Germline study of AR gene of Indian women with ovarian failure

BINEET PANDA1, LAKSHMI RAO1, DURGADATTA TOSH1, HRIDESH DIXIT1, VENKATA PADMALATHA1, MURTHY KANAKAVALLI1, TURLAPATI RASESWARI1, MAMATA DEENADAYAL2, NALINI GUPTA3, BAIYANATH CHAKRABARTY3, PRATIBHA NALLARI4, & LALJI SINGH1

1Centre for Cellular and Molecular Biology, Hyderabad-500007, India, 2Infertility Institute and Research Centre, Hyderabad, India, 3Institute of Reproductive Medicine, Kolkata, India, and 4Department of Genetics, Osmania University, Hyderabad, India

(Received 17 February 2010; revised 15 June 2010; accepted 16 June 2010)

Abstract

Objective. Present study was designed for carrying out the mutational analysis of the entire Androgen receptor (AR) gene including two microsatellite (CAG) and (GGN)n promoter region in cases of premature ovarian failure (POF) and primary amenorrhea (PA).

Design. Previous reports of AR knockout mouse model showed POF phenotype, this draws an attention on the role of AR gene in the aetiology of POF for the case–control association studies in POF samples (n = 133), PA samples (n = 63) and control samples (n = 200).

Results. We identified six mutations including four novel mutations, i.e. c.636G > A, c.1885 + 9C > A, c.1948A > G, c.1972C > A, and two previously reported mutations, i.e. c.639G > A, c.2319 – 78T > G. Repeat length variation was noted in the two microsatellite regions CAG and GGN, located in the coding region of exon 1 at the N-terminal region of the AR gene. The CAG repeat length was homogenously distributed with the same frequency and no association among all cases and controls. The GGN repeat showed a significant association among the SS and SL allele with p = 0.0231 and p = 0.0476, respectively, among the POF/control samples.

Conclusions. Thus, AR gene mutations may play a role in the genetic cause of POF. Identification of the underlying genetic alteration of the AR gene is important for a proper diagnosis of POF subjects.

Keywords: Androgen receptor gene, ovarian failure, mutations, CAG repeat, GGN repeat

Introduction

Clinically, premature ovarian failure (POF, OMIM311360) is described as an unexplained amenorrhea (>6 months) with elevated levels of gonadotropins (FSH > 40 IU/l) and low levels of estrogen, before the age of 40 years. The occurrence of POF is approximately 1% amongst the female population [1]. A majority of patients that are diagnosed as POF in the Indian clinics are reported as sporadic cases, while a few are familial cases. Extensive studies have been carried out at the chromosomal as well as at the gene level to explain the aetiology of POF, but yet still there is not many studies illustrating significant genetic associations. The X-chromosome is considered to be very important in maintenance and functioning of female reproductive system. Many reports have shown abnormalities in various regions of the X-chromosome in association with ovarian failure [2]. The significance of X-chromosome abnormalities recently reported in Turkish population indicates thorough genetic evaluation of women with POF should be performed regardless of clinical features suggestive of chromosomal abnormalities [3].

Androgens play an instrumental role in the development and functional maintenance and development of the male and female reproductive organs [4,5]. Androgens are primarily considered as male hormones; therefore a little is known about the action of androgens in female physiology. Reports have shown correlation between excessive androgen production and polycystic ovary syndrome with abnormal menstrual cycle [6,7]. Androgens such as androstenedione and testosterone get converted into estrogen by aromatase activity in the granulosa cells of the ovary. Estrogen is a female-specific hormone and plays a major role in growth and differentiation of female reproductive organs. Apart from acting as estrogen precursor, androgens play several other physiological roles in the ovary [8,9].

Androgens act through the androgen receptor (AR). The AR gene is located on the chromosome Xq11.2-q12, and is the only sex hormone receptor gene located on the X-chromosome. AR expression has been described in the growing follicles [10]. The crucial functional significance of AR expression in differentiating female reproductive organs prompted us to study genetic association of AR gene and POF.

Recently generated AR-null mutant mouse line using Cre/Lox P system [11] is used to demonstrate that inactivation of AR resulted in the arrest of testicular development and spermatogenesis, failure of brain masculinisation at perinatal stage, high-turnover osteopenia and late onset of obesity in males [12], whereas no overt physical or growth abnormalities were observed in female AR−/− mice. Characterisation of the female reproductive system of AR−/− mice showed POF phenotype. Three weeks old AR−/− female mice have apparently normal ovaries with follicles similar to those in the wild-type females. Thereafter...
the number of healthy follicles in the AR^{-/-} mice gradually declines and there is a marked increase in the number of atretic follicles. By 32 weeks defects in folliculogenesis become profound, with fewer healthy follicles and more of atretic follicles, and >40% of the AR^{-/-} females are already infertile. By 40 weeks no healthy follicles are detectable in the ovaries of AR^{-/-} mice. The AR^{-/-} females become completely infertile with POI-like defects [13]. The POI-like defects observed in female AR^{-/-} mice indicate a possible link between the onset of POI and impaired AR-mediated signalling in hereditary and acquired disorders.

Based on above results, the present study was designed with an objective of carrying out germline analysis of the AR gene and association study among the cases of POI and primary amenorrhoea. This is the first ever report of the mutational analysis of the entire AR gene including two microsatellite (CAG)n, (GGN) and the promoter region among Indian women with ovarian failure.

Materials and methods

Patient and control recruitment

One hundred and ninety-six patients with ovarian failure were recruited, which include non-familial POI cases (n=121), familial POI cases (n=12), non-familial primary amenorrhoea (PA) cases (n=56) and familial PA cases (n=7). These patients were recruited at the Infertility Institute and Research Centre (IIRC), Hyderabad, and the Institute of Reproductive Medicine (IRM), Kolkata. Patients disclaiming any familial history of ovarian failure were categorised as non-familial/sporadic cases. Patients with the family history of ovarian failure in the same generation or parental generation were considered as familial cases. The diagnostic criteria for POI following the definition include at least 6 months of amenorrhea before the age of 40 years, with high serum FSH levels (>40 IU/l). The criterion for PA is PA is defined as a condition with complete absence of menses (no onset of menstrual cycle) or only induced menses. All the patients were assessed clinically, with complete medical and gynaecological history, including the menstrual history, menopausal age, serum FSH levels (three times at 1-month interval), LH levels, TSH levels, and with any history of autoimmune disease. Respective consent forms from these patients were collected by the concerned clinic. Karyotyping with a high-resolution GTG banding was carried out for all the patients and controls for chromosomal anomalies. Patients and controls with chromosomal abnormalities were excluded from the study. Normal healthy females with regular menstrual history, normal FSH levels and successful pregnancies were recruited as controls (n=200). Recruitment of the controls was entirely population-based to support the study. The Institutional Review Board of Centre for Cellular and Molecular Biology (CCMB), Hyderabad, approved the study.

DNA extraction and karyotyping

A 5 ml of peripheral blood was collected in EDTA vacutainers for genomic DNA isolation, and another 5 ml of peripheral blood was collected in heparin vacutainers for chromosomal analysis. DNA was extracted using the Nucleon BACC2 DNA extraction kit (Amersham Pharmacia Biotech, Piscataway, NJ) following the manufacturer’s protocol. Chromosomal analysis was performed on phytohaemagglutinin (PHA)-stimulated peripheral lymphocyte cultures using standard conventional cytogenetic methods.

PCR

The AR gene comprises of 8 exons. Primers for all the exons were used from Singh et al. 2006 [14] and the primers for promoter region ARPRF and ARPRR were designed using GeneTool software. All the PCR conditions included initial denaturation at 94°C for 5 min, cyclic denaturation at 94°C for 45 s and extension at 72°C for 1 min with a total of 35 cycles. All the primer sequences, with their corresponding annealing temperatures, are summarised in Table I. The presence of all sequence variants was confirmed by performing three independent PCR reactions and subsequent DNA sequencing.

DNA sequencing and analysis

All the PCR products were obtained using the above-mentioned primers, amplifying the corresponding exon and flanking regions of the gene. Sequencing was performed using the Big Dye terminator sequencing protocol, supported by Applied Biosystems using an ABI prism 3730 × 1 DNA analyser. The obtained sequence data were analysed using ‘Sequence Analysis’ and autoassembler software in MacOS. The reference AR sequence was downloaded from Ensemble database (www.ensembl.org/Homo_sapiens/index.html). The Human Genome Variation Society (HGVS) guidelines were followed for all the variant nomenclature and genotype representation throughout the text and tables.

CAG and GGN repeat length analysis

The CAG repeat region of the AR gene was amplified with a pair of primers, forward: 5’ FAM-TCACAAGCTCTGTTCCAGACGGTGCTGC-3’, reverse: 5’-GCTTGAAAGTTGCTGTTCCTC-3’ flanking the repeat region. PCR reaction mixture consisted of 1.0 μl PCR buffer (10×), 1.0 μl MgCl₂ (25 mM), 1.0 μl dNTPs (10 mM), 1.0 μl of each primer, 0.5 units AmpliTaq Gold™ DNA polymerase and 20 ng genomic DNA. PCR was done under the following conditions; initial denaturation at 94°C for 12 min followed by 30 cycles of 94°C for 1 min, 60.5°C for 1 min and 72°C for 1 min with a final extension at 72°C for 30 min. GGN repeat was amplified with a pair of primers; forward 5’FAM-CCGCTTCCTCAGTCTGGCCTGCAC 3’ and reverse 5’GCCGCGAGGGTGCCACAGC 3’ flanking the repeat region. PCR reaction mixture included 1.0 μl PCR buffer (10×), 1.0 μl MgCl₂ (25 mM), 1.0 μl dNTPs (10 mM), 1.0 μl DMSO (100%), 1.0 μl glycerol (100%), 1.0 μl of each primer, 0.5 units AmpliTaq Gold™ DNA polymerase and 20 ng genomic DNA. PCR conditions consisted of denaturation at 96°C for 15 min, followed by 40 cycles of 96°C for 1.5 min, 55.5°C for 1 min and 72°C for 3 min and a final extension at 72°C for 20 min. For GeneScan, 3.0 μl of PCR product was mixed with 0.2 μl of ROX500 and 6.8 μl Hi-Di formamide. Upon denaturation for 5 min at 96°C and cooling for 5 min on ice, samples were run on 3730DNA analyser (Applied Biosystems, USA). PCR and genotyping were repeated for all samples to confirm the number of the repeats. The raw data were further analysed using GeneMapper software (Applied Biosystems, USA).
This study reveals six sequence variants, which includes four novel variants confirmed by the replication of three independent PCR and sequencing results. The novel variants include 3 exonic and 1 intronic variant. The details of all the variants are summarised in Table II.

The exon 1 revealed a novel mutation c.636G > A in a control sample and previously reported polymorphism c.639G > A in both patients and control population. The variant c.636G > A did not alter amino acid sequence and is the first report of variation at this location of the AR gene. The silent mutation c.639G > A (snp ID rs6152) was present in 17.85%, i.e. 35 out of 196 patients and 15%, i.e. 30 out of 200 controls. The genotypic distribution of the alleles was similar among patients and controls. The Fisher’s exact analysis revealed $p = 0.4982$ (OR = 1.2319, CI = 0.3994 for POF). The patient having $+$ revealed a novel variation c.1885A in both patients and control population. The patient having $+$ revealed a novel variation c.1885A in both patients and control population.

The exon 4 revealed a missense variant c.1948A > G in the exon 3. Another novel missense variant c.1972C > A in the exon 4. The exon 3 revealed a novel mutation c.2037C > T which was present in 17.85%, i.e. 35 out of 196 patients and 15%, i.e. 30 out of 200 controls. The genotypic distribution of the alleles was similar among patients and controls. The Fisher’s exact analysis revealed $p = 0.3994$ for POF. The patient having $+$ revealed a novel variation c.1885A in both patients and control population. The patient having $+$ revealed a novel variation c.1885A in both patients and control population.

### Table I. List of primers with their annealing temperature and product length.

<table>
<thead>
<tr>
<th>Primer sequence</th>
<th>Product size</th>
<th>Annealing temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1AF</td>
<td>AAGAGAAGGGGAGGCAGGGAAG</td>
<td>584 bp</td>
</tr>
<tr>
<td>AR1AR</td>
<td>CAGGCGAGAGTTGGCCGAGGTG</td>
<td>592 bp</td>
</tr>
<tr>
<td>AR1RF</td>
<td>CCGAGAGAGGTTGGCAGTC</td>
<td>739 bp</td>
</tr>
<tr>
<td>AR1BR</td>
<td>TGGGTCAGGCTTTAACGAG</td>
<td>424 bp</td>
</tr>
<tr>
<td>AR1CR</td>
<td>ATCCCCTTATCTCATGTC</td>
<td>457 bp</td>
</tr>
<tr>
<td>AR1DR</td>
<td>GGAGAATGAGAAGGGAAAG</td>
<td>378 bp</td>
</tr>
<tr>
<td>AR2F</td>
<td>GTGATTTTTCAATGAG</td>
<td>424 bp</td>
</tr>
<tr>
<td>AR2R</td>
<td>AATGCGATAAGCTGAGACT</td>
<td>318 bp</td>
</tr>
</tbody>
</table>

### Results

Sequencing analysis was done for the coding regions of AR gene for all the cases and controls that are subsequently analysed for mutations. In the POF cases, the mean age of attaining amenorrhoea was 26 years (range 14–40 years), mean FSH level was 59.3 IU/l and mean LH level was 33.7 IU/l. Women with regular menstrual cycle, successful pregnancies and normal FSH level (4–11 IU/l) were recruited as controls (age range 30–45 years). In the control group, mean FSH level was 7.3 IU/l and mean age was 37 years.

### Sequencing results

This study reveals six sequence variants, which includes four novel variants confirmed by the replication of three independent PCR and sequencing results. The novel variants include 3 exonic and 1 intronic variant. The details of all the variants are summarised in Table II.

The exon 1 revealed a novel mutation c.636G > A in a control sample and previously reported polymorphism c.639G > A in both patients and control population. The variant c.636G > A did not alter amino acid sequence and is the first report of variation at this location of the AR gene. The silent mutation c.639G > A (snp ID rs6152) was present in 17.85%, i.e. 35 out of 196 patients and 15%, i.e. 30 out of 200 controls. The genotypic distribution of the alleles was similar among patients and controls. The Fisher’s exact analysis revealed $p = 0.4982$ (OR = 1.2319, CI = 0.3994 for POF). The patient having $+$ revealed a novel variation c.1885A in both patients and control population. The patient having $+$ revealed a novel variation c.1885A in both patients and control population.


Two novel missense variants were present in the 4th exon which codes for the hinge region of the androgen receptor, i.e. c.1948A > G This mutation altered the threonine amino acid at 650 position into alanine (PThr650Ala) ($p = 0.3994$ for POF). The patient having this mutation attained menopause at the age of 28 with 76 IU/l FSH levels, 42 IU/l LH and 1 mIU/l TSH. The ovaries were visualised as streak ovaries in this patient. Another novel missense variant c.1972C > A in the exon 4 was revealed in a POF patient (ID 184). This patient attained menopause at the age of 17 with 78 IU/l FSH, 49 IU/l LH and TSH 1.4 mIU/l. The mutation altered glutamine at 658 position into lysine. There is no report of this variant elsewhere.
The sequence analysis of intron 5 flanking exon 6 revealed the presence of a previously reported and common variation c.2319-7T > G (snp ID = rs1337076) among all individuals which could be a variant for Indian population. Apart from these mutations, no other mutations were found in the promoter region of the AR gene.

**Gene scan results**

In the present study, we have analysed the association of two polymorphic repeats namely CAG and GGN in the androgen receptor gene located on the coding region of the 1st exon.

A total of 20 different CAG alleles were analysed ranging from 10 to 31 (Mean ± SD 18.63 ± 3.11, Median 19) (Figure 1). The repeat size 25 and 30 was absent in all the cases as well as controls. Allele 31 was not present in any of the cases and among all the controls allele 10, 13, 26, 28, 29 were absent. Among all cases of POF allele 20 was observed with highest frequency of 19.29%, whereas in PA cases allele 19 had a highest frequency of 17.71%. Among the control, allele 18 had a highest frequency of 17.64%. However the mean variation in the CAG repeat between cases (mean = 18.54) and control (mean = 18.85) was found to be statistically non-significant as evident from the Z test.

The GGN repeat ranged from 13 to 28 allele (mean ± SD, 20.95 ± 1.31; median 21) where the allele 14 and 27 did not show any frequency (Figure 2). The AR allele with 21 GGN repeats was predominant with a very low frequency of smaller allele distributed between the ranges of 13–28. AR allele 21 was found in 79.01% of the cases and 66% of the control and rest of the alleles was distributed scarcely among the range of 13–28. Apart from this no allele size of 25 and 26 was found in any control sample.

Classified the samples on the basis of presence of SS (small small), SL (small large) and LL (large large) alleles by using median value as the cut-off [15]. For CAG, the median was taken to be 19 and for GGN repeat the median was 21. CAG allele ≥19 was considered to be short (S) and allele ≤21 was chosen to be long (L) (Figure 3). Analysis revealed that 42.98% of the POF samples, 43.75% of the PA sample and 35.29% of the controls had a SS allele, whereas 37.72% of POF, 37.5% of PA samples and 47.06% of the control samples were in SL allele. Very few alleles were present in LL allele having a similar distribution of 19.3%, 18.75% and 17.65% in POF, PA and control samples, respectively.

For the GGN repeat allele ≥21 was taken to be short (S) and allele ≤22 was chosen to be long (L) (Figure 4). In GGN repeat also the SS allele distribution was predominant with a frequency of 73.87% in POF, 84.32% in PA and 87.88% in control samples. SL and LL showed a very scarce distribution.

**Discussion**

Androgen receptor is a 92-kDa protein and is encoded by a gene with eight exons. The exons encode specific regions of

<table>
<thead>
<tr>
<th>S. No.</th>
<th>db SNP Reference ID</th>
<th>Mutation</th>
<th>Position</th>
<th>AA Change</th>
<th>POF</th>
<th>PA</th>
<th>CNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Novel 1</td>
<td>c.636G &gt; A</td>
<td>Exon 1</td>
<td></td>
<td>0/133</td>
<td>0/63</td>
<td>1/200</td>
</tr>
<tr>
<td>2</td>
<td>rs6152</td>
<td>c.639G &gt; A</td>
<td>Exon 1</td>
<td></td>
<td>26/133</td>
<td>9/63</td>
<td>30/200</td>
</tr>
<tr>
<td>3</td>
<td>Novel 2</td>
<td>c.1885 + 9C &gt; A</td>
<td>Intron 3</td>
<td></td>
<td>1/133</td>
<td>0/63</td>
<td>0/200</td>
</tr>
<tr>
<td>4</td>
<td>Novel 3</td>
<td>c.1948A &gt; G</td>
<td>Exon 4</td>
<td>p.Thr650Ala</td>
<td>1/133</td>
<td>0/63</td>
<td>0/200</td>
</tr>
<tr>
<td>5</td>
<td>Novel 4</td>
<td>c.1972C &gt; A</td>
<td>Exon 4</td>
<td>p.Gln658Lys</td>
<td>1/133</td>
<td>0/63</td>
<td>0/200</td>
</tr>
<tr>
<td>6</td>
<td>rs1337076</td>
<td>c.2319-7T &gt; G</td>
<td>Introns 5</td>
<td></td>
<td>133/133</td>
<td>63/63</td>
<td>200/200</td>
</tr>
</tbody>
</table>

Figure 1. Distribution of CAG repeats of AR gene in women with ovarian failure.
the gene such as transactivation region (1st exon), DNA-binding domain (2nd and 3rd exons), hinge region (4th exon) and ligand-binding domain (5th to 8th exons) that binds to testosterone (T) or dihydrotestosterone (DHT) [16].

We analysed the coding region of the AR gene including two highly polymorphic repeat regions viz. CAG and GGN repeats as well as the promoter region of the AR gene. The polymorphic CAG and GGN repeats are present within the first exon and encode variable lengths of polyglutamine and polyglycine stretches, respectively, in the N-terminal region of the protein. The CAG repeat can vary in length from 8 to 35 repeats. The GGN repeat is a complex repeat represented by $(\text{GTT})_3\text{GGG(GGT)}_2\text{GGC}_n$, which varies from 10 to 30 repeats. The variations in the CAG repeat length have been associated with diverse clinical conditions [17]. These studies along with in vitro assays have shown an inverse correlation between CAG repeat length and AR transactivation functions; however, GGN repeat length variation has been relatively less studied and necessitates more population studies for deriving conclusive observations. However a recent in vitro study by Nenonen et al., has shown that AR protein with 22CAG genotype has the highest activity compared to AR with 16CAG and 28CAG [18].

Understanding the growth, development and differentiation aspects of ovaries, oocytes or more broadly the reproductive system at functional and genetic level is crucial for the aetiology of female infertility. The null mice for the AR gene (AR$^{-/-}$) expressed phenotypes of POF, which is suggestive of the androgen receptor’s role in the folliculogenesis. There are significant human population-based genomic studies that display distinct phenotypic outcomes of various gene studies in null mice. The present study is an important contribution to corroborate the association of the AR gene with the aetiology of POF in humans.

An extensive sequence analysis was carried out for androgen receptor gene encompassing the promoter version and all the exonic regions. The analysis revealed a few mutations both in the exons and their flanking intronic regions. One control sample revealed a novel mutation c.636G>A in the first exon; however, this mutation was not found in any other sample. Another previously reported variant c.639G>A (snp ID = rs6152) in the 1st exon was almost equally prevalent among both the patient and control samples independent of disease aetiology. One POF patient (ID = 32) had a mutation c.1885 + 9C > A in the intron 3 region. This patient was diagnosed with menopause at the age of 18, streak ovaries during clinical investigations and hormonal profile of 50 IU/l FSH and
also showed a significantly skewed pattern towards (POF/control) as a result not showing the same conclusion could be because of various reasons. Another missense mutation c.1972C>A was found in one POF patient (ID = 184) who attained menopause at the age of 17 with 78 IU/L FSH, 49 IU/L LH levels. The 5th intronic region revealed a previously reported polymorphism variant c.2319–78T>G (p.L1337Q) in all individuals, this variant is predominantly found in most of the Asian populations.

Overall, the mutational analysis revealed only the rare presence of mutations, excluding population variants, in the patient population. Three patients had novel mutations, of which two were missense in nature. All three patients had significantly high FSH and LH levels, and two of them also presented with streak ovaries, although our patient population includes several other patients diagnosed with the presence of streak ovaries.

The repeat length polymorphism analysis was also carried out for the CAG and the GGN repeats among all the cases and controls. The CAG allele was heterogeneously distributed among the entire population ranging from 10 to 31. The CAG repeat length with allele size 19 was most frequently observed in the patient population whereas the allele size 18 was found most frequent in the control population. The varying sized allele distribution for CAG repeats was not having significant differences among the case populations and the control population.

The GGN repeat revealed a slight difference in the allele distribution pattern over CAG repeat in terms of CAG allele size 21 being commonly the most frequent in both the case and control population. Classifying the CAG and GGN repeats into SS, SL, and LL alleles revealed that majority of the cases and controls had their allele distribution in the range of SS allele was more frequent than SL allele frequency and LL allele being the least frequent, except for the control samples in CAG whose frequency was higher in the SL allele.

The distribution frequency of the SS allele for the CAG repeat was observed to be 42.98% among the POF population, 43.75% among the PA sample and 35.29% among the controls. The distribution of SS allele showed a significantly skewed pattern towards the control population over the (control/POF) was observed for the SL allele with \( p = 0.0231 \) and \( p = 0.0476 \), respectively, implicating its association with the aetiology of ovarian failure. We conclude that the AR gene mutations may play a role in the genetic cause of POF. Identification of the underlying genetic alteration of the AR gene is important for a proper diagnosis of POF subjects. It will be of potential interest to identify the downstream factors responsive to AR in the human ovary to understand the role of androgens and AR in the pathophysiology of POF.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

15. Westberg L, Baghaei F, Rosmond R, Hellstrand M, Landén M, Janson M, Holm G, Björntorp P, Eriksson E. Polymorphism of the androgen receptor gene and the estrogen receptor beta gene are associated with androgen...


