Mapping of the POF1 locus and identification of putative genes for premature ovarian failure

Rina M.Davison1,3, M.Fox2 and G.S.Conway1

1Cobbold Laboratories, Division of Endocrinology, Department of Medicine, University College London, WC1N 8AA, and 2Department of Biology, The Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK
3To whom correspondence should be addressed

We have identified a breakpoint on the X chromosome which is associated with premature ovarian failure (POF). Using polymerase chain reaction (PCR) probes of polymorphic microsatellites and fluorescent in-situ hybridization (FISH), this breakpoint has been narrowed to a region of 300 kb spanned by two P1 artificial chromosomes (PAC). Computer exon prediction and gene homology programs revealed three genes in this area. Our results suggest that two of these genes, HS6ST and E2F, and LINE 1 elements may be involved in ovarian development. Interruption of these genes could be the cause of POF. This study demonstrates how various molecular techniques and bioinformatic searches can complement each other in order to solve a clinical problem.

Keywords: E2F/HS6ST/long interspersed nuclear elements/premature ovarian failure

Introduction

Premature ovarian failure (POF), defined as the cessation of menses for >6 months and associated with elevated gonadotrophin concentrations, before the age of 40 years (Coulam et al., 1986) affects 1% of women. The aetiology of POF is usually obscure. A genetic contribution is suggested by the occurrence of families with many affected women and an association with X chromosome defects (Therman et al., 1990). Two specific regions on Xq have been defined as POF loci: POF1 Xq26-qter (Tharapel et al., 1993), and POF2 Xq13.3–Xq21.1 (Powell et al., 1994). It appears that distal deletions involving the POF1 locus result in POF at ages 24–39 years, (Krauss et al., 1994). It appears that distal deletions involving the POF2 locus cause POF at an earlier age of 16–21 years (Powell et al., 1994).

Complete absence of one X chromosome, as in the case of Turner’s syndrome, karyotype 45XO, results in short stature and ovarian dysgenesis. Ovarian dysfunction in Turner’s syndrome may be the result of a lack of diploid dosage of one or more vital genes, both alleles of which are active in oogenesis (Zinn et al., 1990). Haploinsufficiency of POF genes may explain why deletions and translocations causing disruption of genes at POF1 and POF2 loci affects ovarian function.

At the POF1 locus, there is an association between POF and the fragile X premutation (50–200 trinucleotide repeats in the 5′ prime untranslated region of the FMR1 gene at Xq27.3). Studies have demonstrated that 13–25% of fragile X carriers experience POF (Allingham-Hawkins et al., 1999, Uzielli et al., 1999), and conversely 3–15% of women with idiopathic POF harbour a fragile X premutation (Conway et al., 1998; Uzielli et al., 1999). However the exact molecular role of the FMR1 gene in reproduction is as yet undetermined.

At the POF2 locus, 11 breakpoints were mapped to 15 Mb region at Xq21 (Sala et al., 1997), and came to the conclusion that defects in as many as eight genes would be required to explain the association with POF over such a large area. A gene (DIA) interrupted by a breakpoint in a family with POF was identified as a possible candidate for a POF2 gene (Bione et al., 1998). In the current study, the characterization of a breakpoint at the POF1 locus is presented.

We recently identified a family (Davison et al., 1998) in which POF was transmitted through three generations in association with the karyotype 46X, delX(q26), between microsatellites DXS994 and DXS1062. The age range of onset of amenorrhoea in six affected females in this family was 28–34 years. The index case was 26 years at the time of presentation and her mother, who experienced POF at the age of 28 years, shared the deletion. In view of this genetic evidence, the index case was advised that menopause may occur imminently and she chose to undergo IVF with embryo freezing. We postulated that a putative POF1 gene was disrupted at the breakpoint in this family. It might be that genes below the breakpoint are also deleted in this family and hence may be the cause of POF. However, evidence against this hypothesis (Therman et al., 1990), showed that nine out of 12 patients with a breakpoint at Xq26-28 had normal ovarian function, implying that specific genes are interrupted at breakpoints to cause POF. Another author (Sala et al., 1997) reported a normal female with a breakpoint in the POF2 locus suggesting that not all breakpoints in the region would produce POF, only those where POF genes are disrupted.

Materials and methods

Eight polymorphic markers were identified which lie within the breakpoint region identified by cytogenetics. Polymerase chain reac-
tion (PCR) with fluorescent labelling of these markers and an automated laser DNA analyser was used to narrow the breakpoint interval from 7 Mb to a region of 500 kb. Eight P1 artificial chromosomes (PAC) were found which spanned the interval identified by the polymorphic markers. Fluorescent in-situ hybridization (FISH), using PAC spanning this smaller region, was used to narrow the breakpoint to two PACs (200 kb). Computer analysis of these two PACs was performed in order to predict the location of exons of putative POF genes.

**Polymerase chain reaction**

Genomic DNA from the index case and her parents was prepared. For each of the microsatellites (Table I) the forward primer from each set was labelled with a fluorescent dye (FAM; Perkin Elmer, Warrington, UK). The PCR products were separated and analysed using an automated laser DNA analyser (ABI Prism 310; Perkin Elmer) and the appropriate software (Genescan version 2.0.2 and ALF Fragment manager V1.1). With the Prism 310, 1 µl of each fluorescent PCR product mixed with 12 µl of formamide and 0.5 µl of size standard (Genescan 500-TAMRA) were run through a capillary (15 kV, 24 min at 60°C). The sizes of the specific PCR products were determined according to the position of fluorescent peaks. Different peaks represented alleles at the same locus.

**FISH mapping**

FISH was carried out by modifications of a previously described technique (Pinkel et al., 1986). PAC DNA was labelled with biotin-14-dATP by nick translation (Bionick kit; Gibco-BRL Life Technologies, Uxbridge, UK) and the labelled probe was precipitated together with Cor1 DNA and herring sperm DNA. This was then resuspended in hybridization mixture containing 50% formamide, 10% dextran sulphate and 2× SSPE, pH7. Human metaphase chromosomes were obtained from lymphocyte cultures synchronized by addition of sodium citrate (SSC) for 5 min at 75°C. The sizes of the speciﬁc hybridization peaks were determined according to the position of fluorescent peaks. Different peaks represented alleles at the same locus.

**Results**

The PCR results are shown in Table I. Two alleles are present above the breakpoint, whereas below this point is a single allele. Hence it is evident that the breakpoint was between microsatellite DXS692 and DXS1254, and that the female index case inherited the paternal intact X chromosome and maternal X(del).

Eight PACs span this interval, and each was used to perform FISH. The breakpoint lies at the point where a double hybridization signal from both normal and deleted X chromosomes, switches to a single signal only on the normal X and this was found to be between PACs 97K10 (156000 bp) and 358H7 (184974 bp). Representative metaphase spreads are shown in Figure 1. Although FISH is not quantifiable, the signal of PAC97K10 on the X(del) was always less intense when compared with that on the normal X, (Figure 1b). Separate batches of this PAC were used and the phenomenon was reproducible. This might imply that PAC 97K10 was bridging the breakpoint and so only part of it was able to hybridize.

The results of NIX and BLAST analysis of the two breakpoint PACs revealed four genes. Two exons immediately adjacent to the breakpoint corresponded to two genes, heparan sulphate-6-sulphotransferase (HS6ST) and an E2F-related transcription factor. The EST database searches for homology to our genomic sequence also revealed matches with human and Cricetulus griseus HS6ST (2051 bp) sequences. The PAC
genomic sequence had an 82% nucleotide sequence homology and 65% amino acid homology with HS6ST. BLAST analysis of HS6ST with Drosophila genes revealed a nucleotide match with egl (59% identity, 59% similarity). Also the PAC 435D1, which is centromeric to PAC 97K10 had a 75% nucleotide sequence match and a 89% amino acid match with human HS6ST. BLAST analysis of the second exon region revealed a 95% nucleotide match and a 73% similarity with the protein sequence of the human single exon gene (1440 bp) human E2F related transcription factor.

Between exons 2 and 3 of HS6ST, BLAST analysis of another predicted exon region corresponds to long interspersed nuclear elements (LINE 1) elements with a 97% match over 2000 bp with LINE 1 elements 2,5,6,8,12,14,15,19,20,25,33 and 39 (which are homologous) – a LINE 1 element cluster.

Finally BLAST analysis of an exon on PAC 97K10 centromeric to the HS6ST exon revealed a 92% match over 200 bp with a partial sequence of the human GPC3 gene.

Discussion

We have characterized the organization of genes at the breakpoint in this family with POF, and ascertained that there are three genes: HS6ST, E2F, GPC3 that may be interrupted at the breakpoint (Figure 2). In addition LINE 1 elements are inserted within this region.

Heparan-sulphate–6-sulphotransferase (HS6ST)

HS6ST catalyses the transfer of sulphate from 3′-phophoadenosine 5′-phosphosulphate to position 6 of the N-sulphoglucosamine residue of heparan sulphate. The complete human cDNA has been cloned and partially cloned in Chinese hamster ovary cells (Habuchi et al., 1998).

Heparan sulphate proteoglycans (HSPG) are ubiquitously present on the cell surface and in the extracellular matrix. They are known to interact with a variety of proteins such as heparin binding growth factors, extracellular matrix components, protease inhibitors and lipoprotein lipase. HSPG are implicated in morphological regulation during development, cell proliferation, differentiation, adhesion, migration, and also physiological phenomena such as viral infection and blood coagulation.

Heparan sulphate is required for high affinity binding of fibroblast growth factors to their receptors, as well as binding to other cell growth factors. It is postulated that different kinds of sulphotransferases affect the fine structure of heparan sulphate, in turn affecting morphological regulation and cell growth factors. With reference to the ovary, it is possible that vital follicular growth factors are inhibited through disruption of the HS6ST gene and its effect on heparan sulphate ligand binding.

A particular role for sulphation of proteoglycans has been shown in the Drosophila oocyte (Sen et al., 1998). An enzyme similar to HS6ST has been expressed in follicle cells, and plays a pivotal role in the mechanism that establishes oocyte and furthermore embryonic polarity. A parallel role for HS6ST may occur within the human oocyte. The match of HS6ST with the Drosophila gene, egl, is intriguing, as it appears this gene is required for oocyte microtubule polarity. Immunoprecipitation experiments show that egl protein localizes to the oocyte at all stages of oogenesis from oocyte differentiation to its requirement for axis determination of the embryo (Mach et al., 1997).

E2F-related transcription factor

The 1440 bp mRNA sequence of E2F(DP1) showed a 75% match with PACdJ358H7 and a high similarity with the 408 amino acid sequence (70% identity, 73% similarity). The E2F family of mammalian transcription factors plays a critical role in the expression of genes that are required for cell cycle progression. They interact with proteins and bind to DNA as heterodimers. The E2F complex is capable of activating transcription and initiating apoptosis, as shown in fibroblasts, myeloid cells and glioma cells. In fact, the massive apoptosis that occurs when E2F expression is induced in glioma cells is now being considered as a therapeutic strategy for cancer gene therapy (Gomez et al., 1997).
Recently, a specific role for E2F in the *Drosophila* ovary has been described (Royzman *et al*., 1999). E2F regulates the expression of ORC2 protein (Origin Recognition Complex) in follicle cells, such that mutations in E2F cause female sterility by affecting cell cycle transition and chorion gene amplification. E2F is a single exon gene and it may be that the breakpoint interrupts its coding sequence or promoter. With regards to the ovary, POF may be caused by accelerated apoptosis of germ cells. On the other hand, E2F factors are neither ovary nor X-specific, and their effects are ubiquitous.

**Long interspersed nuclear elements (LINEs or L1)**

LINE 1 elements are mobile elements that occupy ~15% of the genome and are concentrated in AT rich regions (Kazazian *et al*., 1998). These retrotransposon elements are sequences that can be transcribed into RNA, reverse transcribed into cDNA and then reintegrated as such into the genome at a new location. LINE 1 insertions were first recognized as potential causal agents of human disease in 1988 when two separate insertions of truncated LINE 1s were found to disrupt the factor VIII gene, resulting in haemophilia (Kazazian *et al*., 1988). Since then at least six other retrotransposed LINE 1 insertions have been described: another in the factor 8 gene (Woods-Samuels *et al*., 1989), three in the Duchenne muscular dystrophy gene (Narita *et al*., 1993; Holmes *et al*., 1994), one in APC (Miki *et al*., 1992) and the other in the β globin gene. The insertions usually occurred in the germ line or during early development, with six of the eight L1 insertions appearing in genes located on the X chromosome. This apparent predilection for the X chromosome may be due to the hemizygosity of X-linked genes in males. Since L1 insertions occur in the germline and are scattered on the X, the clustering of L1s around the breakpoint may have a similar significance in the human, by inducing mutations and diminishing fertility.

Alternatively, LINE 1 elements may have a direct role in the ovary. *Drosophila melanogaster* I factor belongs to the class of LINEs and actively transposes in the germline of the female progeny from crosses between reactive females and inducer males. This process of L1 insertion is known as I-R hybrid dysgenesis and gives rise to reduced fertility, increased frequency of mutations and X chromosome loss (Dimitri *et al*., 1997). Recent studies (Goto *et al*., 1999) have demonstrated that LINE elements are differentially expressed in human female primordial germ cells (as compared to male) at the time of entry into meiosis, suggesting a role in ovarian development.

**Glypican 3 (GPC3)**

GPC3, a member of the glypican family, is an 8 exon gene spanning 500 kb, which has been mapped to human Xq26 by FISH. GPC3 is expressed ubiquitously in the embryo but is restricted to the colon and ovary in the adult. It has been shown (Lin *et al*., 1999) that GPC3 induces apoptosis in MCF-7 breast cancer cells *in vitro*, and it may also function as a tumour suppressor in the ovary. However there is no evidence that GPC3 is involved the regulation of apoptosis in the normal ovary.

Mutations in the GPC3 gene are responsible for Simpson dysmorphia syndrome (Simpson *et al*., 1975) and are inherited as X-linked recessive. The phenotypic effects of this syndrome are a broad stocky appearance, large protruding jaw, widened nasal bridge, enlarged tongue, short hands and fingers together with normal intelligence, or, in more severe cases cleft palate, cardiac anomalies and mental retardation. With regards to ovarian effects, a case report of affected females did not mention any adverse effects (Pilia *et al*., 1996). A mother who was heterozygous for a GPC3 mutation did not have ovarian dysgenesis, although whether an early menopause ensued is unknown (Golabi and Rosen 1984). Although GPC3 fills the criteria for being a large multi-exon gene spanning a region known to be in the POF1 locus, women with POF do not show the phenotype of GPC3 mutations.

**Conclusions**

The aim of this study was to demonstrate the various techniques available to isolate candidate genes at a time when the sequence of the human genome rapidly unfolds. In summary we have identified three genes adjacent to a breakpoint at the POF1 locus, as well as a LINE 1 element cluster. Amongst these, we currently feel that HS6ST is the strongest candidate gene for POF. It is a gene which spans across four exons and hence is likely to be interrupted by a number of translocations in addition to deletions, as in this index family. In addition there...
is a strong similarity with the Drosophila gene egl which has a role in fertility as well as the evidence for sulphotransferase activity in Drosophila oocytes.

Identification of POF genes would enable us to understand the mechanisms of ovarian development. On a practical level, the ability to provide an explanation to the mechanism of POF is therapeutic. Further, in our particular family the gross deletion at Xq26 makes it possible to predict POF prior to the onset of infertility. However mutation screening might reveal smaller defects in a POF1 gene which would be equally informative in women with familial or sporadic POF who are cytogenetically normal.

Acknowledgements
We thank the Sanger Centre for the supply of PACs. Dr Davison is supported by the Wellcome Trust.

References


Received on September 27, 1999; accepted on January 17, 2000