A genome-wide linkage scan in a Dutch family identifies a premature ovarian failure susceptibility locus

R.A. Oldenburg1, M.F. van Dooren1, B. de Graaf1, E. Simons1, L. Goyaerts1, S. Swagemakers2, J.M.H. Verkerk2, B.A. Oostra1 and A.M. Bertoli-Avella1,3

1Department of Clinical Genetics, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, the Netherlands; 2Department of Bioinformatics, Erasmus Medical Center, Rotterdam, the Netherlands
3Correspondence address. Tel: +31-10-7044628; Fax: +31-10-7044736; E-mail: a.bertoliavella@erasmusmc.nl

BACKGROUND: Premature ovarian failure (POF) is characterized by elevated gonadotrophins and amenorrhea before the age of 40 years and occurs approximately in 1% of women. POF etiology is highly heterogeneous with a wide spectrum of etiological pathogenic mechanisms including genetic causes. These mostly involve numerical, structural or monogenic defects on the X-chromosome. Mutations in a small number of autosomal genes (such as FOXL2 and NOBOX) have been identified as a cause of POF. However, in most cases, the disease underlying mechanisms are largely unknown.

METHODS: We performed a genome-wide linkage analysis in a relatively large Dutch family with seven patients suffering from POF, showing a dominant pattern of inheritance. A genome-wide analysis, using 50K single nucleotide polymorphism arrays, was combined with conventional parametric linkage analysis.

RESULTS: We identified three genomic regions on chromosomes 5, 14 and 18 yielding suggestive linkage (multipoint LOD score of 2.4 for each region). After inclusion of one elder unaffected family member, only the region on chromosome 5 remains as a putative POF locus. In addition, we investigated a second family (three living patients over three generations) for the regions on chromosome 5, 14 and 18. Haplotype analysis supported only the locus on chromosome 5q14.1–q15.

CONCLUSION: We performed the first genome-wide linkage search in familial POF and identified a region on chromosome 5q14.1–q15, which may harbor a novel POF susceptibility gene.

Keywords: premature ovarian failure; linkage analysis; familial; SNPs; genome-wide search

Introduction

The average age for menopause in Western populations of women is ~51 years. Premature ovarian failure (POF) can be defined as the development of amenorrhea due to cessation of ovarian function before the age of 40 years (Coulam et al., 1982).

The diagnosis is based on the elevated FSH levels in the menopausal range (usually above 40 IU/l) detected on at least two occasions a few weeks apart (Conway, 2000). POF affects ~1% of women under the age of 40 years and 0.1% under the age of 30 years (Coulam et al., 1986). Women with POF suffer from anovulation and hypoestrogenism and present with primary or secondary amenorrhea, infertility, sex steroid deficiency and elevated gonadotrophins (Kalantaridou et al., 1998). Also, it has been reported that an early loss of ovarian function has major health implications with a nearly 2-fold age-specific increase in mortality rate (Snowdon et al., 1989).

The etiology of POF is highly heterogeneous and a wide spectrum of pathogenic mechanisms may lead to its development, including autoimmune, metabolic (galactosaemia), infectious (mumps), iatrogenic (anticancer treatments) and genetic causes (Laml et al., 2000; Goswami and Conway, 2005). Genetic causes mostly implicate the X-chromosome, and range from a numerical defect, such as a complete deletion of one X chromosome (Turner syndrome) and trisomy X, to partial defects, such as deletions, isochromosomes, balanced X-autosomal translocations and monogenetic defects, such as heterozygous FMR1 premutations, POF1B and BMP15 mutations (Cronister et al., 1991; Schwartz et al., 1994; Partington et al., 1996; Zinn and Ross, 2001; Di Pasquale et al., 2004; Lacombe et al., 2006). Additionally, based on candidate gene approaches, an increasing number of autosomal candidate genes are being reported in the etiology of POF (Goswami and Conway, 2005). Most of those genes are associated with syndromic forms and mutations in non-syndromic POF are rarely found. Recently, mutations in the NOBOX gene (located at 7q35) were identified as a cause of non-syndromic POF (Qin et al., 2007). In most cases of POF, the underlying mechanisms are largely unknown.

The overall incidence of familial cases among women with POF seems to be ~4%, although there are conflicting data...
from various studies; in some studies this proportion seems to be as high as 30% (Cramer et al., 1995; Torgerson et al., 1997) while in others it is much lower (Vegetti et al., 1998). Most studies of affected families show an X-linked inheritance with incomplete penetrance (Coulam et al., 1983; Mattison et al., 1984; Christin-Maitre et al., 1998; Vegetti et al., 1998).

Here, we describe a relatively large Dutch family with POF, showing a dominant pattern of inheritance, with complete penetrance and possible anticipation, given that the subsequent generations are developing POF at an earlier age. We performed a genome-wide linkage analysis and identified a region on chromosome 5q14.1–q15, which may harbor a novel non-syndromic POF susceptibility gene.

Materials and Methods

Family selection
Two Dutch families with a total of 10 affected women were ascertained through the Department of Clinical Genetics at the Erasmus Medical Center in Rotterdam. Pathology reports and/or medical records were retrieved when available. Blood samples for DNA isolation and cytogenetic analyses were collected after obtaining written informed consent.

Fig. 1 depicts the pedigree of family EMC1, where seven cases of POF within two generations were identified, having a clear dominant pattern of inheritance. Fig. 2 represents the pedigree of family EMC2 where three cases of POF in three generations were identified. Although an autosomal dominant inheritance pattern is most likely, a dominant X-linked pattern cannot be excluded.

Genotyping
Genomic DNA was isolated from peripheral blood using a DNA purification kit (Gentra Systems). The genome-wide linkage search was conducted in 12 members of family EMC1 using the Affymetrix GeneChip Mapping 50K HindIII Array, containing 57 244 Single Nucleotide Polymorphism (SNP) markers. Samples were processed according to the manufacturer’s instructions (Affymetrix GeneChip Mapping Assay).

Arrays were stained, washed and scanned using an Affymetrix Fluidics Station 450 with images obtained by using an Affymetrix GeneChip 7G 3000 scanner. Affymetrix GCOS software v1.4, GTYPE software v4.1 and the dynamic model algorithm were used to derive SNP genotypes.

Linkage analysis and loci identification
PedCheck (v1.1) and Merlin (v1.0) were used to search for unusual patterns of Mendelian inheritance or unusual double recombinants consistent with potential genotyping errors. All SNPs showing inconsistency in transmission were removed from further analyses. The statistical package EasyLinkage Plus v5.02 (Hoffmann and Lindner, 2005) designed to perform automated linkage analyses using large-scale SNP data was used to perform all analyses. Allegro v1.2c software (incorporated in the EasyLinkage Plus package) was used to perform fully automated single point and multipoint parametric linkage analysis. As an obvious male phenotype for this disease does not exist, all male family members were incorporated in the analysis with an unknown affection status. LOD scores were obtained using a dominant model of inheritance, with a penetrance of 98%, a phenocopy rate of 0.01 and a disease allele frequency of 0.001. Allele frequencies of all genotyped SNPs were derived from the Affymetrix website based on Caucasian individuals previously genotyped. Map order and genetic inter-SNPs distances were taken from the Affymetrix website (Marshfield sex-averaged genetic map).

Currently available linkage software for multipoint analyses assumes that all markers are in linkage equilibrium. However, for closely spaced SNP markers, this assumption may not always be correct. Therefore, whole genome linkage analyses were performed with predefined spacing of 0.2–0.4 cM, in blocks of 90 and 100 SNPs. Then, single chromosomes showing positive linkage signals were independently analyzed under the same conditions and haplotypes were inferred using Allegro.

Figure 1: Pedigree of family EMC1. Squares indicate males and circles indicate females. Black symbols indicate affected women. Family members who are likely carriers are indicated with a question mark (?), whereas those who are likely unaffected are in white. An arrow points to the index case.
To facilitate inspection and analyses, graphical visualization of haplotypes was performed with HaploPainter v029.5 (Thiele and Nürnberg, 2005).

Microsatellite markers mapping to the ‘candidate’ genomic regions were selected based on their information content and were tested to confirm and where possible to decrease the size of the regions. PCRs were performed with M-13 tailed primers and addition of fluorescent labels (FAM and VIC). Products were run on an ABI Prism 3100 genetic sequencer (Applied Biosystems) and analyzed using the GeneMapper software v.3.0 (Applied Biosystems). Haplotypes were constructed based on the minimum number of recombinations.

**Ingenuity analysis**

Genes from the linkage area with corresponding Entrez id numbers were uploaded into the application (Ingenuity Systems, www.ingenuity.com). A pathway was created from the genes, of which information is present in the Ingenuity database, and genes expressed in ovary (28 genes), which were highlighted using the ‘Species and Tissues Highlight’ tool. Next, Ingenuity lists were made using the ‘search for genes/chemicals associated with functions/diseases’ tool and several other search terms, e.g. ovarian follicle; infertility. Separate lists were made for each search term and those lists were added (one by one) to the created pathway containing the genes from the region. Using the ‘Connect’ tool, we tried to connect the genes from the region with genes from the lists. Several genes from the region could be connected to genes from one or more lists.

**Results**

**Clinical findings**

A summary of the clinical findings is presented in Table I. In family EMC1 (Fig. 1), the index case (III-7) was diagnosed with POF at the age of 30 years. She has no cancer treatment history. DNA investigation showed no premutation in the \textit{FMR1} gene. In addition, she has a normal female 46,XX chromosomal pattern. She has two children, a son and a daughter. The daughter (IV-15) was diagnosed with POF at the age of 24. After 2 years of infertility and ovulation induction treatment, she had a daughter.

The only brother (III-8) of the index case has two daughters (IV-18 and IV-19), both diagnosed with POF at 15 and 14 years of age, respectively. The eldest daughter (IV-18) had her menarche at the age of 12 years. At 15 years of age, she developed amenorrhea. Chromosome analysis showed a normal female chromosomal pattern (46,XX). Antibodies against thyroid and adrenal gland were negative. The younger daughter (IV-19), born with bilateral clubfeet, had her menarche at 10 years of age. Three years later, she presented high FSH and low estradiol levels indicating POF. Chromosome analysis has not been performed. Antibodies against thyroid and adrenal gland were negative.

A sister (III-11) of the index case had her menopause when she was 50-year old. She has two healthy daughters. Another sister (III-13) was diagnosed with POF when she was 25-year old. She has one daughter (IV-22) who had her menarche at the age of 13. She presented elevated FSH levels at the age of 17. Diagnosis POF was established when she was 18-year old. A third sister (III-14) of the index case was investigated because of primary infertility. At an age of 27, the diagnosis POF was established. She did not have children. The mother (II-5) of these patients presented her menopause when she was 50-year old. None of the examined family members presented any obvious dysmorphic features.
The index case of family EMC2 (person III-1) developed POF at age 32 (Table I, Fig. 2). DNA was analyzed for a premutation in the FMR1 gene but no abnormalities were found. Her chromosomal pattern was 46,XX [48]/45,X[2] indicative of a possible Turner syndrome mosaicism. She did not give permission for a skin biopsy analysis. However, besides a small height (152 cm, −2.5 SD), she had no stigmata of Turner syndrome. In addition, her sister (III-2), diagnosed with POF at the age of 37, had a normal chromosomal pattern (46,XX) and a small height as well (158 cm, −2 SD). DNA investigation in individual III-2 showed no premutation in the FMR1 gene and antibodies against thyroid and adrenal gland were not present. Their mother (II-2) had her last pregnancy at the age of 39. At the age of 40, she was diagnosed with endometrial cancer, followed by a hysterectomy and ovariectomy. At the age of 59, she developed small cell bronchial carcinoma (with a smoking history). She died at the age of 60. From her medical records and medical history, it is not clear whether she had POF or not. Her mother (I-2) had developed POF at the age of 33. The medical records were not available. Again, there were no clear dysmorphic features associated with the occurrence of POF in this family.

**Genome-scan, linkage analysis and haplotyping**

We performed a genome-wide linkage analysis in 12 members of family EMC1 with seven women affected with POF using the 50K HindIII Affymetrix mapping arrays.

Overall 56 957 SNPs markers were available, from which 17 165 were uninformative (but included in the multipoint analyses). A total of 1707 SNPs (3%) were removed from the analysis due to unlikely genotypes or possible genotyping errors. After setting inter-marker spacing to 0.2, 0.3 and 0.4 cM, the EasyLinkage program selected 11 028; 8371 and 6725 SNPs for further linkage analyses, respectively.

When parametric multipoint linkage analysis was performed, genomic regions on three chromosomes, namely chromosome 5 (mLOD = 2.4; 86–103 cM), chromosome 14 (mLOD = 2.4; 92–118 cM) and chromosome 18 (mLOD = 2.4; 13–39 cM) displayed LOD scores above 2 (Fig. 3). All

---

**Table 1. Summary of the clinical data from affected family members of family EMC1 and EMC2.**

<table>
<thead>
<tr>
<th>Family</th>
<th>Individual</th>
<th>Age of amenorrhea</th>
<th>FSH U/l</th>
<th>Estradiol pmol/l</th>
<th>Karyotype</th>
<th>FMR1-premutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMC1</td>
<td>III-7</td>
<td>30</td>
<td>NMA</td>
<td>NMA</td>
<td>46,XX</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>III-13</td>
<td>25</td>
<td>31</td>
<td>&lt;10</td>
<td>46,XX</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>III-14</td>
<td>27</td>
<td>40</td>
<td>326</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>IV-15</td>
<td>24</td>
<td>NMA</td>
<td>NMA</td>
<td>46,XX</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>IV-18</td>
<td>15</td>
<td>34</td>
<td>&lt;200</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>IV-19</td>
<td>14</td>
<td>44</td>
<td>&lt;200</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>IV-22</td>
<td>18</td>
<td>63</td>
<td>17</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>EMC2</td>
<td>I-2</td>
<td>33</td>
<td>NMA</td>
<td>NMA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>III-1</td>
<td>32</td>
<td>NMA</td>
<td>NMA</td>
<td>46,XX[48]/45,X[2]</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>III-2</td>
<td>37</td>
<td>75</td>
<td>91</td>
<td>46,XX</td>
<td>No</td>
</tr>
</tbody>
</table>

NMA, no medical records available; NA, not analyzed; No, no carrier of a FMR1 premutation.

---

*Figure 3:* Results of the genome-wide linkage analysis per chromosome (X-axis). Multipoint parametric LOD scores (pLOD MPT) are in the Y-axis.
three regions showed very similar results regardless of the inter-marker spacing used in the analysis.

We then constructed haplotypes using Allegro on chromosome 5, 14 and 18. All informative SNPs in each chromosome were included.

Using the SNP data, we observed in each of the three regions a haplotype shared by all patients that was inherited from the grandfather (II-4). The grandmother (II-5) went through menopause at the age of 50. Individual III-8 (male) was carrying the same haplotypes that were transmitted to his affected daughters (IV-18 and IV-19).

On chromosome 5, recombinations in person III-7 between rs234910 and rs1316753 at 78.3 Mb and person IV-15 between rs261212 and rs38032 at 96.3 Mb delimited the shared disease haplotype (Fig. 4, SNPs are 0.3 cM spaced). When extensive haplotypes were made using all informative SNPs in the region, the area was reduced to an interval between rs10491456 (79.2 Mb) and rs440342 (95 Mb) (data not shown). In patient III-7, a recombination was observed on chromosome 14 (between rs7147709 and rs2282265 at 94.5 Mb). The haplotype extended until the telomere of the long arm of chromosome 14 (data not shown). On chromosome 18, recombinations occurred in person III-14 (between rs1662345 and rs238121 at 3.2 Mb) and person IV-19 (between rs206441 and rs616139 at 10.5 Mb) (data not shown).

Subsequently, microsatellite markers mapping to these areas were tested: on chromosome 5 from 81.9 to 112 cM (D5S424, D5S641, D5S644, D5S433), on chromosome 14 from 92.2 to 134.3 cM (D14S1035, D14S68, D14S65, D14S985, D14S92) and on chromosome 18 from 0 to 41.2 cM (D18S59, D18S63, D18S452, D18S464, D18S53). These markers were tested not only to independently confirm the results observed with the SNP analysis; but also to include all available individuals from this family and from family EMC2. We had a special interest in the genotype of individual II-2 from family EMC1 (Fig. 1). She is a healthy sister (88-year old, she had her menopause at age 55) of family member II-4 who carried the disease haplotype (reconstructed). Following an autosomal dominant pattern of inheritance, individual II-2 should not carry the same haplotype shared by all patients in the family.

After constructing haplotypes using the microsatellite data, the only region that remained as a candidate was 5q14.1–q15, since individual II-2 did not carry the ‘disease’ haplotype for this chromosome, as was the case in the other two regions on chromosomes 14 and 18. We observed that she had the same haplotype as individual III-11, the only unaffected daughter of individual II-4. Based on the SNP data, the shared haplotype extended 20 cM, delimiting a region of 15.8 Mb containing 70 genes (NCBI build 36.2).

**Figure 4:** Pedigree (EMC1) showing SNP haplotypes corresponding to the 5q14.1–q15 region. Rs numbers and physical location (Mb) are shown. Informative SNPs with a spacing of 0.3 cM have been used in this figure. Haplotypes corresponding to individuals II-4 and III-9 were reconstructed by the program (Allegro). A green (olive) bar identifies the disease haplotype. Recombinations occurring in III-7 and IV-15 are delimiting the candidate area.
Furthermore, we tested the same microsatellite markers in available individuals from family EMC2. Interestingly, all patients from the second family shared only on the 5q14.1–q15 locus identical haplotypes (Fig. 2).

Search for candidate genes from the linkage region on chromosome 5

We investigated whether any of the genes located at the chromosome 5 locus could be considered as a candidate gene for POF based on available expression and/or functional data. We searched in the Ingenuity database (Ingenuity Systems, www.ingenuity.com) and found information on 47 of the 70 genes located in the chromosome 5 locus (Supplementary Table 1, column 1, 2 and 3, online).

From these, 28 genes are expressed in ovary (Supplementary Table 1, column 4, online). We tried to find a relationship between any of these genes and known genes involved in pathways related to ovarian failure, female infertility, ovarian follicle and oocyte formation. Genes such as DHFR (dihydrofolate reductase), NR2F1 (nuclear receptor subfamily 2, group F, member 1), MEF2C (MADS box transcription enhancer factor 2, polypeptide C), CCNH (cyclin H), SSBP2 (single-stranded DNA binding protein 2) and CSPG2 (chondroitin sulfate proteoglycan 2) were highlighted in several of these analyses (indicated in bold Supplementary Table 1). Most of the genes are involved in transcription regulation or DNA-binding activity.

Discussion

Mutations in several genes, most of them located on chromosome X, are causing non-syndromic POF: FMR1 (POF1 locus, Xq26–q28), DIAPH2 (POF2A locus, Xq22), POF1B (POF2B locus, Xq21), FOXL2 (POF3 locus, 3q23) and BMP15 (POF4, Xp11.2). Recently, two missense mutations in the NOBOX (7q25) gene were identified as a rare cause of non-syndromic POF in a small subset of patients. It was already known that NOBOX mutations could cause syndromic-POF as part of the Blepharophimosus, Ptosis and Epicanthus Inversus Syndrome (Qin et al., 2007).

Most of the genes causally related to POF have been identified using the ‘candidate gene’ approach. Genome searches based on linkage analysis are practically absent in POF, probably due to the rarity of pedigrees with a large enough number of patients available for analysis.

We performed a genome-wide linkage analysis in a relatively large Dutch family with seven patients suffering from POF. To our knowledge, this is the first genome-wide linkage search performed in POF. We combined genome-wide linkage analysis using 50K SNPS arrays that offer dense genome coverage with conventional parametric linkage analysis. We initially identified three genomic regions on chromosomes 5, 14 and 18 yielding suggestive linkage (Lander and Kruglyak, 1995). After inclusion of one elder unaffected family member, only the region on chromosome 5 remains as a putative POF locus. This assumes that the mutation was transmitted from the great-grandparents (I-1 or I-2). The alternative is that II-4 carried a de novo mutation. In that case, any of the identified regions on chromosome 5, 14 or 18 could harbor the mutated gene. However, we also investigated a second (smaller) family with three living patients and one obligate carrier over three generations. Although the small number of patients limited the statistical power of the analysis, haplotype examination using microsatellite markers supported only the locus on chromosome 5.

In family EMC1 anticipation is suggested, as the age of onset of the disease is lower in successive generations. The mean age at onset of POF in generation IV is 17.8 years compared with 27.3 years in generation III. For family EMC2, this is less obvious as all three cases developed POF in their third decade of life. The observed anticipation in family EMC1 may actually be a bias of ascertainment, because anticipation is very easily mimicked by random variations in severity. However, the suggestion of anticipation in family EMC1 still might help in selecting candidate genes for sequencing. Any of the 70 genes containing trinucleotide repeats can be considered as a candidate, as anticipation is often due to unstable expanding of such repeats.

Functional analysis using the program Ingenuity indicated a number of genes with a possible role in pathways related to ovarian failure. This will help in the selection of genes for further sequence analysis to identify the causal mutation(s).

Our results indicate a region on chromosome 5q14.1–q15 which may explain a small subset of non-syndromic POF cases with an autosomal dominant inheritance pattern.

More studies are needed to replicate and confirm this new locus and reduce the region for gene sequencing.

Supplementary Data

Supplementary data are available at http://humrep.oxfordjournals.org.

Acknowledgements

We thank Tom de Vries-Lentsch for producing the artwork.

Funding

The study was financially supported by The Center for Biomedical Genetics (CBG) and the Center for Medical Systems Biology (CMSB).

References


Submitted on February 15, 2008; resubmitted on April 4, 2008; accepted on April 17, 2008