FEBS 27903 FEBS Letters 555 (2003) 533–538

Identification of a new expanding family of genes characterized by atypical LRR domains. Localization of a cluster preferentially expressed in oocyte

Sébastien Dadé[a], Isabelle Callebaut[b], Pascal Mermillod[a], Philippe Monget[a,*]

[a]Physiologie de la Reproduction et des Comportements, UMR 6073 INRA-CNRS-Université François Rabelais de Tours, 37380 Nouzilly, France
[b]Systèmes Moléculaires et Biologie Structurale, LMCP, CNRS UMR 7590, Universités Paris 6 et Paris 7, case 115, 4 place Jussieu, 75252 Paris Cedex 05, France

Received 11 September 2003; revised 10 November 2003; accepted 10 November 2003

First published online 21 November 2003

Edited by Takashi Gojobori

Abstract **In the present work, we have used the in silico subtraction methodology to identify novel oocyte-specific genes in the mouse. By this way, we have identified in silico a new family of genes composed of more than 80 members. Sequence analysis showed that these genes belong to the superfamily of leucine-rich repeat (LRR) proteins. However, LRRs of this family display some variability in length and in amino acids composition within the β-strands region, as more leucine residues are substituted by other hydrophobic amino acids as compared to canonical LRRs. Interestingly, for nine of these genes, the ESTs were represented almost exclusively in mouse egg libraries. Three of them were selected for experimental study. By RT-PCR and in situ hybridization, we confirmed their specific expression in the mouse oocyte from primary to preovulatory follicles. These three genes are localized in a cluster on mouse chromosome 4, in the vicinity of another recently discovered oocyte specific gene called oogenesin, that we also found to belong to the same family. We thus re-named this latter gene 'oogenesin-1', and the three genes identified here were named oogenesin-2, -3 and -4.**
© 2003 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Key words: Oocyte; Mouse; Expression; Leucine-rich repeat domain; Oogenesin

1. Introduction

Several genes that are specifically expressed in oocytes have recently been identified by mRNA differential display [1] or using in silico subtraction [2]. As shown by knock-out experiments in the mouse, most of these genes play key roles in oogenesis, folliculogenesis or early embryonic development. In particular, GDF-9 and BMP-15 are necessary for folliculogenesis beyond primary follicles in the mouse and the sheep, respectively [3,4]. Moreover, maternal factors such as maternal antigen that embryo required (Mater), Zygotic arrest 1 and nucleoplasmin 2 are required for normal embryonic development beyond the 1- or 2-cell stage (for review: [5,6]). In the present work, we have used the in silico subtraction methodology to identify novel oocyte-specific genes in the mouse.

2. Materials and methods

2.1. In silico identification of oocyte-specific genes

Three cDNA libraries derived from mouse unfertilized eggs (dbEST library ID.14142), 2-cell egg (dbEST library ID.5391) and in vitro fertilized eggs (dbEST library ID.2589) were submitted to digital differential display (DDD) analysis (http://www.ncbi.nlm.nih.gov/UniGene/info/ddd.html) to identify oocyte-specific genes that are not found in several non-normalized cDNA libraries from different adult somatic non-tumoral tissues (brain, kidney, stomach, liver, lung, spleen, muscle, heart, skin, bone marrow, adipose tissue, adrenal gland).

2.2. Sequence analysis

Similarity searches were performed within protein sequence databases using PSI-BLAST [7]. Domain databases (Smart [8], Pfam [9]) were also searched. Comparison of protein sequences was refined using Hydrophobic Cluster Analysis (HCA) [10]. This two-dimensional representation of protein sequences adds to the comparison process information on secondary structures, as hydrophobic clusters defined in such a way mainly correspond to the inner faces of regular secondary structures. The corresponding clusters are much better conserved than sequences themselves and constitute signatures helping the comparison of remote sequences. Figures representing alignment of protein sequences and three-dimensional structure of proteins have been prepared using the ESPript software [11] and SwissPDBviewer [12], respectively.

2.3. RT-PCR analyses

Total RNA was extracted from whole adult tissues (ovaries, testis, liver, kidney, spleen, stomach, ganglion, brain, muscle, skin, lung and heart) using RNAble reagent according to the manufacturer's procedure (Eurobio, Les Ulis, France). Reverse transcription was performed for 1 h at 42°C in a total volume of 25 μl with 2 μg total RNA per sample following standard procedure. Five μl of the cDNA product was amplified by PCR using primers described in Table 1. RT-PCR products were also analyzed by Southern blotting. Briefly, the RT-PCR products were fractionated on 1% agarose gel, transferred to Hybond-N+ membrane (Amersham-Pharmacia) and hybridized with the corresponding cDNA fragment labeled by random priming (1×10^6 cpm/ml) as described previously [13].

To insert cDNAs in cloning vector, PCR products were purified from the agarose using the gel extraction kit QIAEX II (Qiagen, Hilden, Germany) and inserted into pGEM-T vector (Promega, Madison, WI, USA). The selected clones were verified by sequencing.

2.4. In situ hybridization

Frozen ovaries from nine female mice in di-estrus and nine female mice in estrus were serially sectioned (10 μm) with a cryostat to perform in situ hybridization experiments using ^{35}S-labeled cRNAs probes as previously described [14]. Specificity of hybridization was assessed by comparing the signals obtained with the cRNA antisense and the corresponding sense probes. Histological determination of follicular size and degree of atresia was performed on adjacent sections stained with Feulgen [14]

*Corresponding author. Fax: (33)-2-47 42 77 43.
E-mail address: monget@tours.inra.fr (P. Monget).

0014-5793/03/$22.00 © 2003 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/S0014-5793(03)01341-3

Table 1
Primers used for RT-PCR amplification of cDNA fragments of oogenesin-2, -3, -4 and actin

Gene	Primers	Annealing temperature
Oogenesin-2	CTGGCTATGGGACAAAGAAC	55°C
	AAGGGGCTCAAGAAGTAAAT	
Oogenesin-3	TCACAGATTCCCTCAGTATG	54°C
	GCATTTTTATTGTTTATCTCA	
Oogenesin-4	CTGACTGTGGGGTGAATAAA	59°C
	GGATGTAGCAAATGTGACTC	
Actin	ACGGAACCACAGTTTATCATC	60°C
	GTCCCAGTCTTCAACTATACC	

3. Results and discussion

3.1. In silico identification of oocyte-specific genes

By using the DDD software, we found in silico 171 ESTs that appeared specifically expressed or highly enriched in the mouse egg. We systematically verified the specificity of expression of all of these genes by searching in UniGene (http://www.ncbi.nlm.nih.gov/UniGene/) the cDNA sources of all of their ESTs, and finally selected 64 genes for which the main if not exclusive cDNA source was one of the three libraries used in the DDD analysis. Several well known oocyte-specific genes were found in this list, validating this methodology: Dazl, GDF-9, BMP-15, Mater, Obox1, Zfp393, Rfpl4 and Mos. Interestingly, among the other genes, nine (Mm.7786, Mm.7836, Mm.21115, Mm.25370, Mm.26222, Mm.28722, Mm.31283, Mm.148643 and Mm.208554) were shown to be structurally similar. By using standard PSI-BLAST analysis (http://www4.ncbi.nlm.nih.gov/BLAST/), we found that they belong to a new family of genes 'similar to the Pramel1 gene'. This family is composed of more than 80 genes that did not contain any known functional domain (see below and supplemental data). Using data available on the mouse genome (http://www.ncbi.mapviewer), the Mm.7786 and Mm.21115 sequences were assigned to chromosome 5. Interestingly, Mm.25370, Mm.26222, Mm.28722, Mm. 31283, Mm.148643 and Mm.208554 were positioned on chromosome 4 as a cluster of genes just near Pramel1 (Fig. 1). The Mm.7836 gene is also located on chromosome 4 but not in the same locus (not shown).

For further experimental study, we selected Mm.25370 (accession number BC052839.1), Mm.28722 (accession number BC053700) and Mm.148643 (accession number AK054346.1) based on the high representation of their ESTs in oocyte libraries and the vicinity of their localization in the mouse genome, near the Pramel1 gene (see below). During the course of this study, Minami et al. identified by differential display a novel oocyte-specific gene, named 'oogenesin' [15]. In fact, we show here that oogenesin is only one member of the family of genes characterized in the present work, but is clearly distinct from the three genes that we have studied (see below). So we proposed to rename oogenesin as oogenesin-1, and the three genes that we decided to further study as oogenesin-2 (Mm.25370), oogenesin-3 (Mm.28722) and oogenesin-4 (Mm.148643).

3.2. Sequence analysis and structural features

Using data available on the mouse genome (http://www.ncbi.mapviewer) and sequences obtained from full length cDNAs retrieved from NIA (http://lgsun.grc.nia.nih.gov/cDNA), we reconstructed the full length cDNA as well as the exon-intron boundaries of oogenesin-2, -3 and -4. The predicted encoded proteins are composed of 493, 499 and 502 amino acids respectively, and these, as well as oogenesin-1, share a mean of 57.9% sequence identity (Fig. 2A), the corresponding genes being composed of four exons. Searching within domain databases Smart and Pfam did not reveal the presence of any described domains. Minami and colleagues suggested that oogenesin-1 is composed of leucine zippers [15]. However, a thorough analysis of PSI-BLAST searches combined with HCA clearly predicted that oogenesin genes belong to the superfamily of leucine-rich repeat (LRR) proteins [16–18]. LRRs are 20–29-residue sequence motifs present in a number of proteins with various functions. The crystal structure of ribonuclease inhibitor (RI) [19] yielded the first insight into the structural organization of these repeats, which was enriched afterwards by several other experimental structures, such as those of GTPase-activating protein rna1p, spliceosomal protein U2A′, Rab geranylgeranyltransferase, internalin B and YopM (reviewed in [18]). Each LRR corresponds to a structural unit, including an invariant β-strand. The assembly of several LRRs results in non-globular horseshoe-shaped structures in which the β-sheet formed by the

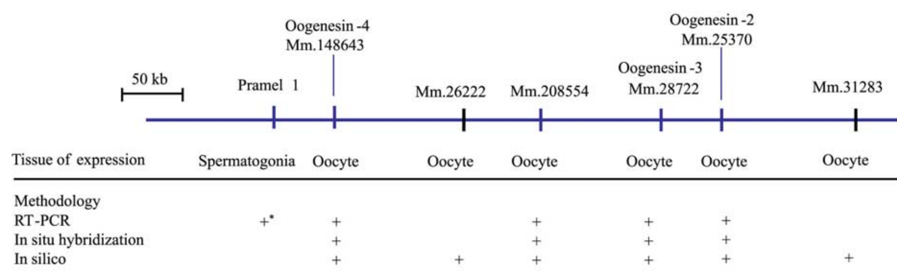

Fig. 1. Representation of the cluster of oogenesin genes in the chromosome 4. The presence of mRNAs or ESTs of the different oogenesin members in male and female germ cells was assessed by experimental (RT-PCR, in situ hybridization) and/or in silico methodologies as described in Section 2. *Data from [22].

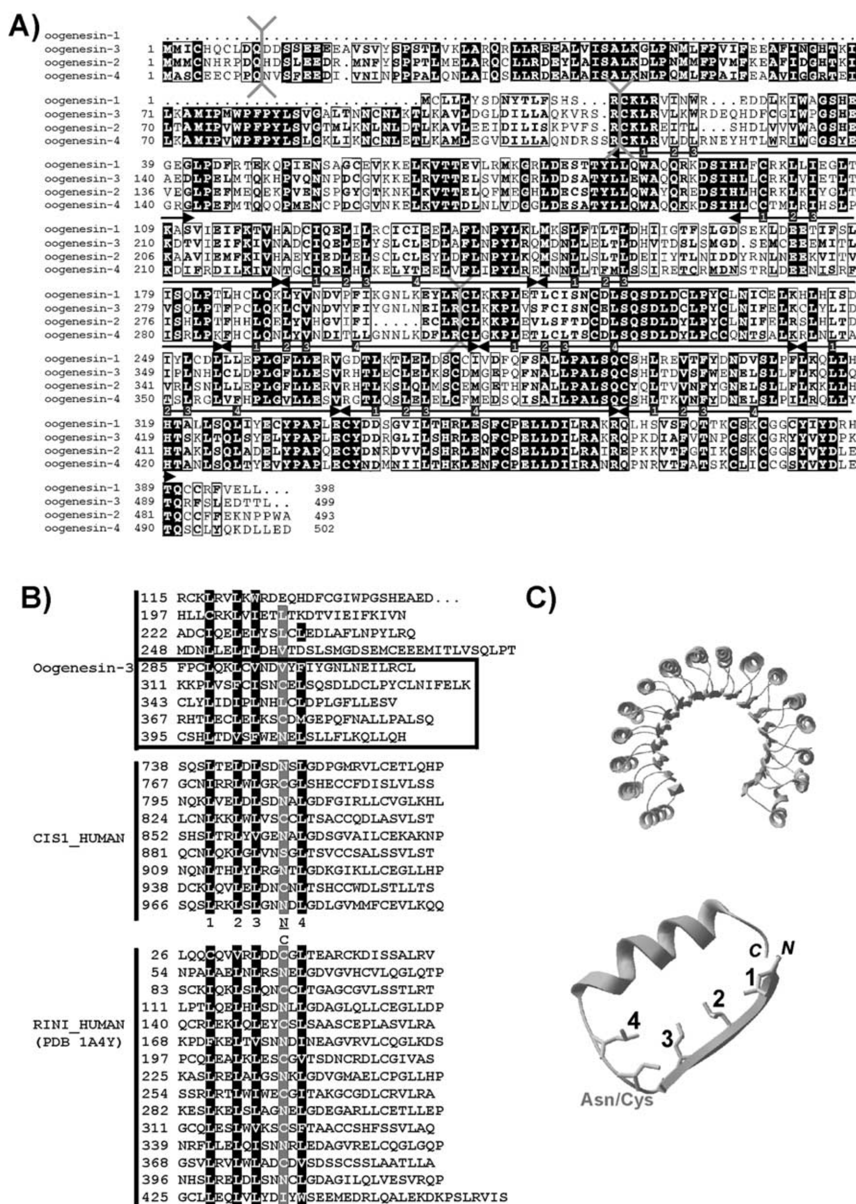

Fig. 2. Sequence analysis of oogenesin family. A: Alignment of the sequences of the *oogenesin* family described in this study. Identities and similarities are shown with black and white boxes, respectively. Positions of the recognized LRR β-region sequences, as reported in panel B for the oogenesin-3 sequence, are indicated with arrows below the alignment, with numbers representing the positions of highly conserved hydrophobic residues (as in panel B). Vertical gray bars specify intron–exon boundaries. This figure has been prepared using the ESPript software [12]. B: Alignment of the recognized LRR β-region sequences of a member of the *oogenesin* family (oogenesin-3) with those of established members of the LRR superfamily: the Cold autoinflammatory syndrome 1 protein (Cryopyrin) or NALP3 (CIS1_HUMAN) and the RI (RINI_HUMAN), whose three-dimensional structure is known (pdb identifier 1A4Y; see panel C). Below the alignment are labeled the positions which are mainly occupied by leucine residues (or hydrophobic ones), and the position frequently occupied by asparagine or cysteine in some LRR (see panel C). Several highly degenerated repeats are suspected to be present before and after the first repeat reported here, but these diverged too much from consensus sequences to be predicted with accuracy. C: Three-dimensional structure of the RI, highlighting the horseshoe shape of the LRR assembly (top). The structural unit of LRR corresponds to a β-strand and of a α-helix, running parallel to the central axis of the horseshoe structure (bottom). Highly conserved positions, mainly occupied by leucine residues (or hydrophobic ones) are highlighted and numbered from 1 to 4, as well as the position which in some LRR is frequently occupied by asparagine or cysteine (see panel B). *N* and *C* indicate the N- and C-terminus, respectively. This figure has been made using SwissPDBviewer [11].

parallel β-strands lines the inner circumference of the horseshoe. The interstrand segments are exposed outwards and are more variable from one LRR family to another, forming either α-helical, as in RI (Fig. 2C), 3_{10}-helical or more irregular extended structures. As a consequence, residues corresponding to the inner β-strand and consecutive loop regions are generally well conserved in LRR proteins, defining a 'LRR consensus sequence' whilst the remaining parts of the repeats may be very different [16,17]. Conserved amino acid positions of the LRR consensus sequence are mainly occupied by leucine residues, which form the buried face of the β-strand (Fig. 2C), in contrast to the propensity of this residue to form α-helices in standard globular domains.

In the oogenesin family, significant similarities were observed with the LRR β consensus sequences from various LRR proteins, in particular those of the Mater protein, the

RIs and of proteins of the NALP (NACHT-, LRR- and PYD-containing Protein) family [20]. These similarities range from aa 285 to aa 420 (based on the oogenesin-3 sequence). Such similarities were further confirmed and refined using HCA (boxed sequences in Fig. 2B). Identical or similar amino acids correspond mainly to residues critical for the LRR fold, which are found highly conserved in LRR multiple alignments (Fig. 2B and SMART profile (SM00370) [8] and buried within LRRs of known three-dimensional structures. It is worth noting that using automatic secondary structure prediction procedures, such sequences are generally predicted as α-helices due to the high content of leucine residues, as proposed for oogenesin-1 by Minami et al. [15]. Rather, in the protein sequences compared in the present study, the hydrophobic clusters corresponding to the LRR β-strands have shapes typical of β-strands, in agreement with experimentally deduced structures where they are associated with more than 80% to β-strands (Le Tuan et al., personal communication). Moreover, a striking feature of LRRs of the *oogenesin* family is that they are 'degenerated': they are variable both in length and sequences forming the β-strands region. This is due to leucine being often substituted by other hydrophobic amino acids as compared to other canonical LRRs. This makes them difficult if not impossible to detect through the domain database profiles. In this regard, oogenesin LRRs might be compared with the LRR L-domain-like family (following the SCOP classification [21]), which consists of less regular structures than the RI-like family, and are made of variable repeats. In the *oogenesin* family as in some other LRR proteins (e.g. RI and NALPs), cysteine and asparagine are found at the position immediately following the β-strands of some LRRs (Fig. 2B,C). In known structures, these residues form specific hydrogen bonds with the free peptide groups at the interior of the structure. However, in contrast to RI or NALPs, this feature is not repeated along the sequence in the *oogenesin* family. No significant alignment could be found in regions separating β-strands; however, these are predicted to contain α-helices, as in RI-like cognate LRRs sequences.

Further examination of sequences suggested that LRRs are also present in the N-terminal parts of the oogenesin proteins, although they are more degenerated and separated by more variable regions. Regions encompassing residues 95–135 and 195–275 (again based on the oogenesin-3 sequence) clearly show one and three LRR consensus sequences respectively, as indicated by several matches with cognate LRR sequences (non-boxed sequences in Fig. 2B). Clusters typical of β-strands, alternating with clusters typical of α-helices, were found in other regions (before and after the region 115–135), but their sequences do not fit well the β-region LRR consensus, making further prediction difficult.

In conclusion, our analysis clearly indicate that oogenesin members might consist of repeated sequences derived from the LRR superfamily, a prediction which now awaits support from experimental data in order to be confirmed and refined.

3.3. Expression analysis of oogenesin-2, -3 and -4

By analysis of RT-PCR products on BET-stained gels, oogenesin-2, -3 and -4 were confirmed to be exclusively expressed in the mouse ovary compared with other tissues (Fig. 3A). To increase the sensitivity of detection of these transcripts, Southern blot experiments were performed on RT-PCR products. Intense signals were only observed in the

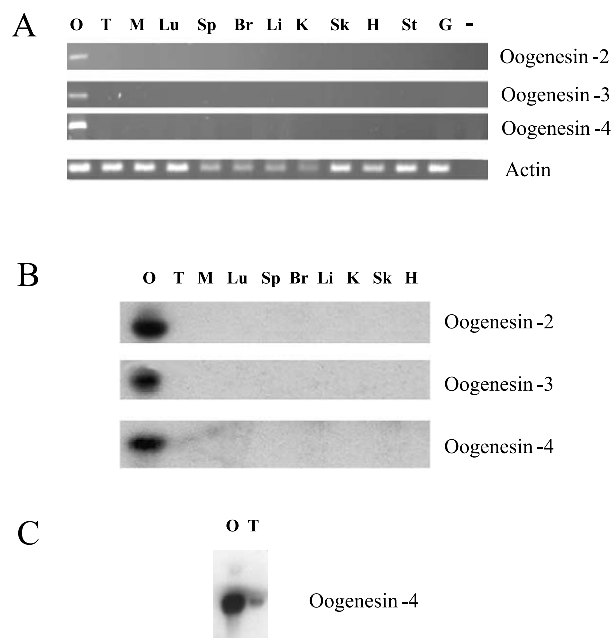

Fig. 3. Expression analysis of oogenesin-2, -3, -4 and actin by RT-PCR in mouse tissues. Total RNA from different tissues were subjected to RT-PCR as described in Section 2. A: BET staining gels. B: Southern blot analysis of the RT-PCR products, short exposure of the autoradiograph. C: Southern blot analysis of the RT-PCR products, long exposure of the autoradiograph. Note the low expression of oogenesin-4 in testis. O, ovary; T, testis; M, muscle; Lu, lung; Sp, spleen; Br, brain; Li, liver; K, kidney; Sk, skin; H, heart; St, stomach; G, ganglion; −, negative control.

ovary, except a faint band in the testis that was observed for oogenesin-4 after overexposure of autoradiographs (Fig. 3B,C).

We then performed in situ hybridization to study the localization of these transcripts in the ovary. Due to the high degree of identity between the three cDNA sequences studied, we have first verified the specificity of each RNA probe by slot-blotting (data not shown). In situ hybridization on mouse ovarian sections confirmed that oogenesin-2, -3 and -4 were specifically expressed in oocytes (Fig. 4). Transcripts were not detected in primordial follicles, clearly appeared in follicles with two layers of granulosa cells (Fig. 4a–f), and were present in early (Fig. 4g–l) as well as large antral follicles (Fig. 4m–r).

3.4. A role of oogenesin gene family in oogenesis and folliculogenesis?

Interestingly, six of the nine genes from the oogenesin family that we have identified in silico as preferentially expressed in oocyte, are positioned within a locus of approximately 500 kb on chromosome 4 (Fig. 1). In addition to oogenesin-2, -3 and -4, we have verified by RT-PCR that the transcript of Mm.208554 is also exclusively found in the ovary (data not shown), strongly suggesting that this gene is also specifically expressed in the oocyte. Moreover, the Pramel1 gene has been identified as preferentially expressed in spermatids by subtractive hybridization by Wang et al. [20]. Of note, we found here that low levels of oogenesin-4 transcripts are also present in testis. Moreover, several ESTs from Mm.31283 have also been found in testis libraries. Overall, these data suggest that all of these oogenesin genes are localized in a locus of genes highly

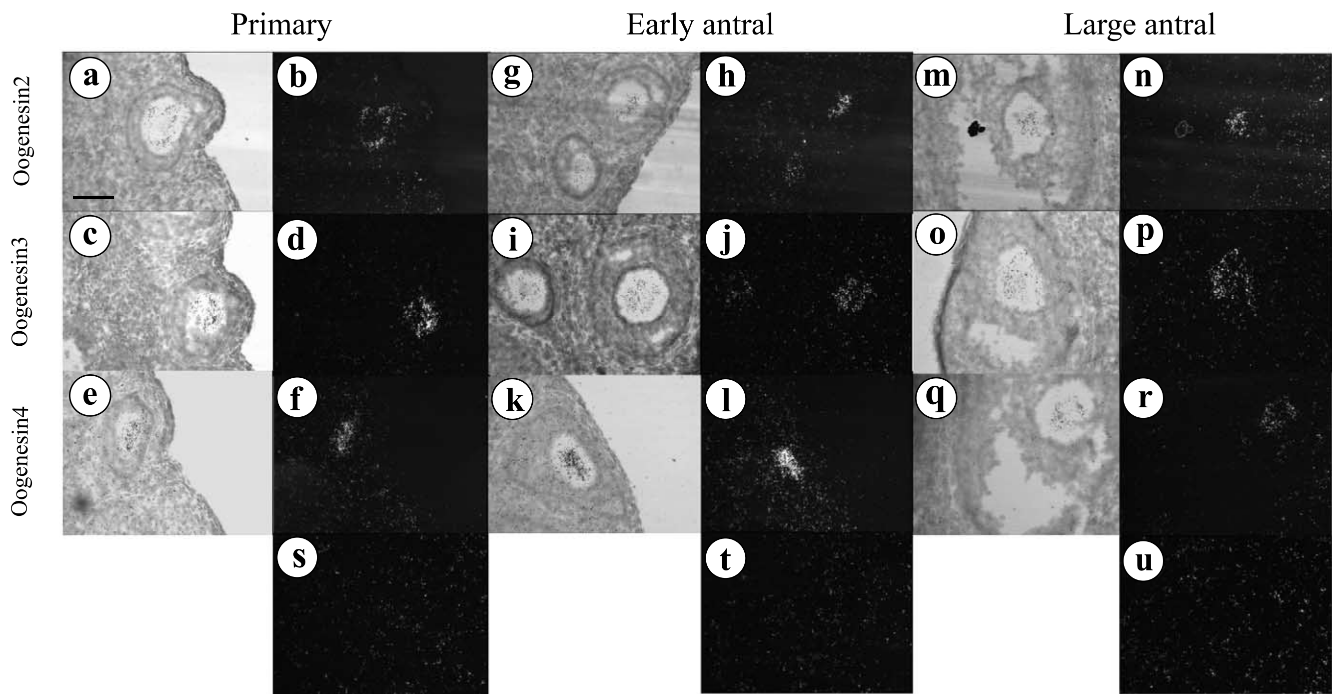

Fig. 4. Localization of oogenesin-2, -3 and -4 mRNAs by in situ hybridization in primary follicles (a–f), early antral follicles (g–l) and large antral follicles (m–r). Bright-field (a,c,e,g,i,k,m,o,q) and dark-field (b,d,f,h,j,l,n,p,r) photomicrographs of follicles hybridized with murine [^{35}S]oogenesin-2, -3 and -4 cRNA antisense probes. s,t,u: Dark-field photomicrographs of follicle hybridized with oogenesin-2, -3 and -4 sense cRNA probes. Scale bar: 50 μm.

if not specifically expressed in germ cells, suggesting the presence of specific genomic regulatory regions in the vicinity of these genes. Presently, the oogenesin-1 gene has not been clearly localized in the mouse genome. Whether it is also localized in this locus remains to be investigated.

Minami et al. have shown that oogenesin-1 mRNA is expressed in ovaries from 15.5 dpc fetus, newborn and adults [15]. In adult ovaries, both oogenesin-1 mRNA and protein were localized in oocytes from primordial to preovulatory follicles. In contrast, expression of oogenesin-2, -3 and -4 mRNA was observed in oocytes from primary to preovulatory follicles, but was not detected in primordial follicles (our present data). Whether the mRNAs of these three genes are translated into proteins, and at what follicular stage, remains to be investigated. Nevertheless, these results suggest that proteins from oogenesin family could play a role in oogenesis and/or folliculogenesis. Moreover, it was shown that oogenesin-1 protein is expressed until the four-cell stage embryo after fertilization, and that it is localized in nuclei at the early two-cell stage, consistent with a role in the activation of zygotic genome. However, a role in meiotic process cannot be ruled out since expression of some elements of this family has been found in spermatogonia [20] and in testis (oogenesin-4, our present work).

In contrast to Minami et al. who suggested that oogenesin-1 contains a canonical leucine-zipper domain [15], we have clearly shown here that the repeats of the oogenesin family obviously appear to be divergent members of the LRR superfamily [17], suggesting that they might have specific functions that remain to be determined. Given the degenerate nature of the observed sequence similarities with LRR proteins, the present prediction however awaits further support from experimental data. Whatever the role of oogenesin genes family, one can hypothesize that the specific expression of such a redundant family of genes in germ cells is significant of a crucial role in gametogenesis and or early development. Investigating a role for these genes will require functional studies such as identification of partners and/or transgenesis studies.

In conclusion, we have used the in silico subtraction methodology to identify novel oocyte-specific genes in the mouse. We have identified a new family of LRR proteins characterized by 'degenerated' β-strands. This family is composed of more than 80 genes, nine being oocyte-specific in an in silico screen. We have verified by RT-PCR and in situ hybridization that three of them are specifically expressed in oocytes from primary to large antral follicles. These three genes are localized in a cluster on mouse chromosome 4.

Acknowledgements: We thank Claude Cahier and his team for expert animal care. We wish also acknowledge Joëlle Dupont, Eric Jean-Pierre, Thierry Delpuech and Claudine Pisselet for excellent technical assistance. We thank Rozenn Dalbies-Tran, Svetlana Uzbekova, and Sophie Pennetier for helpful discussion. Grant support: This work was supported by an INRA grant. S.D. was supported by a fellowship from the Institut National de la Recherche Agronomique and Région Centre.

References

[1] Zeng, F. and Schultz, R.M. (2003) Biol. Reprod. 68, 31–39.
[2] Rajkovic, A., Yan, M.S.C., Klysik, M. and Matzuk, M. (2001) Fertil. Steril. 76, 550–554.
[3] Dong, J., Albertini, D.F., Nishimori, K., Kumar, T.R., Lu, N. and Matzuk, M. (1996) Nature 383, 531–535.
[4] Galloway, S.M., McNatty, K.P., Cambridge, L.M., Laitinen, M.P., Juengel, J.L., Jokiranta, T.S., McLaren, R.J., Luiro, K., Dodds, K.G., Montgomery, G.W., Beattie, A.E., Davis, G.H. and Ritvos, O. (2000) Nat. Genet. 25, 279–283.

[5] Dean, J. (2002) J. Reprod. Immunol. 53, 171–180.
[6] Varani, S. and Matzuk, M.M. (2002) Ernst Schering Res. Found. Workshop 41, 63–79.
[7] Altschul, B., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Nucleic Acids Res. 25, 3389–3402.
[8] Letunic, I., Goodstadt, L., Dickens, N., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R., Ponting, C. and Bork, P. (2002) Nucleic Acids Res. 30, 242–244.
[9] Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L. (2000) Nucleic Acids Res. 28, 263–266.
[10] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J.-P. (1997) Cell. Mol. Life Sci. 53, 621–645.
[11] Guex, N. and Peitsch, M.C. (1997) Electrophoresis 18, 2714–2723.
[12] Gouet, P., Robert, X. and Courcelle, E. (2003) Nucleic Acids Res. 31, 3320–3323.
[13] Gay, E., Seurin, D., Babajko, S., Doublier, S., Cazillis, M. and Binoux, M. (1997) Endocrinology 138, 2937–2947.
[14] Besnard, N., Pisselet, C., Monniaux, D., Locatelli, A., Benne, F., Gasser, F., Hatey, F. and Monget, P. (1996) Biol. Reprod. 55, 1356–1367.
[15] Minami, N., Aizawa, A., Ihara, R., Miyamoto, M., Ohashi, A. and Imai, H. (2003) Biol. Reprod. 69, 1736–1742.
[16] Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 374, 183–186.
[17] Kajava, A.V. (1999) J. Mol. Biol. 277, 519–527.
[18] Kobe, B. and Kajava, A. (2003) Curr. Opin. Struct. Biol. 11, 725–732.
[19] Kobe, B. and Deisenhofer, J. (1993) Nature 366, 751–756.
[20] Tschopp, J., Martinon, F. and Burns, K. (2003) Nat. Rev. 4, 95–104.
[21] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) J. Mol. Biol. 247, 536–540.
[22] Wang, P.J., McCarrey, J.R., Yang, F. and Page, D.C. (2001) Nat. Genet. 27, 422–426.