be considered moderate risk pathogenic variants with probably partial and very low penetrance and/or expressivity. In contrast, p.(Gly91Trp) and p.(Gly152Arg) variants could be considered pathogenic variants with a moderate functional impact.

Keywords *NOBOX* · Premature ovarian insufficiency · Penetrance

Introduction

Premature ovarian insufficiency (POI) is a condition defined by the loss of ovarian function before the age of 40, characterized by disruptions in the menstrual cycle with elevated FSH (follicle-stimulating hormone) and low estradiol levels in blood [1]. It affects approximately 1% of women and can have profound implications on fertility. The *NOBOX* (*newborn ovary homeobox*) gene is considered one of the major genetic causes of POI [2, 3]. *NOBOX* is an ovary homeobox gene involved in first stages of folliculogenesis. In knockout *Nobox* female mice, Rajkovic et al. observed fibrous tissues replacing follicles, causing similar phenotype to non-syndromic ovarian failure [4]. NOBOX deficiency results in a defect in cyst breakdown that leads to abnormal connections between somatic cells and between oocytes [5, 6].

Mouse and human NOBOX share roughly 50% amino-acid sequence homology and present similar functional domains. Both proteins contain a homeodomain (amino-acids 272–363) required for DNA binding and co-factor interactions with a putative nuclear localization domain signal (within the homeodomain) responsible for the nuclear localization of NOBOX. They directly bind the growth and differentiation factor 9 (*GDF9)* and POU class5 homeobox 1 (*Pou5f*) promoters. Additionally, they contain a putative Src-homology 3 binding domain (amino-acids 273–280). The carboxy-terminal region of both proteins is proline-rich (amino-acids 392–668). NOBOX functions could be

Pénélope Jordan, Camille Verebi, Laila El Khattabi, and Thierry Bienvenu contributed equally to this work.

Extended author information available on the last page of the article

modulated by posttranslational modifications (such as sumoylation) [7].

Several molecular studies have been performed to evaluate the prevalence of *NOBOX* variants in patients diagnosed with POI and found *NOBOX* variants in 1.2–9% of patients with POI, suggesting that *NOBOX* variants appear to be the most frequent genetic cause of POI [8–11]. These variants can disrupt the normal functioning of NOBOX, leading to aberrations in folliculogenesis and ultimately, ovarian insufficiency. Interestingly, studies have shown that *NOBOX* variants can exhibit varying effects depending on the specific variant and its location within the gene. In 2007, Qin et al. (2007) discovered seven known single-nucleotide polymorphisms and four novel variants in 96 white women with POI [12]. They demonstrated that the p.(Arg355His) variant disrupted NOBOX homeodomain binding to NOBOX DNA-binding element and had a dominant negative effect on the binding of wild-type NOBOX to DNA [12]. In 2011, a French group sequenced *NOBOX* in a cohort of 178 women with idiopathic POI and identified 19 heterozygous variants including one nonsense (c.907C > T; p.(Arg303Ter)) and four missense (c.271G > T, p.(Gly91Trp); c.349C > T, p.(Arg117Trp); c.1025G > C, p.(Ser342Thr); c.1048G > T, p.(Val350Leu)) in 12 of these patients [2]. The authors demonstrated that these variants compromised the ability of the proteins to bind to and transactivate the *GDF9* promoter [2]. The same group analyzed 213 patients with POI and reported 3 novel and 2 recurrent heterozygous missense NOBOX rare variants found in 12 other patients [13]. Using a functional assay on the *GDF9* promoter and on *KIT-L*, they demonstrated that the missense variants (p.(Gly91Thr), p.(Gly111Arg), p.(Arg117Trp), p.(Lys371Thr), and p.(Pro619Leu)) were deleterious for protein function [13]. All these studies suggest that *NOBOX* heterozygous variant leads to premature ovarian depletion by a dominant negative effect or by haploinsufficiency. In 2017, two novel homozygous frameshift variants were identified, emphasizing loss-of-function effect as the most probable mechanism linking *NOBOX* variants to POI [8, 9].

Furthermore, research has revealed ethnic-specific variations in the prevalence and nature of *NOBOX* variants among women with POI. For example, the rare molecular studies performed from North- or Sub-Saharian Africa found a high proportion of *NOBOX* variants in patients with POI: in a Tunisian POI cohort of 125 unrelated subjects, three known missense variants (p.(Arg117Trp), p.(Gly91Trp), and p.(Pro619Leu)) were identified in eight patients [10]. It was also observed in two other studies a high frequency of *NOBOX* variants in POI patients from sub-Saharian origin [2, 11].

To analyze pathogenicity of different variants according to specific ethnic origins, we compared the prevalence of *NOBOX* variants identified in our cohort of 810 unrelated patients with POI from different ethnical-geographical origins with the reported frequency in the general population in the gnomAD database.

Material and methods

Patients

We recruited 810 women under 40 years old in a period of 5 years, with idiopathic, sporadic or familial POI, normal karyotype (46,XX), primary or secondary amenorrhea, and elevated FSH (> 25 UI/L) in two measurements. Pregnancy was not excluded in the patient group. All participants signed an informed consent letter previously approved by the local ethics committee, and the study has been performed in accordance with the ethical standards as laid down in the Declaration of Helsinki. According to their ethnic origin, these women were divided in four groups: European ($n = 468$), North-African ($n = 123$), Sub-Saharian Africa ($n = 214$), and Asian ($n = 5$).

We used the genome aggregation database (gnomAD) (https://gnomad.broadinstitute.org/) (v2.1.2) to compare frequencies in general population. We also used the results from a control population of 200 Tunisian women with normal reproductive history [10].

Next-generation sequencing

Genomic DNA was extracted from peripheral blood samples of POI individuals and purified using the Chemagic 360 nucleic acid extractor (Perkin Elmer) and the Wizard genomic DNA Promega Kit (Promega Corporation). A customized target capture array covering 7 selected known causative genes, including *NOBOX* (NM_001080413.3), was designed by Roche NimbleGen to cover all exons and at least 25 bp of flanking intron sequence based upon NCBI Build37/UCSC version hg19. Based on the variant frequencies identified previously, 7 causative genes with confirmative functional evidence were included in the target panel, including one meiosis gene (*MSH5*), three ligands and receptors associated genes (*BMP15*, *FSHR*, *and GDF9*), and three transcriptional factors preferentially expressed in the ovaries (*FOXL2*, *NOBOX*, *and NR5A1*). The selected genes satisfied at least one of the following requirements: (1) pathogenic variants of the gene have been identified in women with POI; (2) functional studies have been performed to confirm that the genes are involved in ovarian function maintenance [14]. Although the genetic etiology of POI is heterogeneous with more than 100 genes associated with POI, this panel may identify the etiology in ~ 8% of women with POI [3]. The list of the 7 genes known to cause POI can

be found in the supplemental information for this article (Table S1). Genomic DNA (1 µg) was fragmented using an enzymatic approach. The sample library was prepared using a KAPA Library preparation kit (KAPA, KR0935). The target regions were captured using the SeqCap EZ library kit (Roche NimbleGen) and then subjected to multiplex sequencing with the Illumina MiSeqSystem to a mean coverage of 200X according to the manufacturer's recommendations. The paired-end reads were processed with Polyweb (http://www.polyweb.fr/), an in-house Paris Descartes bioinformatics platform pipeline following GATK best practices recommendations and variants annotation (http://www.broadinstitute.org/gatk/guide/topic?name=bestpractices). Sequences were aligned to the human genome hg19 (NCBI build37/hg19 version) with BWA and SNVs and indels were called using three different tools: genome analysis toolkit (GATK) unified genotyper, SAMtools, and FreeBayes. All calls with a read coverage $\leq 5\times$ and a Phred-scaled SNP quality of ≤ 20 were filtered out. Variants were annotated with an in-house Paris Descartes bioinformatics platform pipeline based on the Ensembl database (release 67). We excluded from the analysis all *NOBOX* variants with a MAF > 3%.

In silico *analysis*

In silico analysis was implemented with CADD (combined annotation dependent depletion) tool and with Alamut v.2.4 (Interactive Biosoftware) (https://www.interactive-biosoftware.com/alamut) to predict pathogenic effects.

Mutation validation

All variants were confirmed by Sanger sequencing. Amplified PCR products were sequenced with specific primers on an ABI3130xl genetic DNA analyzer (Applied Biosystems, Life technologies) using BigDyedideoxy terminator chemistry (Applied Biosystems, Foster City, CA, USA). Primer sequences and positions, PCR conditions, and product sizes are available on request.

Statistical analysis

The statistical analyses were done using the SAS software (SAS institute, Inc. Cary, NC) and R environment (https://www.R-project.org/.) The Fisher exact test was used to assess the differences between the cohort and the general population and to study the estimated odds ratio (OR). The data are presented as a percentage or mean and SD for quantitative variables. $P < 0.05$ was considered statistically significant.

Results

Screening of the whole coding region of the *NOBOX* gene allowed us to identify 35 different variants: 2 frameshift variants, 2 nonsense variants, 1 consensus splice site variant, 1 inframe deletion, and 29 different missense variants (Fig. 1; Table 1). Among these variants, 26 were not reported yet (Table S2). In total, 171 patients with POI (21%) carried at least one *NOBOX* variant. Five patients carried at two different

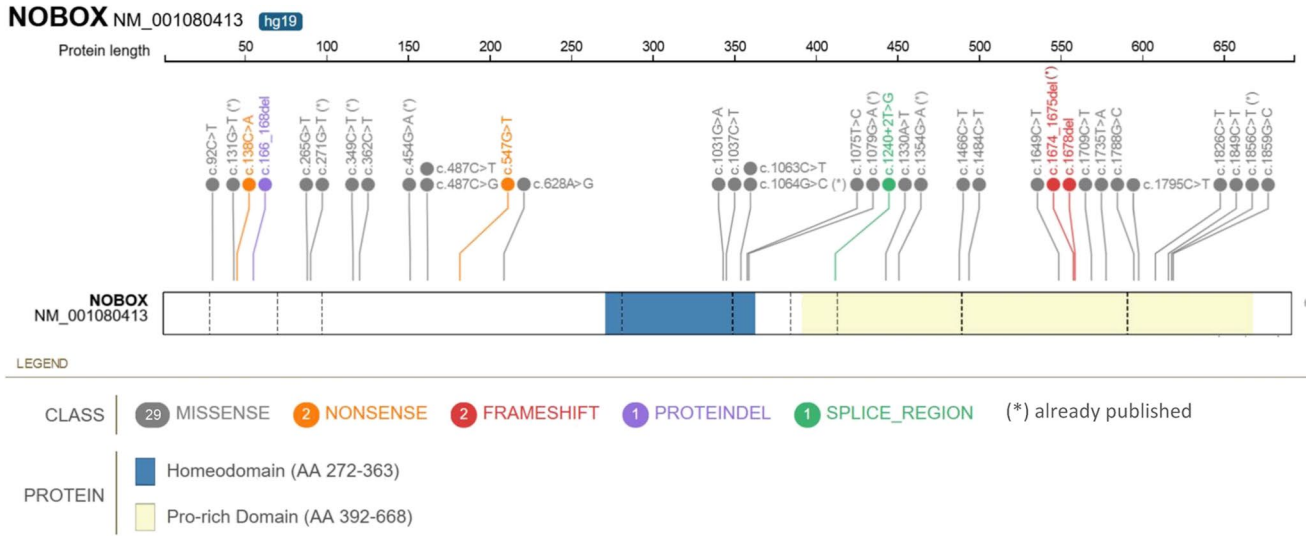

Fig. 1 Structure of the human NOBOX protein and *NOBOX* variants identified in our cohort of 810 women with POI. Figure was generated using St. Jude PeCan Data Portal

Table 1 *NOBOX* variants identified in the cohort of 810 individuals. Results of prediction of pathogenicity using different softwares (CADD, SIFT, Polyphen-2, Revel) are indicated for each variant. Results of frequency in gnomad database are indicated. "In association" means that the *NOBOX* variant was identified with another *NOBOX* variant in the same individual (without available segregation)

Sequence variation	Case count	Case count (in association)	NM_001080413.3 (hg19)	Location (exon, E; intron, I)	Aminoacid change	GnomaADvs3	GnomAD (women)	GnomADvs3 (African/African American)	SIFT	Polyphen-2	CADD (phred)	REVEL
c.92C > T	1		chr7:g.144101767G > A	E2	p.(Pro31Leu)	nd			Tolerated	Benign	0.19	Uncertain
c.131G > T	14	3	chr7:g.144101728C > A	E2	p.(Arg44Leu)	718/152178	389/77842	679/41428	Tolerated	Benign	0.37	Benign
c.138C > A	1		chr7:g.144101721G > T	E2	p.(Tyr46Ter)	15/152190	9/77846	1/41456			25.50	
c.166_168del	1		chr7:g.144101695_144101697del	E2	p.(Phe56del)	13/151980	8/77770	12/41404			1.18	
c.265G > T	3		chr7:g.144098989C > A	E3	p.(Val89Leu)	223/152282	114/77784	212/41272	Tolerated	Benign	7.50	Benign
c.271G > T	32	7	chr7:g.144098983C > A	E3	p.(Gly91Trp)	1192/152122	610/77810	1111/41410	Tolerated	Probably damaging	17.53	Uncertain
c.349C > T	36	4	chr7:g.144098634G > A	E4	p.(Arg117Trp)	3362/152182	1765/77832	3148/41426	Damaging	Probably damaging	16.49	Uncertain
c.362C > T	4	1	chr7:g.144098621G > A	E4	p.(Pro121Leu)	226/152168	122/77834	206/41488	Tolerated	Benign	10.1	Benign
c.454G > A	12		chr7:g.144098529C > T	E4	p.(Gly152Arg)	503/152142	223/77812	38/41428	Damaging	Possibly damaging	8.51	Uncertain
c.487C > T	2	1	chr7:g.144098496G > A	E4	p.(Arg163Cys)	23/152050	14/77784	19/41372	Tolerated	Possibly damaging	13.32	Uncertain
c.487C > G	3		chr7:g.144098496G > C	E4	p.(Arg163Gly)	39/152050	9/77784	3/41372	Tolerated	Benign	1.1	Uncertain
c.547G > T	2		chr7:g.144098436C > A	E4	p.(Glu183Ter)	nd					33	
c.628A > G	1		chr7:g.144098355 T > C	E4	p.(Asn210Asp)	8/152162	2/77820	1/41436	Tolerated	Benign	9.65	Benign
c.1031G > A	1		chr7:g.144097219C > T	E5	p.(Arg344His)	4/152246	2/77866	0	Tolerated	Benign	5.55	Uncertain
c.1037C > T	1		chr7:g.144097213G > A	E5	p.(Ala346Val)	71/152198	39/77846	67/41460	Tolerated	Benign	0.08	Uncertain
c.1063C > T	1		chr7:g.144096941G > A	E6	p.(Arg355Cys)	1/152198	0/77846	0	Damaging	Probably Damaging	33	Damaging
c.1064G > C		1	chr7:g.144096940C > G	E6	p.(Arg355Pro)	nd			Damaging	Probably damaging	32	Damaging
c.1075 T > C	1		chr7:g.144096929A > G	E6	p.(Trp359Arg)	nd			Damaging	Probably Damaging	29.66	Damaging
c.1079G > A	5	1	chr7:g.144096925C > T	E6	p.(Arg360Gln)	45/152152	29/77822	5/41432	Damaging	Probably damaging	29.20	Damaging
c.1240 + 2 T > G	1		chr7:g.144096488A > C	I7	p.(?)	nd					33	
c.1330A > T	2		chr7:g.144096182 T > A	E8	p.(Ser444Cys)	nd			Tolerated	Probably damaging	24.70	Damaging
c.1354G > A	26	1	chr7:g.144096158C > T	E8	p.(Asp452Asn)	1273/150002	700/76926	70/40562	Tolerated	Benign	6.81	Benign
c.1466C > T	1		chr7:g.144096046G > A	E8	p.(Thr489Ile)	nd			Damaging	Probably damaging	23.8	Benign
c.1484C > T	1		chr7:g.144095665G > A	E9	p.(Pro495Leu)	nd			Damaging	No prediction	19.73	Benign
c.1649C > T	1		chr7:g.144095500G > A	E9	p.(Thr550Met)	18/152142	7/77828	10/41422	Tolerated	No prediction	4.23	Benign
c.1674_1675del	1		chr7:g.144095480_144095481del	E9	p.(Phe559Tyrf-sTer39)	1/152158	1/77830	0			24	

Table 1 (continued)

Sequence variation	Case count	Case count (in association)	NM_001080413.3 (hg19)	Location (exon, E; intron, I)	Aminoacid change	GnomADvs3	GnomAD (women)	GnomADvs3 (African/African American)	SIFT	Polyphen-2	CADD (phred)	REVEL
c.1678del	1		chr7:g.144095471del	E9	p.(Met560Cysfs*Ter24)	nd						
c.1709C>T	2		chr7:g.144095440G>A	E9	p.(Ser570Leu)	40/152198	21/77852	1/41444	Tolerated	No prediction	17.25	Benign
c.1735 T>A		1	chr7:g.144095414A>T	E9	p.(Ser579Thr)	12/152142	9/77836	1/41424	Tolerated	No prediction	0.34	Benign
c.1788G>C	1		chr7:g.144094621C>G	E10	p.(Trp596Cys)	nd			Damaging	No prediction	27.10	Uncertain
c.1795C>T		3	chr7:g.144094614G>A	E10	p.(Pro599Ser)	nd					22.40	
c.1826C>T	3		chr7:g.144094583G>A	E10	p.(Pro609Leu)	205/152210	107/77854	165/41448	Damaging	No prediction	20.40	Uncertain
c.1849C>T	1		chr7:g.144094560G>A	E10	p.(His617Tyr)	89/189188	44/77828	12/41452	Tolerated	No prediction	3.00	Benign
c.1856C>T		3	chr7:g.144094553G>A	E10	p.(Pro619Leu)	557/152170	307/77848	530/41438	Tolerated	No prediction	4.14	Benign
c.1859G>C	1		chr7:g.144094550C>G	E10	p.(Gly620Ala)	4/152216	3/77852	4/41458				

heterozygous variants (i.e., p.[(Arg44Leu)];[(Pro619Leu)], p.[(Gly91Trp)];[(Arg117Trp)], p.[(Gly91Trp)];[(Arg163Cys)], p.[(Arg355Pro)];[(Arg360Gln)], and p.[(Asp452Asn)];[(Ser579Thr)]), and two were homozygous for a missense variant (p.(Gly91Trp) and p.(Arg117Trp)).

Although several variants (such as p.(Arg117Trp), p.(Gly91Trp), p.(Pro619Leu), p.(Gly152Arg), p.(Asp452Asn)) showed variable degrees of in vitro functional impairment, including defects in transcriptional activity, autophagosomal degradation, nuclear localization, or protein instability [2, 13, 15, 16], the pathogenicity of these variants remains unclear. Using data from our POI cohort, we compared the frequency of the *NOBOX* variants (with MAF < 0.03) in our population of POI patients and in the general population.

When we compared the frequency of our variants in our POI population and in the general population taking into account only women, no significant differences were observed for 15 rare variants (Table 2). However, all lost-of-function variants (nonsense, frameshift, and splice) were found to be significantly overrepresented in POI except c.138C > A (p.(Tyr46Ter)) (Table 2). Moreover, we observed a significant excess of four missense variants (c.131G > T, p.(Arg44Leu); c.271G > T, p.(Gly91Trp); c.454G > A, p.(Gly152Arg); c.1354G > A, p.(Asp452Asn)) in the affected population (Table 2). OR of the mutant carriers varied between 1.98 (c.1354G > A) and 3.123 (c.271G > T) (Table 2).

Nevertheless, several *NOBOX* variants were detected worldwide with uneven geographic/ethnic distribution. Thus, variants p.(Arg44Leu), p.(Gly91Trp), and p.(Arg117Trp) were most frequently reported in African/African-American whereas variants p.(Asp452Asn) and p.(Gly152Arg) were predominantly reported in subjects of European origin. Considering these differences of distribution, we chose to perform an analysis of *NOBOX* variants with ethnic subgrouping (African, European, and North-African origin).

If we considered the subgroup of African origin, we observed a significant overrepresentation of five variants in the affected POI population compared with the African population (from gnomAD) (Fig. 2). However, no significant differences were observed for p.(Arg44Leu), p.(Arg117Trp), and p.(Gly152Arg) and the other rare variants. OR varied between 2.61 (c.271G > T) and 5.41 (c.1354G > A) for the most frequent *NOBOX* variants (Fig. 2).

If we considered the subgroup of European origin, we also observed a significant overrepresentation of 12 variants in the affected POI population compared with the European population (from gnomAD) (Fig. 3). No significant differences were observed for the 23 other rare variants. OR varied between 2.50 (c.454G > A) and 24.23 (c.271G > T) for the most frequent *NOBOX* variants (Fig. 3).

Table 2 Risk of POI associated with *NOBOX* variants in population-based studies (POI cohort versus controls from the gnomAD general population). A light grey box denotes that the enrichment reaches significance. Note: OR, odds ratio; CI, confidence interval (min = minimal, max = maximal)

Variant	Count.1_POIcohort	Count.1_GnomAD	Count.2_POIcohort	Count.2_GnomAD	p-value	OR	CI.min	CI.max
c.92C>T	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.265G>T	3	223	1617	152059	0.516698608556709	1.2650767214003	0.258653520106966	3.75180579889747
c.271G>T	39	1192	1581	150930	2.63768278410838e-09	3.12333952522714	2.20114508345259	4.31429063387068
c.349C>T	40	3362	1580	148820	0.445055208106628	1.12063921683822	0.796019626562491	1.5362863370119
c.362C>T	5	226	1615	151942	0.0988799590051486	2.08144123583673	0.668435952406749	4.93929541443761
c.454G>A	12	503	1608	151639	0.0137010089532167	2.2497466729746	1.15224830515539	3.97618223528861
c.487C>G	3	39	1617	152011	0.00988770827244743	7.23238659874573	1.42813130431048	22.7976708338689
c.487C>T	3	23	1617	152027	0.00253673404502032	12.262319909883	2.35455340739763	40.6183924603712
c.547G>T	2	0	1618	152064	0.000111047140052813	Inf	17.6383230996278	Inf
c.628A>G	1	8	1619	152154	0.0909134215351792	11.7468748091206	0.264585470631688	87.7285767258965
c.1031G>A	1	4	1619	152242	0.0515469518931769	23.5174567829235	0.477065820192834	237.260317958916
c.1037C>T	1	71	1619	152127	0.533499940148448	1.32342521715155	0.0330266420925711	7.62575591839202
c.1063C>T	1	1	1619	152197	0.020953000941232	93.9802963996824	1.19685039348216	6850.65893078651
c.1064G>C	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1075T>C	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1079G>A	6	45	1614	152107	1.62724287824283e-05	12.5646948073863	4.37417286030927	29.5929827219825
c.1240+2T>G	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1331G>T	17	718	1603	151460	0.00289039588750862	2.23709678880709	1.29264653897788	3.61900811040018
c.1330A>T	2	0	1618	152064	0.000111047140052813	Inf	17.6383230996278	Inf
c.1354G>A	27	1273	1593	148729	0.0015159422072947	1.98021588006881	1.29563038612177	2.90607136915813
c.138C>A	1	15	1619	152175	0.155847111064911	6.26590458997851	0.148758321909214	40.7789413562329
c.1444G>A	14	26259	1606	125641	7.05443357299252e-109	0.0417098821980462	0.0227220997370021	0.0702012916157795
c.1466C>T	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1484C>T	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1649C>T	1	18	1619	152124	0.18229474274326	5.22041412239493	0.12521254477003	33.1334519521171
c.166_168del	1	13	1619	151967	0.137954081536119	7.22140267397909	0.169796318142337	48.1247941988534
c.1674_1675del	1	1	1619	152157	0.0209584222934777	93.9555828256745	1.19653584351747	6848.99815820673
c.1678del	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1709C>T	2	40	1618	152158	0.0723832889215135	4.70190245730195	0.550037418211828	18.1566797854412
c.1735T>A	1	12	1619	152130	0.128637277202005	7.83033034604057	0.183066435628237	52.9910186395137
c.1788G>C	1	0	1619	152064	0.0105411103302881	Inf	2.40699755050815	Inf
c.1795C>T	3	0	1617	152064	1.169130233895e-06	Inf	38.8472678352687	Inf
c.1796C>A	4	19253	1616	132807	5.48574925585717e-87	0.0170814754969256	0.00462625056066567	0.0438197623338642
c.1826C>T	3	205	1617	152005	0.485439140844756	1.37564418770125	0.281072841436504	4.08437086754716
c.1849C>T	1	89	1619	189099	0.535856235296825	1.31235350824595	0.032841430707626	7.5127817142433
c.1856C>T	3	557	1617	151613	0.299530044329889	0.505002773042486	0.103718469715221	1.48615434631903
c.1859G>C	1	4	1619	152212	0.0515567926338943	23.5127577399505	0.4769718121417	237.214015689886
c.1991A>G	1	4575	1619	147599	3.628947587428e-20	0.0199279846072886	0.000508779853308317	0.111234396123635

Finally, if we considered the subgroup of North-Africa, we also observed a significant overrepresentation of 3 variants and a significant underrepresentation of only one variant in the affected POI population compared with the African population (from gnomAD) (in absence of specific database for North-Africa) (Fig. 4A). When we compared to the European population (from gnomAD), we observed a significant overrepresentation for 6 variants (Fig. 4B). No significant difference was observed for the other variants. OR varied between 19.73 (c.349C > T) and 173 (c.131G > T) for the most frequent *NOBOX* variants (Fig. 4B).

For each variant, if we considered the population with the highest prevalence, comparable significate OR values were calculated for variants p.(Gly152Arg) (OR 2.50 European population; OR 2.25 / all populations) and p.(Gly91Trp) (OR 2.61 / African population; OR 3.12 All populations). For the other most frequent *NOBOX* variants (p.(Arg44Leu), p.(Asp452Asn), p.(Arg117Trp)), no significant difference was observed.

Discussion

NOBOX variants are often described as the most common genetic explanation for POI. This appears to be true in African and European populations but not in Asian populations [10, 12, 13]. However, frameshift, nonsense, and splice site variants are very rare in the general population as well as in POI cohorts (Table S2). Up to now, 8 frameshift [8, 9], 1 splice variant, and 4 nonsense variants have been identified in women with POI (Table S2) [2, 16]. In this study, we identified three novel loss-of-function variants (1 frameshift, 1 nonsense, and 1 splice site) (Table 1). Interestingly, most of the loss-of-function variants described here and in the literature in women with POI map toward to the 3'end of the *NOBOX* transcript (within or after the homeodomain; 12/15 variants) (Table S2). However, the great majority of *NOBOX* variants were missense variants (86%) for which pathogenicity remains elusive especially since few missense variants are frequently observed in the general population. To help classifying these variants, we analyzed the OR values according to their molecular consequences and to their ethnic frequency.

According to the significant OR values, we distinguished two groups of risk variants: variants with a high-risk (OR > 5) (see Table 2) and variants with a moderate risk (OR < 5). All nonsense, frameshift and splice variants except c.130C > A (p.(Tyr46Ter) are in the high-risk subgroup which is in accordance with NOBOX loss-of-function being associated with POI.

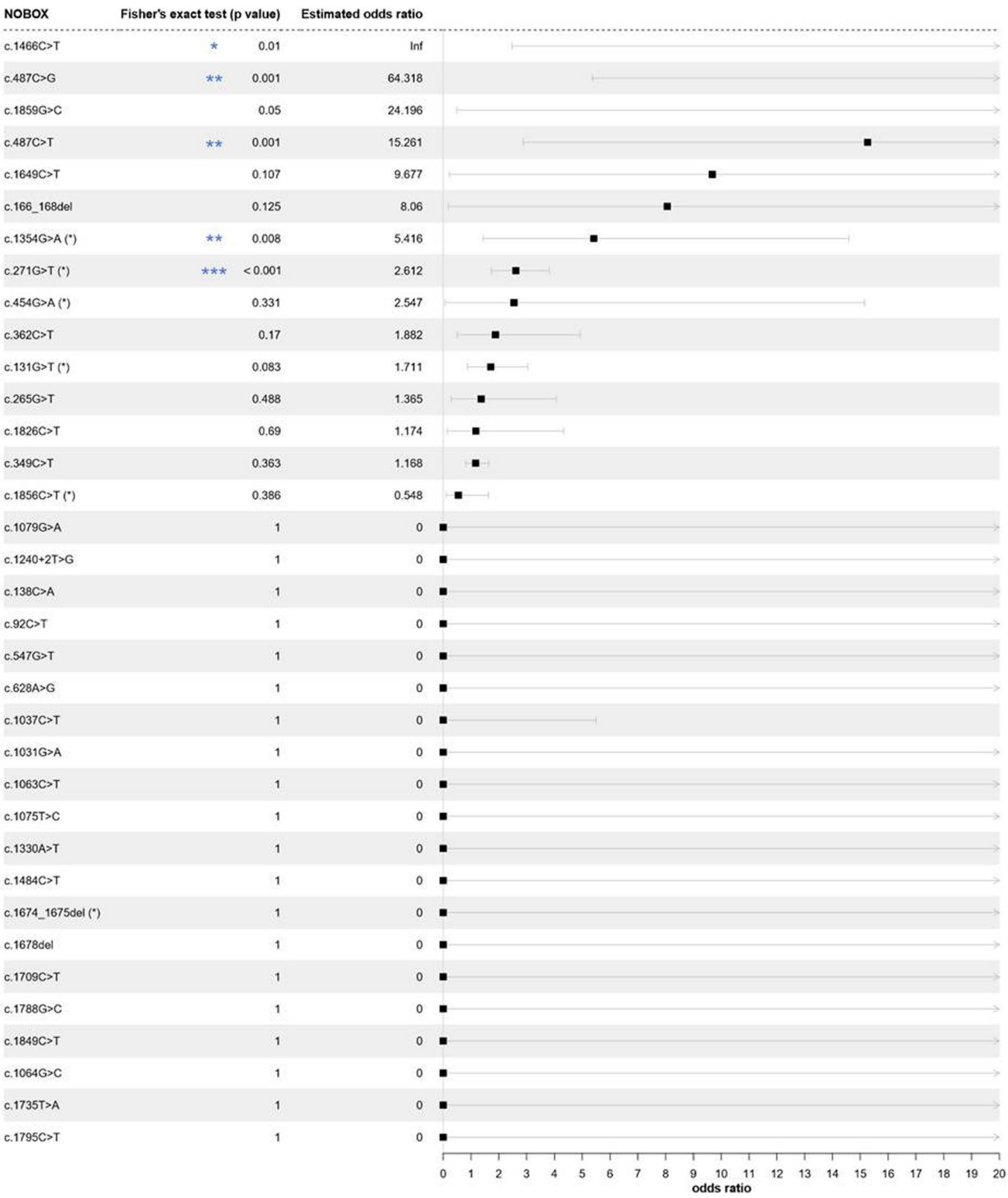

Fig. 2 Forest plot showing the odds' ratios for risk of POI among women from African origin (as compared to gnomAD population from African origin). Previously published variants are followed by (*). Note: OR, odds ratio; CI, confidence interval. Significance of Fisher's test is highlighted with blue stars

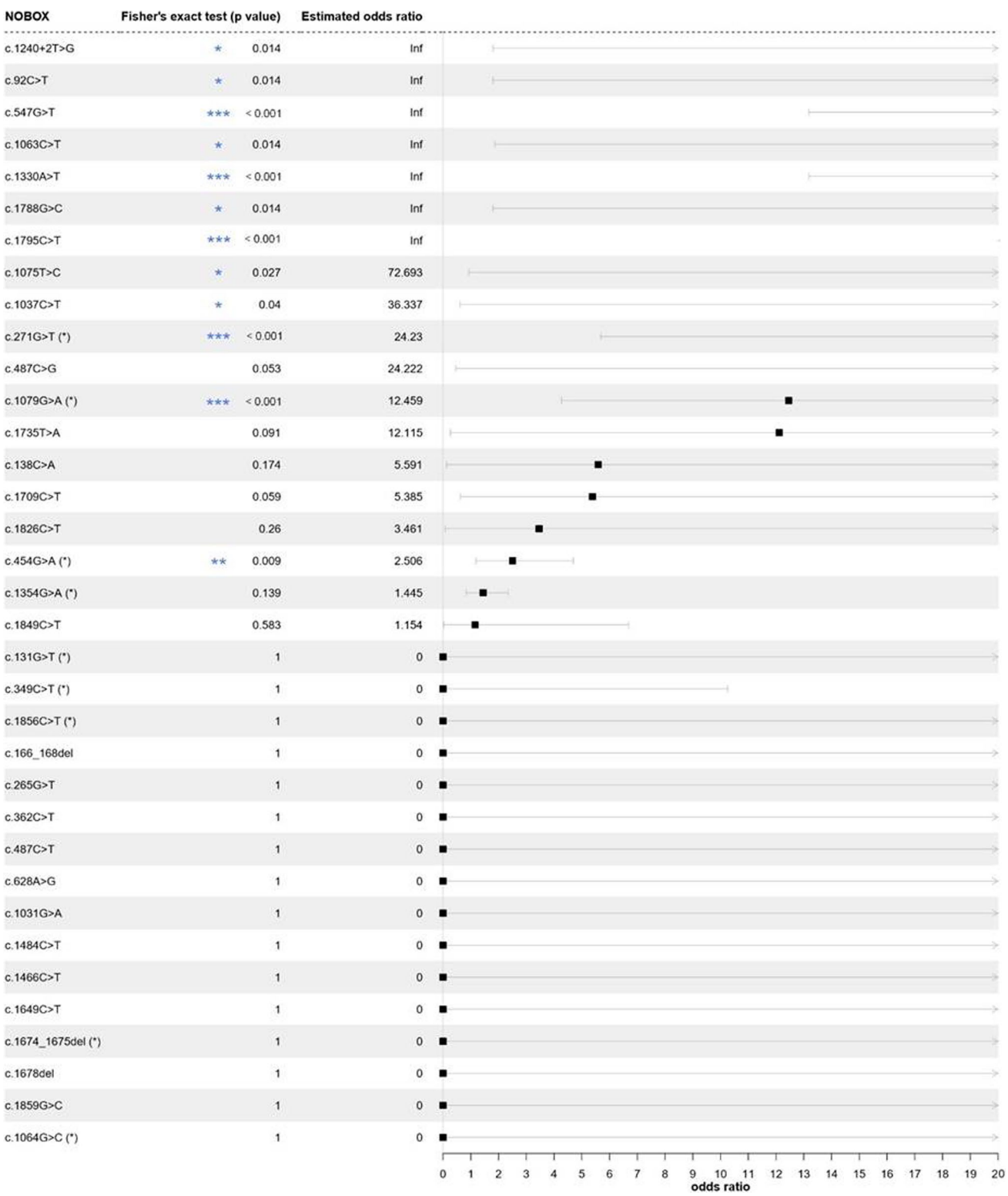

Fig. 3 Forest plot of the calculated odds ratio for each *NOBOX* variant among women from European origin (as compared to gnomAD population from European origin). Previously published variants are followed by (*). Note: OR, odds ratio; CI, confidence interval. Significance of Fisher's test is highlighted with blue stars

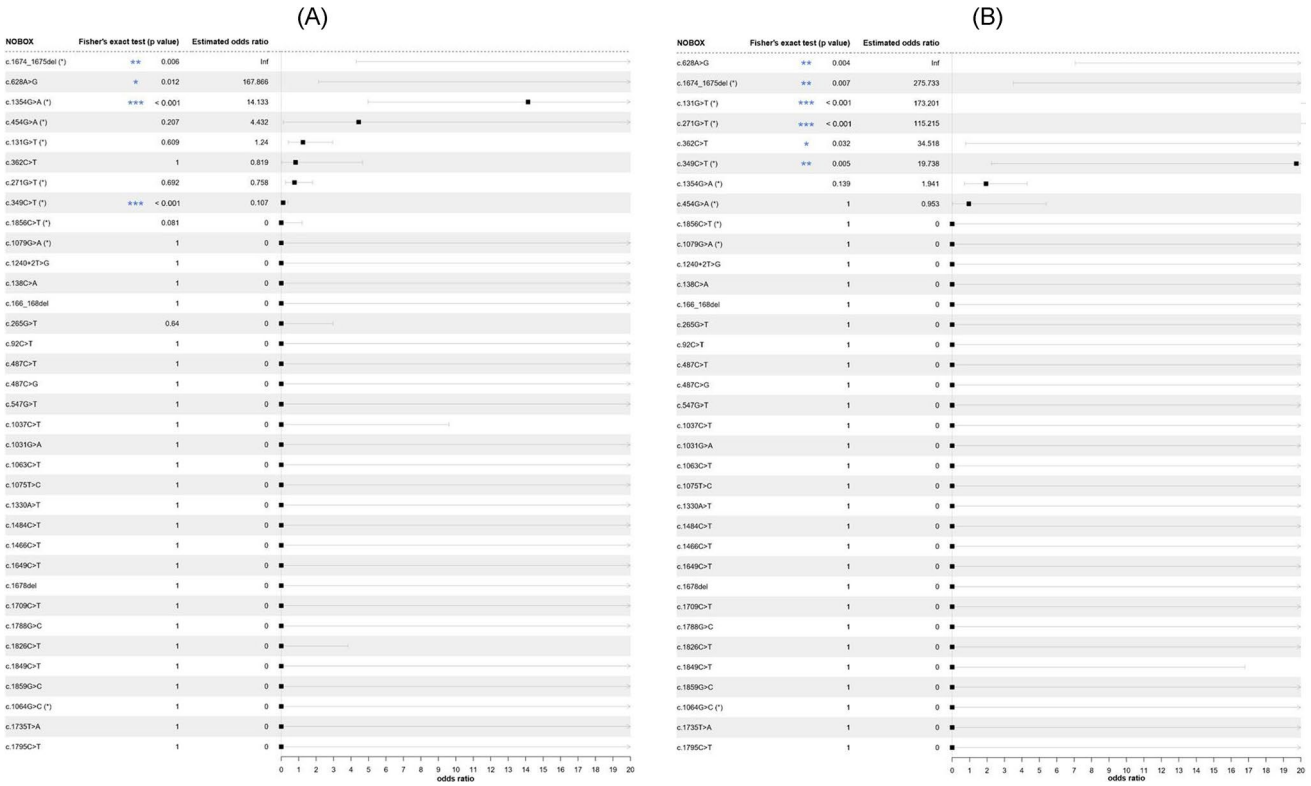

Fig. 4 Forest plot of the calculated odds ratio for each *NOBOX* variant among women from North-Africa origin (as compared to gnomAD population from African origin (**A**) or from European origin (**B**)). Previously published variants are followed by (*). Note: OR, odds ratio; CI, confidence interval. Significance of Fisher's test is highlighted with blue stars

Regarding missense variants, we observed a significant overrepresentation of the most frequent ones (c.131G > T, p.(Arg44Leu); c.271G > T, p.(Gly91Trp); c.454G > A, p.(Gly152Arg); c.1354G > A, p.(Asp452Asn)) in our 810 POI patients as compared to the general population using data from the gnomADvs3 database, except for p.(Arg117Trp) which appeared to have the same frequency in both. However, taking into account the ethnic origin of the individuals, we observed no significant OR difference for p.(Arg44Leu) and p.(Arg117Trp) in African subgroup and for p.(Asp452Asn) in European subgroup. These population data suggest that these variants could be considered benign variants with a mild or very mild functional impact. In contrast, we detected a significant OR difference for p.(Gly91Trp) in African/African-American subgroup and p.(Gly152Arg) in European subgroup suggesting that these variants could be considered variants with a moderate functional impact, OR being 2.61 and 2.50, respectively.

Moreover, the functional consequences of p.(Arg44Leu), p.(Gly91Trp), p.(Arg117Trp), p.(Gly152Arg), and p.(Asp452Asn) were previously characterized in cell culture experiments with transfected HEK 293 T cells and CHO cells [2, 13, 15]. In accordance with our results, a reporter assay based on the ability to stimulate *KIT-L* or *GDF9* as a NOBOX-inducible promoter driving the luciferase expression showed that p.(Gly91Trp) and p.(Gly152Arg) affected the transcriptional activity whereas p.(Arg44Leu) had no effect [2, 13, 15]. In addition, the variants p.(Gly91Trp) and p.(Gly152Arg) impaired the nuclear localization of the protein [15]. Although these variants reported as benign/VUS in ClinVar, all these functional and population data suggest that the p.(Gly91Trp) and p.(Gly152Arg) variants could be considered pathogenic variants, whereas p.(Arg44Leu) should be considered a benign variant. It is highly probable that these pathogenic variants present an incomplete penetrance effect.

However, we observed an imperfect correlation between the risk effects and the functional loss of *NOBOX* for two variants p.(Asp452Asn) and p.(Arg117Trp). Previous reporter assays based on the ability to stimulate *GDF9* or *OCT4* promoter showed that p.(Asp452Asn) and p.(Arg117Trp) altered the transcriptional activity [2, 8, 13], and p.(Asp452Asn) impaired the nuclear localization of the protein [15]. Our population study results showed that p.(Arg117Trp) was not more frequent in our POI cohort as compared to general population, even with ethnical subgrouping, and that the p.(Asp452Asn) was also not

more frequent in our POI cohort as compared to European population. These data suggest that these variants could be considered moderate risk pathogenic variants with probably partial and very low penetrance and/or expressivity. The impact of these variants may vary depending on the genetic background and/or the environment. We can suppose that homozygotes and compound heterozygotes for these *NOBOX* variants are expected to impart substantially higher risk; however, we need larger data set to conclude [10].

As noted above, several *NOBOX* variants are more frequently observed in patients from African origin. This excess may explain the higher prevalence of POI in the African American population than in Japanese women [17]. In a large study of women, the authors showed that POI was reported in 1.0% of Caucasian women, 1.4% of African American women, 1.4% of Hispanic women, while only 0.5% of Chinese and 0.1% of Japanese women were affected. The differences in frequency across ethnic groups were statistically significant. Hence, POI prevalence could be associated with variable frequency of *NOBOX* variants and /or of other genes involved in the disorder in different ethnic subgroups. Recently, it was shown that the *SIX6OS1* gene was involved in POI from Pakistan or Chinese origin [18].

However, in our study we can note several limitations. First, in the group of women with fertility disorders, there is a great phenotypic heterogeneity. Several of these women may have at least one child before POI or after POI, with or without assisted reproduction. Second, the gnomAD database aggregates data from over 195,000 individuals through a world-wide collaborative effort on data sharing. Most of the sequence data is generated for case–control studies of common adult-onset disease and may include "healthy" women with known or unknown POI (~ 1% in the age range 35–40 years). Finally, using whole exome or genome sequencing, more than 100 genes have been reported to be associated with POI, and we cannot exclude the presence of additional pathogenic variants in these other genes [14, 19]. Despite these limits and unresolved questions, our analysis offers strong support for the notion that several *NOBOX* variants including missense variants are relatively high risk factors for POI and argues for routine genetic screening of patients in the clinical setting. Moreover, our study demonstrated that the high frequency of POI in certain ethnic populations such as African population may be due to the high frequency of some *NOBOX* variants (such as (p.Gly91Trp)) specific to these populations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10815-023-02981-y.

Author contribution LEK and TB co-conceptualized and designed the study. SP, PJ, DMG, EP, SJC, PT, SCM, and GPB reviewed medical records and collected patient data. CF performed molecular analysis and provided data analysis, tables, and figures. PJ, TB, CV, JMD, and LEK have written content toward the first draft of the manuscript. All authors reviewed and revised the manuscript and approved the final version as submitted and agree to be accountable for all aspects of the work. All authors are responsible for the accuracy and integrity of the work.

Data availability The data that support the findings of the study are available from the corresponding author and the first author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Christin-Maitre S, Givony M, Albarel F, Bachelot A, Bidet M, Blanc JV, Bouvattier C, Brac de la Perrière A, Catteau-Jonard S, Chevalier N, Carel JC, Coutant R, Donadille B, Duranteau L, El-Khattabi L, Hugon-Rodin J, Houang M, Grynberg M, Kerlan V, Leger J, Misrahi M, Pienkowski C, Plu-Bureau G, Polak M, Reynaud R, Siffroi JP, Sonigo C, Touraine P, Zenaty D. Position statement on the diagnosis and management of premature/primary ovarian insufficiency (except Turner Syndrome). Ann Endocrinol (Paris). 2021;82(6):555–71. https://doi.org/10.1016/j.ando.2021.09.001.
2. Bouilly J, Bachelot A, Broutin I, Touraine P, Binart N. Novel NOBOX loss-of- function mutations account for 6.2% of cases in a large primary ovarian insufficiency cohort. Hum Mutat. 2011;32(10):1108–13. https://doi.org/10.1002/humu.21543.
3. Jiao X, Ke H, Qin Y, Chen ZJ. Molecular genetics of premature ovarian insufficiency. Trends Endocrinol Metab. 2018;29(11):795–807. https://doi.org/10.1016/j.tem.2018.07.002.
4. Rajkovic A, Pangas SA, Ballow D, Suzumori N, Matzuk MM. NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. Science. 2004;305(5687):1157–9. https://doi.org/10.1126/science.1099755.
5. Lechowska A, Bilinski S, Choi Y, Shin Y, Kloc M, Rajkovic A. Premature ovarian failure in nobox-deficient mice is caused by defects in somatic cell invasion and germ cell cyst breakdown. J Assist Reprod Genet. 2011;28(7):583–9. https://doi.org/10.1007/s10815-011-9553-5.
6. Venturella R, De Vivo V, Carlea A, D'Alessandro P, Saccone G, Arduino B, Improda FP, Lico D, Rania E, De Marco C, Viglietto G, Zullo F. The genetics of non-syndromic primary ovarian insufficiency: a systematic review. Int J Fertil Steril. 2019;13(3):161–8. https://doi.org/10.22074/ijfs.2019.5599.
7. Patton BK, Madadi S, Briley SM, Ahmed AA, Pangas SA. Sumoylation regulates functional properties of the oocyte transcription factors SOHLH1 and NOBOX. FASEB J. 2023;37(2):e22747. https://doi.org/10.1096/fj.202201481R.
8. França MM, Funari MFA, Lerario AM, Nishi MY, Pita CC, Fontenele EGP, Mendonca BB. A novel homozygous 1-bp deletion in the NOBOX gene in two Brazilian sisters with primary ovarian failure. Endocrine. 2017;58(3):442–7. https://doi.org/10.1007/s12020-017-1459-2.
9. Li L, Wang B, Zhang W, Chen B, Luo M, Wang J, Wang X, Cao Y, Kee K. A homozygous NOBOX truncating variant causes

defective transcriptional activation and leads to primary ovarian insufficiency. Hum Reprod. 2017;32(1):248–55. https://doi.org/10.1093/humrep/dew271.
10. Bouali N, Francou B, Bouligand J, Lakhal B, Malek I, Kammoun M, Warszawski J, Mougou S, Saad A, Guiochon-Mantel A. NOBOX is a strong autosomal candidate gene in Tunisian patients with primary ovarian insufficiency. Clin Genet. 2016;89(5):608–13. https://doi.org/10.1111/cge.12750.
11. Eskenazi S, Bachelot A, Hugon-Rodin J, Plu-Bureau G, Gompel A, Catteau-Jonard S, Molina-Gomes D, Dewailly D, Dodé C, Christin-Maitre S, Touraine P. Next generation sequencing should be proposed to every woman with "idiopathic" primary ovarian insufficiency. J Endocr Soc. 2021;5(7):bvab032. https://doi.org/10.1210/jendso/bvab032.
12 Qin Y, Choi Y, Zhao H, Simpson JL, Chen ZJ, Rajkovic A. NOBOX homeobox mutation causes premature ovarian failure. Am J Hum Genet. 2007;81(3):576–81. https://doi.org/10.1086/519496.
13. Bouilly J, Roucher-Boulez F, Gompel A, Bry-Gauillard H, Azibi K, Beldjord C, Dodé C, Bouligand J, Mantel AG, Hécart AC, Delemer B, Young J, Binart N. New NOBOX mutations identified in a large cohort of women with primary ovarian insufficiency decrease KIT-L expression. J Clin Endocrinol Metab. 2015;100(3):994–1001. https://doi.org/10.1210/jc.2014-2761.
14. Luo W, Ke H, Tang S, Jiao X, Li Z, Zhao S, Zhang F, Guo T, Qin Y. Next-generation sequencing of 500 POI patients identified novel responsible monogenic and oligogenic variants. J Ovarian Res. 2023;16(1):39.
15. Ferrari I, Bouilly J, Beau I, Guizzardi F, Ferlin A, Pollazzon M, Salerno M, Binart N, Persani L, Rossetti R. Impaired protein stability and nuclear localization of NOBOX variants associated with premature ovarian insufficiency. Hum Mol Genet. 2016;25(23):5223–33. https://doi.org/10.1093/hmg/ddw342.
16. França MM, Funari MFA, Lerario AM, Santos MG, Nishi MY, Domenice S, Moraes DR, Costalonga EF, Maciel GAR, Maciel-Guerra AT, Guerra-Junior G, Mendonca BB. Screening of targeted panel genes in Brazilian patients with primary ovarian insufficiency. PLoS One. 2020;15(10):e0240795. https://doi.org/10.1371/journal.pone.0240795.
17. Luborsky JL, Meyer P, Sowers MF, Gold EB, Santoro N. Premature menopause in a multi-ethnic population study of the menopause transition. Hum Reprod. 2003;18(1):199–206. https://doi.org/10.1093/humrep/deg005.
18. Fan S, Jiao Y, Khan R, Jiang X, Javed AR, Ali A, Zhang H, Zhou J, Naeem M, Murtaza G, Li Y, Yang G, Zaman Q, Zubair M, Guan H, Zhang X, Ma H, Jiang H, Ali H, Dil S, Shah W, Ahmad N, Zhang Y, Shi Q. Homozygous mutations in C14orf39/SIX6OS1 cause non-obstructive azoospermia and premature ovarian insufficiency in humans. Am J Hum Genet. 2021;108(2):324–36. https://doi.org/10.1016/j.ajhg.2021.01.010. (Erratum in: Am J Hum Genet. 2022 Jul 7;109(7):1343).
19. Ke H, Tang S, Guo T, Hou D, Jiao X, Li S, Luo W, Xu B, Zhao S, Li G, Zhang X, Xu S, Wang L, Wu Y, Wang J, Zhang F, Qin Y, Jin L, Chen ZJ. Landscape of pathogenic mutations in premature ovarian insufficiency. Nat Med. 2023;29(2):483–92. https://doi.org/10.1038/s41591-022-02194-3.

Authors and Affiliations

Pénélope Jordan[1] · Camille Verebi[1] · Sandrine Perol[2] · Sarah Grotto[2] · Corinne Fouveaut[1] · Sophie Christin-Maitre[3] · Aude Brac de la Perrière[4] · Virginie Grouthier[5] · Sophie Jonard-Catteau[6] · Philippe Touraine[7] · Geneviève PLU-Bureau[2] · Jean Michel Dupont[1] · Laila El Khattabi[1,8] · Thierry Bienvenu[1]

✉ Thierry Bienvenu
thierry.bienvenu@inserm.fr

Pénélope Jordan
Penelope.jordan@aphp.fr

Camille Verebi
Camille.verebi@aphp.fr

Sandrine Perol
Sandrine.perol@aphp.fr

Sarah Grotto
Sarah.grotto@aphp.fr

Corinne Fouveaut
Corine.fouveaut@aphp.fr

Sophie Christin-Maitre
Sophie.christin-maitre@aphp.fr

Aude Brac de la Perrière
Aude.brac-delaperriere@chu-lyon.fr

Virginie Grouthier
Virginie.grouthier@chu-bordeaux.fr

Sophie Jonard-Catteau
Sophie.jonard@chu-lille.fr

Philippe Touraine
Phillipe.touraine@aphp.fr

Geneviève PLU-Bureau
Genevieve.plu-bureau@aphp.fr

Jean Michel Dupont
Jean-michel.dupont@aphp.fr

Laila El Khattabi
Laila.elkhattabi@aphp.fr

[1] Service de Médecine Génomique Des Maladies de Système Et d'Organe, Hôpital Cochin, APHP, Centre Université de Paris Cité, 123 Boulevard de Port-Royal, 75014 Paris, France

[2] Unité de Gynécologie Médicale, APHP, Centre Université de Paris Cité Cité, Hôpital Cochin Port-Royal, 75014 Paris, France

[3] Service d'endocrinologie, Diabétologie Et Médecine de La Reproduction, APHP, Sorbonne Université, Hôpital Saint-Antoine, 75012 Paris, France

[4] Service d'Endocrinologie, de Diabétologie Et Des Maladies Métaboliques A, Hospices Civiles de Lyon, 69000 Lyon, France

[5] Service de Gynécologie Médicale, CHU de Bordeaux, 33000 Bordeaux, France

[6] Département d'assistance Médicale À La Procréation, Hôpital Jeanne de Flandre, 59000 Lille, France

[7] Département d'Endocrinologie Et Médecine de La Reproduction, APHP. Sorbonne Université, Pitié-Salpêtrière Hospital, Center for Rare Endocrine and Gynecological Disorders, ERN-HCP, Paris, France

[8] Present Address: Département de Génétique Médicale, Unité Fonctionnelle de Génomique Chromosomique, APHP. Sorbonne Université, Hôpital Armand Trousseau, 75012 Paris, France