Association of 8q22.3 locus in Chinese Han with idiopathic premature ovarian failure (POF)

Yingying Qin¹, Han Zhao¹, Jianfeng Xu², Yongyong Shi³, Zhiqiang Li³, Jie Qiao⁴, Jiayin Liu⁵, Chunrong Qin⁶, Chune Ren⁷, Jie Li⁸, Shiling Chen⁹, Yunxia Cao¹⁰, China POF Study Group¹, Joe Leigh Simpson¹¹ and Zi-Jiang Chen¹

¹Center for Reproductive Medicine, Provincial Hospital Affiliated to Shandong University, National Research Center for Assisted Reproductive Technology and Reproductive Genetics, The Key Laboratory for Reproductive Endocrinology of Ministry of Education, Shandong Provincial Key Laboratory of Reproductive Medicine, 324 Jingwu Road, Jinan 250021, China, ²Fudan-VARI Center for Genetic Epidemiology, Fudan University, Shanghai 200433, China, ³Bio-X Center, Shanghai Jiao Tong University, Shanghai 200240, China, ⁴Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, China, ⁵The First Affiliated Hospital with Nanjing Medical University, Nanjing 210029, China, ⁶Maternal and Child Health Hospital, Shenzhen 518028, China, ⁷Center for Reproductive Medicine, Hospital Affiliated to Weifang Medical College, Weifang 261053, China, ⁸The First Affiliated Hospital, Sun Yat-sen University, Guangdong 510080, China, ⁹Nanfang Hospital, Guangdong 510515, China, ¹⁰Reproductive Medicine Center, First Affiliated Hospital, Anhui Medical University, Hefei 230032, China and ¹¹College of Medicine, Florida International University, Miami, FL 33199, USA

Received July 26, 2011; Revised September 27, 2011; Accepted October 4, 2011

Premature ovarian failure (POF) is a complex heritable disorder known to be caused by chromosomal abnormalities and to date a limited number of known mutations, often autosomal. We sought to identify additional genetic loci associated with POF by performing the first large-scale genome-wide association study (GWAS). GWAS, using Affymetrix SNP 6.0 chip, was conducted in an initial discovery set of 391 well-documented (follicle-stimulating hormone >40 IU/ml) Chinese Han POF patients, compared with 895 unrelated Chinese female controls. A replication study on the most significant loci was then performed in an independent set of 400 cases and 800 controls. Suggestive significant associations were observed at 8q22.3. Replication of eight single-nucleotide polymorphisms (SNPs) (rs10464815, rs10808365, rs3847152, rs3847153, rs3847154, rs3843552, rs10955242, rs3843555) (P ≤ 3.86 x 10⁻⁶) was confirmed in verification sets. No specific candidate gene was found in the immediate region of 8q22.3. This GWAS, involving by far the largest sample of POF cases accumulated to date, revealed heretofore unrecognized association between POF and a novel genetic locus or region of unknown nature on 8q22.3. We speculate existence of a long-distance regulatory region that has relevance to the control of ovarian differentiation or oogenesis. Given failure to find association with any of the other autosomal regions known to harbor genes causing ovarian failure, our findings also underscore the likelihood of considerable genetic and etiologic heterogeneity in POF and the need for additional approaches like whole-genome sequencing.

INTRODUCTION

Premature ovarian failure (POF) is defined as cessation of menstruation before the expected age of menopause. Diagnosis is confirmed by elevated serum follicle-stimulating hormone (FSH) levels (>40 IU/l) prior to the age of 40 years. Approximately 1% of the Caucasian population is affected.

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
before age 40; ≤ 0.1% are affected prior to age 30 and 0.01% <20 years of age (1). Unfortunately, there are no accurate incidence data concerning POF in Chinese population. Diverse etiologies have been claimed, including infection, chemotherapeutic treatments, pelvic surgery and autoimmune diseases; however, genetic factors are generally considered paramount. Heritability has been well established in POF, and many genetic causes recognized (2). Loci on the X chromosome and autosomes are pivotal for ovarian development and oogenesis.

Deletions resulting in loss of certain regions on the X chromosome have long been accepted as causative. Terminal deletions arising at Xp11 or Xq13 more often cause complete than POF, whereas more distal deletions (Xp21 → pter; Xq27 → qter) usually cause POF (3). Interstitial X-deletions are also reported (4). Reciprocal (X-autosomal; autosomal) translocations explain certain cases, but whether reflective of a perturbed gene or reflective of non-specific meiotic breakdown remains unclear. Certain regions of the chromosome (designated POF1, POF1B, POF2 and POF3) have been identified for association with POF, but these associations are now realized as possibly unreliable and further do not unambiguously identify causal genes (5–7). Copy-number-variant duplications as well as deletions have also unexpectedly been observed in near-equal numbers (8,9), although usually without parental studies that are necessary to exclude familial transmission indicative of a non-functional familial polymorphism. Among several X-candidate genes, only premutation involving the fragile X (Xq27) explains a significant minority of cases, 1–5% sporadic and only still slightly more familial cases (10,11).

Autosomal genes play significant roles in ovarian development than traditionally believed. In fact, the autosomal gene FOXL2 is considered the locus most pivotal for female sex differentiation (12). Ovarian failure is one component of many syndromes caused by mutant autosomal genes. Here we focus on nonsyndromic (isolated) POF, for which several dozen autosomal genes have already been subjected to detailed candidate gene analysis by us and others. Except for FSHR in Finnish women (13), none explains >1–2% of POF in any given ethnic group. Other genes interrogated by candidate gene approaches and considered perturbed include TGF family members like BMP15 and GDF9 (14,15), inhibin A (16), homeobox genes like NOBOX (17), transcription factors FIGLA (18) and POU5F1 (19), meiotic genes MSH5 and DMC1 (20), and RFLP4 (21). We failed to show perturbations in the RNA-binding protein NANSOS3 (22), the G protein receptor GPR3 (23) and homeobox LHX8 (24). Other groups found no perturbations in nonsyndromic POF when interrogating PTEN (25), NOG (26), KIT (27), FOXO1A (28) and AMH (29). A single FOXL2 mutation was found by Harris et al. (30) among 290 interrogated nonsyndromic POF cases (31). In aggregate, no more than 10% of POF cases can be explained etiologically.

Given the above, POF becomes an ideal disorder to be subjected to genome-wide association studies (GWASs) for the detection of significant autosomal loci. In contrast to common adult onset disorders (e.g. bipolar disorder, schizophrenia), POF has the advantage of being defined by crisp laboratory-based diagnostic criteria. In normal women FSH is usually <10 IU/l; levels exceeding 10.8–17 IU/l are indicative of diminishing ovarian reserve (32), whereas levels >40 IU/l invariably mean ovarian failure. As oocyte depletion begins to occur, FSH levels rapidly rise beyond 10 IU/l to reach 40 IU/l or higher, levels never observed in normal menstruating women. Thus, a distinct quantitative end point applies for the diagnosis of ovarian failure. Only two GWAS reports exist (Supplementary Material, Table S1). One showed association with PTH-responsive B1 gene (PTHB1) in a small discovery set of 24 women and 24 controls (33); the other showed association with ADAMTS19 in the discovery set of 99 Dutch women and 181 controls (34), although not in the replication set. In addition, 5q14.1–q15 was found to be a susceptibility region in a single Dutch POF family subjected to genome-wide linkage scan (35). However, in no study did sample size exceed 100, all studies thus lacking statistical power sufficient to detect a modest association while evaluating hundreds of thousands of SNPs. We have thus conducted a GWAS using Affymetrix SNP 6.0 chip to identify predisposing genetic risk factors. Our report utilized a well-defined cohort encompassing in aggregate 791 POF patients and 1695 controls, all Chinese Han.

RESULTS

Genome-wide association study

Principal component analysis shows that the genetic substructure was comparable between our cases and controls (Fig. 1). Similarly, the Q-Q plot of adjusted P-value indicates no systematic bias, with an inflation factor of 1.019 (Fig. 2). The most significant locus was found to be located on 8q22.3. This locus includes eight SNPs in linkage disequilibrium ($r^2 = 0.98$); P-values ranged from $1.6 \times 10^{-6}$ to $3.86 \times 10^{-6}$ (Table 1). We did not observe any SNP that reached a genome-wide significant level at $5 \times 10^{-8}$ (Fig. 3) in the

![Figure 1. Principal component analysis of GWAS sample. The plot presents the top two eigenvectors identified by principal component analysis using the Eigenstrat software.](https://academic.oup.com/hmg/article-abstract/21/2/430/662720/10-February-2019)
GWAS stage, and all SNPs with \( P < 5 \times 10^{-6} \) mapped to 8q22.3 (Fig. 4).

**Replication**

To further validate the GWAS findings, we chose a smaller subset of SNPs for further replication in an independent sample set including 400 cases and 800 controls. The association results of the eight SNPs genotyped in the replication stage are presented in Table 1. All of the eight SNPs are significantly associated with POF (\( P < 0.05 \)) and with the same direction of association.

**Combined analysis**

In addition to the association test with the replication set, we also carried out a combined analysis with GWAS data using meta-analysis. The Mantel–Haenszel method was used to estimate the pooled odds ratio for all strata as a fixed effects model. After combining the GWAS and replication data, the \( P \)-value of one SNP (rs3847153) almost reached genome-wide significant level (8.85 \( \times \) \( 10^{-8} \)). The combined \( P \)-values for the remaining seven SNPs range from 1.15 \( \times \) \( 10^{-6} \) to 1.71 \( \times \) \( 10^{-7} \). Results of the combined analysis strongly support the variants at 8q22.3 being associated with susceptibility to POF.

**DISCUSSION**

In the POF susceptibility region discovered—8q22.3—no known gene match exists. This region has not previously been described to be related to POF, but it is noteworthy that two reports have described ovarian dysgenesis associated with balanced autosomal translocations involving chromosome 8, regions 8q22.1 and 8p21.2, respectively (36, 37). Chromosomal rearrangement involving 8q22 is in particular consistent with the rearrangement involving 8q22.3 (Fig. 4).

### Table 1. The meta-analysis of the association of POF with SNPs on 8q22.3

<table>
<thead>
<tr>
<th>SNP</th>
<th>BP</th>
<th>A1</th>
<th>A2</th>
<th>GWAS Case (n = 391), control (n = 895)</th>
<th>Odds ratio</th>
<th>F_A</th>
<th>F_U</th>
<th>P-value</th>
<th>Replication Case (n = 400), control (n = 800)</th>
<th>Odds ratio</th>
<th>F_A</th>
<th>F_U</th>
<th>P-value</th>
<th>Meta-analysis Case (n = 791), control (n = 1695)</th>
<th>Odds ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10464815</td>
<td>101</td>
<td>877</td>
<td>327</td>
<td>C A 0.2782 0.375 2.00E-06</td>
<td>0.6424</td>
<td></td>
<td></td>
<td></td>
<td>0.3135 0.3583 0.0333</td>
<td>0.8177</td>
<td></td>
<td></td>
<td></td>
<td>1.15E-06 0.7237 0.4839</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10808365</td>
<td>101</td>
<td>879</td>
<td>667</td>
<td>C T 0.2782 0.3756 1.76E-06</td>
<td>0.6409</td>
<td></td>
<td></td>
<td></td>
<td>0.3031 0.3578 0.0087</td>
<td>0.7807</td>
<td></td>
<td></td>
<td></td>
<td>1.71E-07 0.7067 0.0717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3847152</td>
<td>101</td>
<td>882</td>
<td>250</td>
<td>C C 0.2775 0.375 1.64E-06</td>
<td>0.6401</td>
<td></td>
<td></td>
<td></td>
<td>0.3008 0.3583 0.0053</td>
<td>0.7703</td>
<td></td>
<td></td>
<td></td>
<td>8.85E-08 0.7020 0.0371</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3847153</td>
<td>101</td>
<td>882</td>
<td>332</td>
<td>G A 0.2775 0.375 1.64E-06</td>
<td>0.6401</td>
<td></td>
<td></td>
<td></td>
<td>0.3106 0.3502 0.0314</td>
<td>0.8003</td>
<td></td>
<td></td>
<td></td>
<td>6.13E-07 0.7199 0.2569</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3847154</td>
<td>101</td>
<td>882</td>
<td>354</td>
<td>T C 0.2775 0.3726 2.38E-06</td>
<td>0.6466</td>
<td></td>
<td></td>
<td></td>
<td>0.3155 0.3601 0.0254</td>
<td>0.8113</td>
<td></td>
<td></td>
<td></td>
<td>7.12E-07 0.7205 0.2984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3843552</td>
<td>101</td>
<td>884</td>
<td>329</td>
<td>G T 0.2755 0.375 1.64E-06</td>
<td>0.6398</td>
<td></td>
<td></td>
<td></td>
<td>0.3166 0.3596 0.0199</td>
<td>0.806</td>
<td></td>
<td></td>
<td></td>
<td>5.45E-07 0.7189 0.2285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10955242</td>
<td>101</td>
<td>882</td>
<td>806</td>
<td>A G 0.2769 0.3744 1.68E-06</td>
<td>0.6398</td>
<td></td>
<td></td>
<td></td>
<td>0.3086 0.3593 0.0139</td>
<td>0.7958</td>
<td></td>
<td></td>
<td></td>
<td>6.08E-07 0.7200 0.2547</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3843555</td>
<td>101</td>
<td>884</td>
<td>409</td>
<td>C T 0.2788 0.3727 3.86E-06</td>
<td>0.6507</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

F_A and F_U indicate the allele frequency in cases and controls, respectively.

*The Bonferroni correction was based on 419,326 SNPs that passed the QC.*
our GWAS findings because molecular perturbation could have occurred and led to the deletion of a causative gene. Aboura et al. (38) and Ledig et al. (39) used array CGH to detect copy number variants, finding no changes at 8q22.3 but deletions (8q24.3 and 8q24.13) in ovarian failure cases and duplications (8q11.23 and 8q12.1) in POF cases. However, parental transmission could not be excluded; thus, significance of these findings remains unclear. We speculate that the 8q22.3 region we have discovered plays an important regulatory role in oogenesis. Examples of long-distance regulatory effects exist, such as 17q24–q25, the region containing SOX9. Haploinsufficiency of SOX9 causes XY sex reversal (male to female) (40), whereas duplication causes XX sex reversal (female to male) (41). Disturbance of a region up to 1 Mb upstream of SOX9 precludes action of SOX9 in both mice (42) and humans (43). That annotated genes do not exist in the immediate region of 8q22.3 thus does not exclude a significant effect on oogenesis.

A great strength of our study is its well-defined inclusion criteria for both POF subjects (FSH > 40IU/l) and controls (menstruating). Misclassification is thus very unlikely to confound our results. Our discovery and replication sets also presented at a younger age than POF cohorts in other studies (Table 2). This may explain our finding—a significant association with only one SNP (rs6279 in ANKK1) found in the four reported GWAS related to age at natural menopause and early menopause (44–47) (Supplementary Material, Table S2). Our study was further restricted to Chinese Han, minimizing misinterpretation due to ethnic stratification. That significance of association using GWASs was not greater probably reflects the known genetic heterogeneity, autosomal and X chromosomal, accepted as existing in POF (2).

Despite the attractiveness of GWASs for POF, genome searches in a sufficiently large number of patients are practically absent in POF, doubtless reflecting rarity of POF. An unavoidable limitation of our study is its sample size. Although we recruited by far the largest population-based cohort of POF to date, our study still lacked the sample size of thousands required to provide sufficient statistical power to detect modest associations of autosomal loci. However, this limitation becomes almost unavoidable because POF is rare compared with common disorders like type 2 diabetes mellitus or hypertension more typically subjected to GWASs. Our study lacked optimal size in the initial stage to detect other
novel genetic determinants having smaller effects on POF. Given that POF is a rare phenotype, it is possible that point mutations or copy number variants might contribute to the etiology of POF. Indeed, in previous studies, we had sought associations of rare variants with POF at various candidate genes (14,17,18). Future research may elucidate the potential contributions of rare alleles to POF also at 8q22.3.

Providing the etiologic (genetic) explanation for women with POF would be valuable clinically. Identifying genetic causation may further be of diagnostic value to other family members, especially female offspring. If a younger relative has the same genetic perturbation, timeline for achieving pregnancy might be advanced and early child-bearing encouraged. Cryopreservation of oocytes or ovarian slices can be envisioned. Of course, the ultimate goal is generating a therapeutic gene product to replace a deficiency or alternatively ameliorate the deleterious effect of a dominant negative mutation.

This GWAS study of POF, involving by far the largest sample studied to date, used SNP arrays to show for the first time that a region containing eight susceptibility SNPs on 8q22.3 was involved in POF. We speculate existence of an important yet undefined long-distance regulatory region affecting oogenesis. Identification of risk alleles on 8q22.3 justifies further investigations to better define the role played by this region and to identify populations at high risk for POF.

GWAS genotyping and quality control
Genome-wide genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0. Genotype data were called using the Birdseed algorithm (48). We applied the following criteria to perform quality control. For samples, a call rate of 0.95 was used to filter individuals. For SNP filtering (after sample filtering), SNPs with call rates <95% in either case or control samples were removed. SNPs with MAF (minor allele frequency) <1%, or deviated significantly from Hardy–Weinberg equilibrium (P ≤ 0.001), in controls were excluded. A total of 642 069 SNPs passed the quality criteria and were used for the subsequent analysis.

SNP selection for validation and replication genotyping
SNPs with a P-value <5 × 10⁻⁶ in the GWAS stage were selected. SNPs with unambiguous genotype scatter plots

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Discovery set</th>
<th>Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>391</td>
<td>400</td>
</tr>
<tr>
<td>FSH concentration (IU/l)</td>
<td>78.40 ± 28.44</td>
<td>73.88 ± 30.23</td>
</tr>
<tr>
<td>Familial clustering</td>
<td>41</td>
<td>32</td>
</tr>
<tr>
<td>Mother affected</td>
<td>39</td>
<td>27</td>
</tr>
<tr>
<td>Menopause &lt; 40 years</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Menopause &lt; 45 years</td>
<td>30</td>
<td>14</td>
</tr>
<tr>
<td>Siblings affected</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Menopause &lt; 40 years</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Age at diagnosis (years)</td>
<td>30.28 ± 4.46</td>
<td>29.79 ± 5.31</td>
</tr>
<tr>
<td>Amenorrhea primary</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>Amenorrhea secondary</td>
<td>359</td>
<td>228</td>
</tr>
<tr>
<td>Age at menarche (years)</td>
<td>14.71 ± 2.17</td>
<td>14.82 ± 2.11</td>
</tr>
<tr>
<td>Age at onset of menstrual dysfunction (years)</td>
<td>23.41 ± 6.03</td>
<td>23.85 ± 6.78</td>
</tr>
<tr>
<td>Age of amenorrhea (years)</td>
<td>24.85 ± 6.07</td>
<td>24.87 ± 5.88</td>
</tr>
</tbody>
</table>

Table 2. Clinical features of cases in discovery and replication sets, which are not significantly different (P > 0.05)
were excluded from the subsequent stage. We used the Sequenom MassARRAY iPLEX platform for genotyping in the replication stage. The Sequenom protocol involves a multiplex PCR reaction prior to a single-base primer extension reaction. The individual SNPs are identified by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Statistical analysis

The EIGENSTRAT software was applied for the evaluation of population stratification by a principal component approach (49). The top two eigenvectors were plotted in Figure 1 and were adjusted as covariates in the logistic regression analysis. Genome-wide association analysis of a single marker was performed using the PLINK software package (50). The associations of the SNPs with POF status were evaluated using a logistic regression model, assuming additive effects of the alleles (0, 1 and 2). A haplotype was used to generate the Manhattan plot (Fig. 3) (51). The quantile–quantile (Q–Q) plot (Fig. 2) was created using the R qqplot function. The regional plot was generated using LocusZoom. For the replication study, allelic association analysis of the replication sample was conducted using SHEsis (52). The GWAS and replication data were combined using meta-analysis with the R ‘meta’ package. The Mantel–Haenszel method was used to calculate the fixed effect estimate (53).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank all the participants involved in this study. We also particularly thank the China POF Study Group for patients’ recruitment.

Conflict of Interest statement. No potential conflict of interest relevant to this article was reported.

FUNDING

Supported by grants from the National Basic Research Program of China (973 Program—2012CB944700, 2010CB945002, 2007CB947403, 2007CB948103), the National Natural Science Foundation of China (30973170, 81000236, 81000238), science research foundation item of National Natural Science Foundation of China (30973170, 2010CB945002, 2007CB947403, 2007CB948103), the National Basic Research Program of China (2010CB945102, 2009CB944202, 2008CB944602, 2006CB944002, 2005CB942903), and the National Key Basic Research Project (2005CB624600). This work was supported by grants from the National Basic Research Program of China (2006CB504800, 2008CB942900, 2009CB944200), the National Natural Science Foundation of China (30972205, 30921062), and the Shanghai Natural Science Foundation (10ZR1412300). The EIGENSTRAT software was applied for the evaluation of chromosome genomic DNA sequence copy number variation associated with chromosome breakpoints associated with premature ovarian failure (POF) map to a 15-Mb YAC contig spanning Xq21. Genomics, 40, 123–131.


REFERENCES


