Disponible en ligne sur ScienceDirect
www.sciencedirect.com
Elsevier Masson France
EM|consulte
www.em-consulte.com
Annales d'Endocrinologie 71 (2010) 218–221
Annales d'Endocrinologie
Annals of Endocrinology

Journées Klotz 2010

Genome-wide association study and premature ovarian failure

Insuffisance ovarienne prématurée et études d'association au génome

S. Christin-Maitre [a,*], G. Tachdjian [b]

[a] *Reproductive Endocrine Unit, Hôpital Saint-Antoine, ER9, Université Paris VI, 75012 Paris, France*
[b] *Inserm U935, Embryology and Cytogenetic Department, Hôpital A.-Béclère, Université Paris XI, 92140 Clamart, France*

Available online 2 April 2010

Presented by Jacques Young

Résumé

L'insuffisance ovarienne prématurée (IOP) est définie par une aménorrhée durant plus de quatre mois associée à une ascension des gonadotrophines habituellement au-dessus de 20 mUI/ml survenant chez une femme avant l'âge de 40 ans. Un certain nombre de gènes candidats ont été identifiés au cours des 15 dernières années comme *FOXL2*, *FSHR*, *BMP15*, *GDF9* et la prémutation *Xfra*. Cependant, l'étiologie de l'IOP demeure inconnue dans plus de 90 % des cas. La première stratégie pour identifier des gènes candidats en dehors de l'étude des gènes impliqués dans les insuffisances ovariennes des modèles animaux repose sur l'analyse des délétions et des translocations du chromosome X chez les patientes. La deuxième stratégie repose sur des analyses de liaison et la troisième sur des études d'hybridation comparative du génome (*comparative genomic hybridization* [CGH] *array*). La dernière stratégie repose sur des études d'association du génome (Genome-Wide Association Studies [GWAS]). Cette technique consiste à rechercher les polymorphismes nucléotidiques uniques (SNPs) chez les patients et des témoins. Jusqu'à maintenant, trois études ont été réalisées et ont identifié différents loci potentiellement liés à l'IOP, comme le PTHB1 et le ADAMTS19. Cependant, des réplications dans des cohortes indépendantes sont nécessaires. Les études de type GWAS sur de larges cohortes de femmes présentant une IOP devraient permettre d'identifier de nouveaux gènes candidats dans un futur proche.

Mots clés : Insuffisance ovarienne prématurée ; *FOXL2* ; *FSHR* ; *BMP15* ; *GDF9* ; *Xfra* ; Délétion du chromosome X ; Translocation du chromosome X ; CGH ; GWAS

Abstract

Premature ovarian failure (POF) is defined as an amenorrhea for more than 4 months, associated with elevated gonadotropins, usually higher than 20 mIU/ml, occurring in a woman before the age of 40. Some candidate genes have been identified in the past 15 years, such as *FOXL2*, *FSHR*, *BMP15*, *GDF9*, *Xfra* premutation. However, POF etiology remains unknown in more than 90% of cases. The first strategy to identify candidate gene, apart from studying genes involved in ovarian failure in animal models, relies on the study of X chromosome deletions and X;autosome translocations in patients. The second strategy is based on linkage analysis, the third one on Comparative Genomic Hybridization (CGH) array. The latest strategy relies on Genome-Wide Association Studies (GWAS). This technique consists in screening single nucleotide polymorphisms (SNPs) in patients and controls. So far, three studies have been performed and have identified different loci potentially linked to POF, such as PTHB1 and ADAMTS19. However, replications in independent cohorts need to be performed. GWAS studies on large cohorts of women with POF should find new candidate genes in the near future.

Keywords: Premature ovarian failure; *FOXL2*; *FSHR*; *BMP15*; *GDF9*; *Xfra* premutation; X chromosome deletions; X ;autosome translocations; CGH; GWAS

Premature ovarian failure (POF) is defined as an amenorrhea during more than 4 months associated with elevated gonadotropins, usually higher than 20 mIU/ml, occurring in a woman before the age of 40. Its prevalence is 1–2% of women

* Corresponding author.
E-mail address: sophie.christin-maitre@sat.aphp.fr (S. Christin-Maitre).

0003-4266/$ – see front matter © 2010 Published by Elsevier Masson SAS.
doi:10.1016/j.ando.2010.02.014

[1]. Its etiology may be linked to chemotherapy and/or radiotherapy, infectious diseases or surgery. Genetic causes might be involved. Some candidate genes have been identified in the past 15 years, such as *FOXL2*, *BMP15*, *GDF9*, *Xfra* premutation [2]. However, POF etiology remains unknown in more than 90% of cases. Hypothetically, gene mutations or gene variants could be involved in the occurrence of POF.

In the past years, cytogenetic studies have identified in some patients with POF, abnormalities in chromosome X numbers, such as X monosomy or 45,X/46,XX mosaicism. Furthermore, X chromosomal deletions, X duplications or X;autosome translocations have been identified in different patients. Those cases have identified two main regions on the X chromosome, named POF1 and POF2, located in Xq26.2-q28 and Xq13.3-q22, respectively [3]. Several groups have characterized different breakpoints and have identified different candidate genes in POF women, such as the human homologue of dia [4] and progesterone receptor membrane component-1 (PGRMC1) in a family with X;autosome translocation [t(X;11)(q24;q13)] [5].

A second strategy in order to identify candidate gene in POF women has been to perform genetic linkage combined with positional cloning, calculating LOD scores [6]. This strategy has identified gene mutations involved in monogenic disease. Those mutations enormously increase the risk of the disease. However, POF is often a sporadic and multigenic disease. Indeed, the incidence of familial cases is reported to be between 4 and 31%. However, thanks to this strategy, Aittomäki et al. have identified in 1995, the FSH receptor as a candidate gene in Finnish families. In those large families, containing at least three cases of primary amenorrhea, Aittomäki et al. showed a linkage with the short arm of chromosome 2 [7], and secondly identified a point mutation in the gene coding for FSH receptor. Several years later, a sibling study was performed in Dutch families. Women tested had very concordant or very discordant age of menopause, within the family. A total of 165 families were analyzed. This study has shown links between menopausal age and the short arm of chromosome X, in Xq21.3 (LOD score 3.1) and chromosome 9q21.3 (LOD score 2.6) [8].

A third stategy relies on array comparative genomic hybridization (array CGH). CGH is a modified fluorescence in situ hybridization technique [9]. Differentially labeled studied and reference DNA segments are cohybridized to normal metaphase spreads or to mapped DNA sequences or oligonucleotides arrayed onto glass slides (array CGH). To perform array CGH, PCR amplified bacterial artificial chromosomes (BACs), cDNAs or synthetic oligonucleotides were spotted as elements on the array. These arrays contain 3000 to 500,000 mapped reporters instead of whole chromosomes. These reporters have multiple biological annotations such as chromosomal location, sequence information, gene name, biological and molecular function. In array CGH, the resolution is, first of all, dependent on the platform used. The higher the number of targets, the higher the potential resolution will be. Thus, the resolution of array CGH is determined by the size and number (density) of sequences spotted onto the array [10]. The study and reference DNA segments are cut into small fragments and labeled by different fluorochromes to distinguish between the sequences hybridized in the two genomes. The hybridization of both types of genomic DNA is conducted simultaneously in order to induce competition between the studied DNA and the reference DNA for specific DNA sequences. Using fluorescence microscopy and digital image analysis, the relative amounts of studied and reference DNA hybridizing to the target DNA are estimated by measuring the green and red fluorescence, at each point. In order to determine whether a chromosomal loss or gain is present, the fluorescence ratio of the green to red intensities is calculated.

Recent advances in CGH array have enabled high resolution comparison of whole genomes for the identification of genetic alterations in cancer and other genetic diseases. We used a resolution of 1 Mb in order to test a patient with X chromosome deletion. This microarray CGH showed that the del(X)(q21.31) was also associated with a Xpter duplication including the SHOX gene [11]. In a second study, we tested 99 POF patients [12]. Eight statistically significantly different CNVs have been identified in chromosomal regions 1p21.1, 5p14.3, 5q13.2, 6p25.3, 14q32.33, 16p11.2, 17q12, and Xq28. In the eight statistically significant CNVs we reported, we found five genes involved in reproduction, thus representing potential candidate genes in POF.

A new strategy is to perform Genome-Wide Association Studies (GWAS). A GWA study is defined by the National Institute of Health as a study of common genetic variation across the entire human genome designed to identify genetic associations with observable traits [13]. The GWA approach is revolutionary because studies of the entire human genome can be performed at very high levels of resolution, in thousands of unrelated individuals, unconstrained by prior hypotheses regarding genetic associations with disease [14]. Those types of analysis consist in screening single nucleotide polymorphisms (SNPs) in patients and controls. They have a much higher resolution than CGH array. They usually test the whole genome. SNPs are the most common genetic variations in the human genome. A single-nucleotide difference between two homologous chromosomes occurs when two different deoxyribonucleic acid (DNA) bases are at the same locus in the two chromosomes. The difference is called a polymorphism if, in the pool of chromosomes in an interbreeding population, both of these different bases are found in at least 1% of the chromosomes. Nearly 12 million unique human SNPs have been catalogued in the Human Genome Single Nucleotide Polymorphism Database (dbSNP) [15]. Thus, these genetic polymorphisms are useful as genetic markers in identifying the disease variants via linkage disequilibrium (LD). This large number of SNPs in the human genome provides the highest density of genetic markers. The human genome can be organized into haplotypes with strong LD among the SNPs. SNPs that are located adjacent to each other within a genomic region, tend to be inherited together more frequently than expected in a block pattern. The first arrays tested hundreds of thousands of SNPs. The arrays used nowadays are able to test millions of SNPs, therefore increasing the resolution of the analysis. They are called high-density microarrays and rely on high-density oligonucleotide genotyping platforms. This screening may be performed in familial cases or in large cohort of patients. The