Identification of differentially expressed markers in human follicular cells associated with competent oocytes

Melanie Hamel1, Isabelle Dufort1, Claude Robert1, Catherine Gravel1, Marie-Claude Leveille2, Arthur Leader2 and Marc-Andre Sirard1,3

1Département des Sciences Animales, Centre de Recherche en Biologie de la Reproduction, Université Laval, Québec, Canada G1K 7P4; 2Ottawa Fertility Centre, Ottawa, Canada; 3Correspondence address. E-mail: marc-andre.sirard@crbr.ulaval.ca

BACKGROUND: The development of an accurate method for selection of high-quality embryos is essential to achieve high pregnancy rates with single embryo transfer in human IVF. The developmental competence of the oocyte is acquired during follicle maturation and strong communication also exists between the follicular cells (FCs) and the oocytes; thus oocyte developmental competence may be determined by markers expressed in the surrounding FCs.

METHODS: From consenting patients (n = 40), FCs were recovered on a per follicle basis by individual follicle puncture. Hybridization analyses using a custom-made complementary DNA microarray containing granulosa/cumulus expressed sequence tags (ESTs) from subtracted libraries and an Affymetrix GeneChip® were performed to identify specific genes expressed in follicles leading to a pregnancy. The selected candidate genes were validated by quantitative-PCR (Q-PCR).

RESULTS: Subtractive libraries prepared from pooled samples representing pregnant versus non-pregnant patients produced 1694 ESTs. Hybridization data analysis discriminated 115 genes associated with competent follicles. Selected candidates were confirmed by Q-PCR: 3-beta-hydroxysteroid dehydrogenase 1 (P = 0.0078), Ferredoxin 1 (P = 0.0203), Serine (or cysteine) proteinase inhibitor clade E member 2 (P = 0.0499), Cytochrome P450 aromatase (P = 0.0359) and Cell division cycle 42 (P = 0.0396).

CONCLUSIONS: Microarray technologies are useful to mine the transcriptome of FCs expressed in follicles associated with competent oocytes and could be used to improve embryo selection with the objective of successful single embryo transfer.

Keywords: competent oocyte; follicle; granulosa cells; multiple pregnancies; gene expression

Introduction

To improve pregnancy rates in human IVF more than one embryo is commonly transferred, leading to an increased risk of multiple pregnancies. Multiple pregnancies are associated with potential harmful consequences to the mother and the babies (Pinborg, 2005) and are considered a major problem in IVF treatments.

To alleviate this problem, there is a need for a method that would predict the viability of the embryos available for transfer. The techniques already used to assess embryo quality are mainly based on subjective microscopic methods such as embryo morphology, cleavage rate, appearance of pronuclei (Van Royen et al., 2001), fragmentation rate, blastomeres number and embryo symmetry (De Neubourg et al., 2004). However, 29% of embryos with good morphological appearance can present chromosomal abnormalities (Munne et al., 1995), and this is likely underestimated in regards to the restricted number of chromosomes assessed (Munne and Cohen, 1998) or the hormonal stimulation protocol applied (Munne et al., 1997). Taken together, these suggest that embryos with proper morphological appearance alone are not sufficient to predict a pregnancy.

Follicular cells (FCs) obtained during oocyte recovery in regular IVF cycles can be divided into two subpopulations, the cumulus cells and the mural granulosa cells. The cumulus cells form a group of closely associated cells that surround the oocyte in the antral follicle and the mural granulosa cells line the follicular wall. It is well known that bi-directional communication between the oocyte and these cells occurs throughout follicul development (Buccione et al., 1990; Eppig et al., 2002; Senbon et al., 2003; Gilchrist et al., 2004; Makabe et al., 2006; Sirard et al., 2006) and is essential in the acquisition of developmental competence in mammalian oocytes (de Loos et al., 1991; Webb et al., 2002; Fair, 2003). Moreover, it is well established that granulosa cells play an essential role in the follicular differentiation process leading to the optimal conditions for the oocyte development, ovulation, fertilization and subsequent implantation (Adashi, 1994). Therefore, the embryo quality may depend on the final maturation of the follicle leading to an oocyte with the ability to give a successful pregnancy.
Our understanding of the specific and temporal changes in gene expression of FCs during follicular growth is far from complete in animals and humans. Some reports have studied the effects of FSH on transcriptional regulation in human granulosa cells (Sasson et al., 2003; Perlman et al., 2006) or by LH/hCG stimulation on the granulosa cells of mice (McRae et al., 2005; Hernandez-Gonzalez et al., 2006; Shimada et al., 2006), cattle (Robert et al., 2001) and rats (Leo et al., 2001). Other studies have investigated the human cumulus cells transcriptome (McKenzie et al., 2004; Zhang et al., 2005; Assou et al., 2006) or the mouse cumulus cell oocyte complex transcriptomes (Hernandez-Gonzalez et al., 2006). These studies provide valuable information about the gene expression profile in FCs during follicular growth and in response to FSH and LH/hCG stimulation. However, the relationship of the expression of specific genes during follicular growth and their ability to determine embryo quality is yet to be thoroughly investigated.

A previous experimentation in our laboratory studied gene expression patterns between follicles bearing an healthy oocyte leading to a blastocyst, and follicles leading to developmental arrest in the bovine (Robert et al., 2001). This study revealed that several genes could be associated with follicular competence, supporting the hypothesis that the transcriptomes of granulosa and/or cumulus cells from individual human follicles that produce embryos resulting in a pregnancy could contain useful markers.

The granulosa cells are usually discarded during oocyte recovery procedures. Gene expression profile in these cells may accurately reflect the fertility potential of an oocyte in a non-invasive way. Identification of oocytes that produce high-quality embryos that more likely to implant may reduce the number of embryos transferred into a recipient without lowering pregnancy rates. Therefore, we have analyzed FCs (mural granulosa and cumulus cell) subtracted libraries obtained from individual aspirated follicles from patients undergoing IVF and pooled samples into two groups based on pregnancy outcome (follicles resulting in a pregnancy and follicles leading to an embryo that fails to develop) and thus identified markers of embryonic quality and competence. In addition, any information about the successful conditions associated with good follicles could be useful to better modulate ovarian stimulation protocols.

**Materials and Methods**

**FC collection**

FCs were obtained with their consent from women (n = 40) that were undergoing IVF treatment at the Fertility Center at the Ottawa Hospital, Canada. These women had endometriosis, tubal or idiopathic infertility diagnosis but not polycystic ovary syndrome. The procedure was performed with approval from the Ottawa Hospital Research Ethics Board. Following ovarian stimulation, follicular fluid, FCs and oocytes from individual follicles were collected by ultrasound-guided follicular aspiration using a double lumen needle. The oocytes and surrounding cumulus cells were removed for IVF treatment. The remaining follicular fluid was centrifuged at 800g for 10 min at room temperature to isolate the FCs containing mural granulosa cells, for each individual follicle. The resulting pellet was suspended in 500 µl of phosphate-buffered saline solution at 4°C and was transferred into a cryovial. After centrifugation at 2000g for 1 min at room temperature, the supernatant was removed and cells were rapidly frozen and stored in liquid nitrogen until RNA extraction.

After the fertilization process, cumulus cells surrounding the oocytes were also recovered on an individual follicle basis using the same protocol as described for FC isolation.

A range of 1–15 follicles were aspirated for an average of 7.48 follicles per patient, and an average of 4.13 embryos was obtained per woman. Data (fertilization, embryo development, embryo morphology, transfer and pregnancy) generated from each follicle was recorded by an embryologist. Depending on the IVF protocol used, one or two (average of 1.4) embryos were transferred at either day 3 (67%) or day 5 (33%) for a total of 34 patients with an overall per transfer pregnancy rate of 53%. Pregnancy was confirmed by the presence of a fetal heartbeat by ultrasound at 6–8 weeks.

**Treatment assignment**

For the mural granulosa cells, three pools of follicles [pool 1 (n = 6), pool 2 (n = 15) and pool 3 (n = 9)] were created from follicles associated with a successful pregnancy, which were called the positive groups 1, 2 and 3, respectively (Table I). Each embryo of the positive group was scored according to the clinic’s embryo selection protocol. The cleavage stage and morphological characteristics were the two main criteria. Following a chart table, transferred embryos were at least 8 cells or 5–7 cells with high scores in morphology grade. Pregnancy was confirmed by the presence of a fetal heartbeat by ultrasound at 6–8 weeks. These pools were used to generate RNA associated with positive groups. Three more pools [pool 1 (n = 6), pool 2 (n = 15) and pool 3 (n = 9)] (Table I) were assigned to the negative groups 1, 2 and 3, respectively, containing follicles resulting in embryos that arrested their development before the 8-cell stage and include one embryo that did not implant (group 1). The qualification of these embryos was done by morphological criteria, according to shape, granularity, presence of multinucleated blastomeres and 3D orientation. These embryos were ‘truly’ arrested for at least 24 h and showed an increase of fragmentation and degeneration. Positive and negative groups from pool 1 was used to make the custom-made complementary DNA (cDNA) microarray, pools 1 and 2 from both positive and negative groups served for array hybridizations and all 3 pools served for quantitative real-time (Q)-PCR analysis. Cumulus cells from the same follicles selected for both the positive and negative group 1 were used separately from mural granulosa cells to elaborate a cumulus cells subtracted library and the custom-made cDNA microarray.

**RNA isolation**

Total RNA from both mural granulosa cells and cumulus cells was extracted with 1 ml of Trizol reagent (Invitrogen, Burlington, Canada) following the manufacturer’s protocol. Both the mural granulosa cells and cumulus cells were extracted separately for each individual follicle. DNase treatment was then applied using the DNase I Amplification Grade kit (Invitrogen) according to the manufacturer’s instructions. Extracted RNA was dissolved in 30 µl of water and quantified by spectrophotometry at 260 nm. Total RNA quality and integrity were verified using an Agilent Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, USA).

**Microarray slide preparation**

Suppressive subtractive hybridizations for granulosa cells and cumulus cells

According to the manufacturer’s instructions for the BD SMART PCR cDNA Synthesis Kit (Clontech, Mountain View, USA), mRNAs from positive group 1 and negative group 1 from granulosa and cumulus cells (1 µg) from pools of total RNA were reverse transcribed.
The suppressive subtractive hybridization (SSH) was performed with the PCR Select cDNA Subtraction Kit (Clontech) according to the manufacturer’s instructions. DNA amplified previously with the SMART kit from positive groups 1 of mural granulosa and cumulus cells served as the tester and negative groups 1 cells for both different groups of granulosa and cumulus cells as the driver. The suppressive hybridization of adapter-ligated testers with drivers was performed according to the manufacturer’s protocol (Clontech), using a tester to driver ratio of 1:60.

### cDNA sequencing

The PCR products were ligated into a vector using the pGEM®-T Easy Vector (Promega, Nepean, Canada) and then transformed into DH5α-T1 Max Efficiency cells (Invitrogen). For both mural granulosa cells subtracted library and cumulus cells subtracted library, bacterial colonies (1050) were randomly picked and grown in 96-well plates containing 200 μl Luria-Bertani Broth medium (BD Biosciences, Mississauga, Canada) with 50 μg/ml ampicillin (Sigma-Aldrich, Oakville, Canada). Colonies were incubated at 37°C with agitation for 6 h and then kept at 4°C until PCR amplification of the inserted fragment. For PCR, 2 μl of bacterial suspension were added to a PCR mix containing 1× PCR buffer, 0.25 mM dNTP, 0.25 mM PCR Nested Primer 1 (5’-AGCGTGGTCCGGCCCGGG-3’), 2R (5’-TCGACGGCGCCGCCGGCAGT-3’) (Clontech) and 1.25 U of HotMaster Taq DNA Polymerase (Eppendorf, Mississauga, Canada). The PCR conditions consisted of a 94°C initial denaturing step for 2 min and 30 cycles consisting of a denaturing step of 20 s at 94°C, an annealing step of 10 s at 65°C and an elongation step of 1 min at 65°C and a final step at 65°C for 7 min. PCR product aliquots (3 μl) were visualized on 1% agarose-ethidium bromide gel to verify cDNA length and quality (single band). Amplicons with more than one band were rejected. The remaining bacterial suspension was stored in 20% glycerol at −80°C.

The PCR products were purified and sequenced as described previously (Vallee et al., 2006). Sequence traces were visualized with the online freeware Chromas 1.45 (http://www.technelysium.com.au/chromas.html) and sequences were loaded into the cDNA Library Manager Program (Genome Canada Bioinformatics, Quebec, Canada), trimmed (http://www.phrap.org/phredphrapconsed.html) and compared against the Genbank database (http://www.ncbi.nlm.gov/BLAST/). The BLAST results were compiled into a report chart for each submitted sequence.

### Custom-made cDNA microarray preparation

Purified PCR products were speedvac-evaporated (SPD SpeedVac Savant), suspended in a solution of equal parts of dimethyl sulfoxide and H2O, and spotted in two replicates in different location on GAPSII glass slides (Corning, Corning, NY, USA), using a VersArray Chip WriterPro robot (Bio-Rad, Mississauga, Canada). In addition to human mural granulosa and cumulus cells subtracted libraries, other libraries previously obtained in our lab were also spotted on the slide in two replicates. These libraries are from a bovine cumulus cell subtracted library (not published) and a bovine competent granulosa cell subtracted library (Robert et al., 2001). A SpotReport Alien and Plant cDNA Array Validation System (Stratagene, Ottawa, Canada) were printed as negative controls. Human actin, tubulin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNAs acted as positive controls and a fragment of the green fluorescent protein (GFP) was used as an exogenous positive control. DNA was then cross-linked with ultraviolet rays (300 mV) and quality control was performed with Terminal Deoxynucleotidyl Transferase Assay (GE healthcare, Quebec, Canada).

### Microarray hybridizations

**Custom-made cDNA microarray preparation**

The custom-made cDNA microarrays were hybridized in an ArrayBooster using the Advacard chart for each submitted sequence.

### Microarray hybridizations

**Custom-made cDNA slide hybridizations**

Total RNA from positive and negative groups 2 of mural granulosa cells were amplified using the RiboAmp™ RNA Amplification kit (Molecular Devices, Mountain View, USA) according to the manufacturer’s instructions. Briefly, total RNA was reversed transcribed with a primer incorporating a T7 RNA polymerase promoter sequence. Double-stranded cDNA was synthesized, column-purified (Qiagen, Mississauga, Canada) and used as a template to drive in vitro transcription using the T7 polymerase. This global amplification was linearly amplified by one round and the resulting UTP-axonol allyl RNA (aaRNA) was column purified and the quantity of aaRNA was estimated by spectrophotometry at 260 nm. Probes were labeled with Alexa Fluor 555 and 647 reactive dye packs (Invitrogen) according to the protocol from Molecular Probes. Slides were hybridized overnight at 55°C (for cDNA) or 50°C (for RNA) and labeled purified probes using the SlideHyb #1 buffer (Ambion, Austin, USA). Hybridizations were performed in an ArrayBooster using the Advacard.

### Table I. Treatment assignments with granulosa cell tissues.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Positive groups</th>
<th>Negative groups</th>
<th>Pool 1</th>
<th>Pool 2</th>
<th>Pool 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Transferred oocyte with pregnancy</td>
<td>Failure in development</td>
<td>Transferred oocyte with pregnancy</td>
<td>Transferred oocyte with pregnancy</td>
<td>Transferred oocyte with pregnancy</td>
</tr>
<tr>
<td></td>
<td>6 follicles</td>
<td>4 patients</td>
<td>15 follicles</td>
<td>9 patients</td>
<td>9 follicles</td>
</tr>
<tr>
<td></td>
<td>Custom-made granulosa array (cumulus and granulosa cells libraries)</td>
<td>Hybridization of the custom-made</td>
<td>Hybridization of the Affymetrix Chip ‘B’</td>
<td>Hybridization of the custom-made</td>
<td>Hybridization of the Affymetrix Chip ‘D’</td>
</tr>
<tr>
<td></td>
<td>Subtracted libraries (cumulus and granulosa cells libraries)</td>
<td>Granulosa array ‘A’</td>
<td>Gene validation by real-time PCR</td>
<td>Gene validation by real-time PCR</td>
<td>Gene validation by real-time PCR</td>
</tr>
<tr>
<td></td>
<td>Hybridization of the Affymetrix Chip ‘C’</td>
<td>Hybridization of the Affymetrix Chip ‘D’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gene validation by real-time PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cumulus cell tissues were used separately to construct a subtracted library and the Custom-made granulosa array.
AC3C (The Gel Company, San Francisco, USA). Slides were then washed twice with 2 x standard saline citrate (SSC)/0.5% sodium dodecyl sulfate (SDS) for 15 min at 55°C (for cDNA) or 50°C (for RNA) and twice with 0.5 x SCC / 0.5% SDS for 15 min at 55°C (for cDNA) or 50°C (for RNA).

**Experimental design for the custom-made cDNA microarray hybridizations**

Two hybridizations were performed using different pools of patients (groups 1 and 2). For the first hybridization, forward-subtracted PCR products from the human mural granulosa cells library (positive and negative groups 1) were used as probes to hybridize custom-made cDNA microarray. For the second hybridization, RNA from both positive and negative groups 2, linearly amplified by one round of T7, were used as probes.

Slides were scanned using the VersArray ChipReader System (Bio-Rad) and analyzed using the ChipReader and ArrayPro Analyzer software (Media Cybernetics, Bethesda, USA). Data analysis was performed as described previously (Vallee et al., 2005; Mourtou et al., 2006). Fluorescence signal intensities for each replicate were log 2 transformed and normalized by the Loess method, and corrected for background. The determination of the background signal threshold was performed with the SpotReport cDNA controls (Stratagene), which determine the background (t = m + 2 x sd, where ‘t’ is the calculated threshold, ‘m’ the mean and ‘sd’ the standard deviation of the negative control data, n = 58). Transcripts above the threshold were considered as present in granulosa cells, whereas the other transcripts were eliminated from the analysis. If one replicate was lower than the background, the clone was completely eliminated from the analysis. For both hybridizations, candidates with a log 2 ratio more than 2 were listed and candidates appearing in both lists were given more attention.

**Affymetrix slide hybridizations**

Two Affymetrix human genome arrays were hybridized with the positive and negative groups 1 and 2, respectively, at the CREMO (Centre de recherche du CHUL, Quebec, Canada). Double stranded cDNA synthesized by reverse transcription was obtained from 250 ng of RNA and amplified twice according to the Affymetrix instructions. Biotin-labeled aaRNA was produced from the cDNA from mural granulosa cells and used to probe the Affymetrix human genome array (HG-U133 Plus 2 array) (Affymetrix, Lexington, USA) (http://www.affymetrix.com/technology/index.affx). This gene chip contains probes for 33 000 well-substantiated human genes (44 700 transcripts). Hybridizations and washes were performed using the Affymetrix GeneChip system according to the manufacturer’s instructions. Average difference and expression level of genes were calculated according to absolute and comparison analysis algorithms as recommended by the manufacturer. A ratio more than 2 (positive groups:negative groups) was used to select candidates.

**Candidate gene selection**

Selection of clones for further analysis was based on the microarray results from the custom-made cDNA array slides and the Affymetrix slides. A total of 115 different markers were then selected and graded according to their number of occurrences in different libraries, their presence in the human granulosa library, their replication in the same library and the signal intensities. After selection and grading, 18 candidate genes were validated by quantitative-PCR (Q-PCR): cytochrome P450 aromatase (CYP19A1); cell division cycle 42 (CDC42); dihydropyrimidinase-like 3 (DPYSL3); 3-beta-hydroxysteroid dehydrogenase (HSD3B1); epiregulin (EREG); serine (or cysteine) proteinase inhibitor clade E member 2 (SERPINE2); serine (or cysteine) proteinase inhibitor, clade A, member 3 (SERPINA3); tumor necrosis factor, alpha-induced protein 6 (TNFAIP6); scavenger receptor class B, member 1 (SCARB1); inhibin beta A (INHBA); sprouty 2 (SPRY2); ferredoxin 1 (FDX1); regulator of G-protein signaling 2 (RGS2); neuropilin 1 (NRP-1); early growth response 1 (EGR1); phosphoglycerate kinase 1 (PGK1); interleukin 6 signal transducer (IL6ST) and basic leucine zipper transcription factor 2 (BACH2) and 2 housekeeping genes [Beta actin (ACTB) and GAPDH] were used as an internal control.

**Real-time PCR**

Primers of each candidate gene were designed with the Primer3 web interface using sequences derived from The National Center for Biotechnology Information (NCBI) corresponding to our library sequences (Table II). Real-time analysis measured and compared the three different groups of mural granulosa cells for the positive and negative groups with the same procedure already published (Vigneault et al., 2004). Briefly, for each sample, a reverse transcriptase was performed using 50 ng granulosa cell RNA (quantified by spectrophotometer) with the Sensiscript kit (Qiagen) according to the manufacturer’s directions. GFP RNA (7 pg) was added to the RNA mixture as an exogenous control for the reaction. To confirm that the right product was amplified, all amplifications were visualized on an agarose gel (2%) and then sequenced.

**Statistical analysis**

Data normalization for each gene in every pool of amplified cDNA was performed by calculation as a ratio to the level of GFP RNA as already published (Dode et al., 2006). Data are presented as mean ± SEM. GAPDH and ACTB housekeeping genes were assessed to verify the stability of RNA quantity. The evaluation of mRNA differences between the positive groups and negative groups was done by a non-parametric two-tailed unpaired t-test. Differences were considered statistically significant at the 95% confidence level (P < 0.05).

**Results**

**Suppressive subtractive hybridization**

Using the SSH technique, we subtracted reverse transcribed mRNA from six follicles that resulted in a successful pregnancy (positive group 1) and six follicles leading to an embryo with failure in development (negative group 1). After cloning and transformation, 1050 clones were selected from each subtracted library, where 852 clones from the human mural granulosa cells subtracted library and 842 clones from the human cumulus cells subtracted library were positive for a single insert (Table III). Clones were then sequenced and analyzed using the nucleotide–nucleotide BLAST (blastn) program on the NCBI database. In the granulosa and cumulus cell subtracted libraries, 89% and 68% of the total sequences were from genes with known functions, respectively, leading to a total of 465 and 645 unique sequences in each library (Table III). These two subtracted libraries share only 58 unique sequences, which constitute evidence that these two populations of cells possess a repertoire of different potential candidate genes.
### Table II. Information and sequences of specific primers used for amplification in real-time PCR.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Primer sequences</th>
<th>GenBank accession number</th>
<th>UniGene accession number</th>
<th>Product size (bp)</th>
<th>Annealing temperature (°C)</th>
<th>Fluorescence acquisition temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP19A1</td>
<td>F-GCACATCTCTAATACTACGAGTC&lt;br&gt;R-CTTGAGGACGATGCACGCAAGA</td>
<td>NM_000103</td>
<td>HS65438</td>
<td>380</td>
<td>56</td>
<td>84</td>
</tr>
<tr>
<td>CDC42</td>
<td>F-ACGACCCTAGTATTCCACCAAC&lt;br&gt;R-ATACCTGAGGAGCACCTCAAGGG</td>
<td>NM_001791</td>
<td>HS69358</td>
<td>262</td>
<td>57</td>
<td>82</td>
</tr>
<tr>
<td>DPYSL3</td>
<td>F-TGTCCTGACCTCTCCTTACGATT&lt;br&gt;R-ATACCAGCCACTCTTCTCATTTC</td>
<td>NM_001387</td>
<td>HS51695</td>
<td>171</td>
<td>58</td>
<td>85</td>
</tr>
<tr>
<td>HS3B1</td>
<td>F-TGTCGGCAGTTCCTTTACAC&lt;br&gt;R-TGTGTTCCAGAGGCTCTTCTTC</td>
<td>NM_008662</td>
<td>HS364941</td>
<td>101</td>
<td>55</td>
<td>83</td>
</tr>
<tr>
<td>EREG</td>
<td>F-GCACTTCTGTCAGAGCTAC&lt;br&gt;R-CTGGGTGTCATCTGTACAGG</td>
<td>NM_001432</td>
<td>HS115263</td>
<td>159</td>
<td>56</td>
<td>82</td>
</tr>
<tr>
<td>SERPINE2</td>
<td>F-TGAAGGGACGCTGAAAGTTG&lt;br&gt;R-ACCTCCAGAACAGGAAACACTTGC</td>
<td>NM_006216</td>
<td>HS38449</td>
<td>451</td>
<td>59</td>
<td>81</td>
</tr>
<tr>
<td>SERPINA3</td>
<td>F-ACAAGTGAGGAAGTGGAAGG&lt;br&gt;R-AGGCTACAGAGGAGCAAG</td>
<td>NM_001085</td>
<td>HS534293</td>
<td>347</td>
<td>59</td>
<td>87</td>
</tr>
<tr>
<td>TNFAIP6</td>
<td>F-GGAATCTCTCTCAATAGGAGT&lt;br&gt;R-AACTCCAGTAATACCTGACC</td>
<td>NM_007115</td>
<td>HS437322</td>
<td>190</td>
<td>56</td>
<td>82</td>
</tr>
<tr>
<td>INHBA</td>
<td>F-TCAGTCTCAAGGACATCGCT&lt;br&gt;R-AACTCGACATGGCTTCTGTT</td>
<td>NM_002192</td>
<td>HS583348</td>
<td>223</td>
<td>57</td>
<td>87</td>
</tr>
<tr>
<td>SPRY2</td>
<td>F-GGAATGCTAAGCAACAGG&lt;br&gt;R-GGACACCTGAGAAGAATGGTC&lt;br&gt;</td>
<td>NM_005842</td>
<td>HS18676</td>
<td>215</td>
<td>57</td>
<td>77</td>
</tr>
<tr>
<td>FDX1</td>
<td>F-TCAACCTGTCACCTCATCTTTG&lt;br&gt;R-AGGGTCATGGCATCAGAC&lt;br&gt;</td>
<td>NM_004109</td>
<td>HS744</td>
<td>168</td>
<td>57</td>
<td>80</td>
</tr>
<tr>
<td>RGS2</td>
<td>F-CTGTGACCCCTGACACAGATT&lt;br&gt;R-CAACACCACTCTCTCTTCTG&lt;br&gt;</td>
<td>NM_002923</td>
<td>HS78944</td>
<td>179</td>
<td>57</td>
<td>81</td>
</tr>
<tr>
<td>NRP1</td>
<td>F-CCCTGTGGTTTTATCTCCGAGG&lt;br&gt;R-GAGACTTGTGGAGCAAGAC&lt;br&gt;</td>
<td>NM_003873</td>
<td>HS131704</td>
<td>191</td>
<td>56</td>
<td>82</td>
</tr>
<tr>
<td>EGR1</td>
<td>F-CCATTACGAGGAGGAGT&lt;br&gt;R-GGGTCAGCATGATGATGG&lt;br&gt;</td>
<td>NM_001964</td>
<td>HS326035</td>
<td>251</td>
<td>58</td>
<td>82</td>
</tr>
<tr>
<td>PGK1</td>
<td>F-ACCTTGCTGCTGGAAGAGC&lt;br&gt;R-TTAAGGGTGTCTGCACCT&lt;br&gt;</td>
<td>NM_000291</td>
<td>HS78971</td>
<td>449</td>
<td>59</td>
<td>86</td>
</tr>
<tr>
<td>BACH2</td>
<td>F-AAGCTACCACTTGAGTGCTG&lt;br&gt;R-CAACAGCTGGAGGCTGG&lt;br&gt;</td>
<td>NM_021813</td>
<td>HS269764</td>
<td>283</td>
<td>59</td>
<td>82</td>
</tr>
<tr>
<td>IL6ST</td>
<td>F-AGGAAACCGTGATACCAAAAGGCC&lt;br&gt;R-AGGAATGCTAAGCAGCAAG&lt;br&gt;</td>
<td>NM_175767</td>
<td>HS530822</td>
<td>378</td>
<td>59</td>
<td>81</td>
</tr>
<tr>
<td>GAPDH</td>
<td>F-ACACAGTCTCATGCTC&lt;br&gt;R-CCACACCTGTTGCT&lt;br&gt;</td>
<td>NM_002046</td>
<td>HS544577</td>
<td>452</td>
<td>56</td>
<td>89</td>
</tr>
<tr>
<td>ACTB</td>
<td>F-CTGGACATAAAGGAGAAGC&lt;br&gt;R-CTCAAGGCGAGAAGATC&lt;br&gt;</td>
<td>NM_001101</td>
<td>HS520640</td>
<td>375</td>
<td>58</td>
<td>88</td>
</tr>
</tbody>
</table>

CYP19A1, Homo sapiens Cytochrome P450, family 19, subfamily A, polypeptide 1; CDC42, Homo sapiens Cell division cycle 42; DPYSL3, Homo sapiens Dyhdrorymidinase-like 3; HS3B1, Homo sapiens Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1; EREG, Homo sapiens EEpiregulin; SERPINE2, Homo sapiens Serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2; SERPINA3, Homo sapiens Serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3; TNFAIP6, Homo sapiens tumor necrosis factor, alpha-induced protein 6; INHBA, Homo sapiens Inhibin, beta A (activin A, activin AB alpha polypeptide); SPRY2, Homo sapiens Sprouty homolog 2 (Drosophila); FDX1, Homo sapiens ferredoxin 1; RGS2, Homo sapiens Regulator of G-protein signalling 2; NRP1, Homo sapiens Neuropilin 1; EGR1, Homo sapiens Early growth response 1; PGK1, Homo sapiens Phosphoglycerate kinase 1; BACH2, Homo sapiens BTB and CNC homology 1, basic leucine zipper transcription factor 2; IL6ST, Homo sapiens Interleukin 6 signal transducer (gp130, oncostatin M receptor); GAPDH, Homo sapiens Glyceraldehyde-3-phosphate dehydrogenase; ACTB, Homo sapiens Bêta actin; F, forward primer; R, reverse primer.

### Custom-made granulosa microarray design

A cDNA microarray of FC expressed sequence tags containing 2278 transcripts coding for more than 1200 different genes was made of human and bovine subtracted granulosa and cumulus cells libraries. Preliminary hybridizations (not shown) demonstrated that human probes can hybridize successfully with both human and bovine cDNA microarrays.

### Microarray hybridizations

Two hybridization experiments were performed with the custom-made granulosa microarray (Fig. 1). A first hybridization was done with the subtracted samples obtained by the SSH technique. A total of 1503 transcripts of the total 2278 transcripts demonstrated strong ratio for the positive group (potential true positive). In the second hybridization, we used a second pool of mural granulosa cells from different groups of patients (positive and negative groups) and obtained 593 transcripts with a strong ratio for the positive group. It is important to note that the material for these two hybridizations was obtained through different amplification procedures, SMART-PCR and T-7 amplification for groups from pool 1 and pool 2, respectively, leading to a more stringent group of genes being positive in both hybridizations. Comparison of the positive clone lists (ratio above 2 times more expressed...
in positive samples compared with negative samples) from the two different pools on the custom-made granulosa microarray resulted in the identification of 288 different genes with a strong intensity (ratio greater than 2) in both hybridizations. It represents 13% of the total transcripts of the custom-made granulosa microarray.

For the human Affymetrix GeneChip®, a total of 325 and 507 genes had a ratio higher than 2 between positive and negative groups for the two hybridizations, respectively (Figure 1). Comparison of the same two pools used for granulosa microarray hybridizations, a log2 ratio higher than 2 for the signal intensity was considered as expressed positive. Different parameters were used for the selection of potential candidates. Clones were selected and categorized according to their known functions, their hybridization intensities, their presence in more than one hybridization and their number of occurrences in the same library. Furthermore, we have selected clones with functions known to be involved in oocyte competence. A set of 18 different potential markers was selected.

### Real-time PCR candidate genes selection

The selection of the competent candidate genes was based on the result of the hybridizations on both platforms. For custom-made microarray hybridizations, a log2 ratio higher than 2 for the signal intensity was considered as expressed positive. There was little overlap between the gene lists and CYP19A1 is the only gene that had a strong ratio in all four hybridizations (Supplementary Table 1).

### Real-time PCR

Real-time PCR was performed with all three pools of human mural granulosa cells from each group (positive and negative groups). The expression of housekeeping genes ACTB and GAPDH was similar ($P > 0.05$) in both groups (Table IV). From the 18 candidate genes selected, five genes [FDX1 ($P = 0.0203$), CYP19A1 ($P = 0.0359$), cdc42 ($P = 0.0396$), SERPINE2 ($P = 0.0499$) and $3\beta$HSD 1 ($P = 0.0078$)] had a statistical difference between mural granulosa positive and negative groups ($P < 0.05$) (Fig. 2). The $3\beta$HSD 1 ($P = 0.0078$) had a higher gene expression in the positive groups ($P < 0.01$) (Fig. 2). Genes such as EGR1 ($P = 0.1117$), PGK1 ($P = 0.1231$), NRP-1 ($P = 0.1424$), RGS2 ($P = 0.1456$) and SERPINA3 ($P = 0.1712$) were not statistically different between the two groups, mainly due to larger variation in the levels measured, but could be considered as potential indicators of follicular competence. No differences in the transcript levels were observed between the two groups in the eight other selected genes (Table IV).

### Discussion

Results presented here have identified five potential follicular markers associated with embryo quality resulting in a successful pregnancy in humans. The markers are FDX1, CYP19A1, cdc42, SERPINE2 and $3\beta$HSD1.

We believe that the markers are likely to originate from mural granulosa cells, although we are aware that with the method of follicular fluid aspiration, it is difficult to obtain a pure sample of granulosa cells. The FCs may contain some cumulus cells, blood cells and perhaps some stroma/theca cells, but our protocols were designed to reduce the chances of contaminates appearing in the candidate genes. First, the subtractive hybridization should have removed any

<table>
<thead>
<tr>
<th>Human subtracted library</th>
<th>Number of sequences</th>
<th>Average insert size (bp)</th>
<th>Unigen target</th>
<th>Sequences with known function</th>
<th>Sequences with uncharacterized functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Granulosa</td>
<td>852</td>
<td>457</td>
<td>645</td>
<td>89%</td>
<td>11%</td>
</tr>
<tr>
<td>Cumulus</td>
<td>842</td>
<td>416</td>
<td>465</td>
<td>68%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Known sequences match with a sequence already characterized; Uncharacterized sequences match with a clone, BAC, RIKEN or hypothetical protein.

Table III. Characteristics of human subtracted libraries.
contaminant clones if present in both the positive and negative groups. Second, for clinical aspects, the cell population present in the analyzed samples must reflect the biological tissue samples recovered in normal IVF; in that case, positive markers could be useful even if not of granulosa cells origin. Lastly, it has been shown that samples with 75% purity were indistinguishable from the pure sample in gene expression profiles using both custom-made arrays and the Affymetrix microarray technology (Szaniszlo et al., 2004).

This study incorporated two platforms, custom-made microarrays and the Affymetrix arrays, with two different species in order to strengthen the criteria for candidate gene selection. Two different hybridizations were performed with the custom-made microarray, first to remove false positives

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mean positive groups (fg)</th>
<th>Positive groups ± SEM</th>
<th>Mean negative groups (fg)</th>
<th>Negative groups ± SEM</th>
<th>Difference between means</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>69.14</td>
<td>9.060</td>
<td>47.27</td>
<td>22.40</td>
<td>21.87 ± 24.37</td>
<td>0.4203</td>
</tr>
<tr>
<td>ACTB</td>
<td>7.371</td>
<td>1.731</td>
<td>6.580</td>
<td>0.4321</td>
<td>0.7911 ± 1.785</td>
<td>0.6805</td>
</tr>
<tr>
<td>EGR1</td>
<td>1.621</td>
<td>0.2788</td>
<td>0.9975</td>
<td>0.09431</td>
<td>0.6231 ± 0.2943</td>
<td>0.1117</td>
</tr>
<tr>
<td>NRP1</td>
<td>0.007751</td>
<td>0.001792</td>
<td>0.002427</td>
<td>0.001847</td>
<td>0.005324 ± 0.002692</td>
<td>0.1424</td>
</tr>
<tr>
<td>PGK1</td>
<td>0.6293</td>
<td>0.1911</td>
<td>0.2459</td>
<td>0.04668</td>
<td>0.3835 ± 0.1967</td>
<td>0.1231</td>
</tr>
<tr>
<td>SERPINA3</td>
<td>1.805</td>
<td>0.5112</td>
<td>0.9310</td>
<td>0.1187</td>
<td>0.8738 ± 0.5248</td>
<td>0.1712</td>
</tr>
<tr>
<td>RGS2</td>
<td>1.506</td>
<td>0.3919</td>
<td>0.7180</td>
<td>0.1926</td>
<td>0.7877 ± 0.4367</td>
<td>0.1456</td>
</tr>
<tr>
<td>DYSKL</td>
<td>0.3093</td>
<td>0.1203</td>
<td>0.2604</td>
<td>0.0974</td>
<td>0.0489 ± 0.1547</td>
<td>0.7677</td>
</tr>
<tr>
<td>EREG</td>
<td>0.07575</td>
<td>0.03159</td>
<td>0.06101</td>
<td>0.02902</td>
<td>0.0147 ± 0.04290</td>
<td>0.7485</td>
</tr>
<tr>
<td>SCARB1</td>
<td>0.2006</td>
<td>0.1638</td>
<td>0.04266</td>
<td>0.03938</td>
<td>0.1579 ± 0.1733</td>
<td>0.4137</td>
</tr>
<tr>
<td>SPRY2</td>
<td>0.2719</td>
<td>0.2468</td>
<td>0.1810</td>
<td>0.1388</td>
<td>0.09082 ± 0.2832</td>
<td>0.7645</td>
</tr>
<tr>
<td>TNFAIP6</td>
<td>0.6810</td>
<td>0.5002</td>
<td>0.4334</td>
<td>0.2343</td>
<td>0.2476 ± 0.5524</td>
<td>0.6772</td>
</tr>
<tr>
<td>IL6ST</td>
<td>0.7127</td>
<td>0.5254</td>
<td>0.4616</td>
<td>0.2535</td>
<td>0.2511 ± 0.5833</td>
<td>0.6891</td>
</tr>
<tr>
<td>INHBA</td>
<td>0.1904</td>
<td>0.0211</td>
<td>0.2462</td>
<td>0.0402</td>
<td>-0.0558 ± 0.04541</td>
<td>0.2865</td>
</tr>
<tr>
<td>BACH2</td>
<td>0.0278</td>
<td>0.0046</td>
<td>0.02916</td>
<td>0.0170</td>
<td>-0.008386 ± 0.01757</td>
<td>0.6580</td>
</tr>
</tbody>
</table>
and then with a new pool of RNA. These two hybridizations shared 288 genes with higher expression in the positive group compared with the negative group (12.64% of total transcripts on the granulosa microarray). With the Affymetrix Chip, only 24 transcripts (0.05% of the total transcripts on the chip) shown a ratio higher than 2 in both hybridizations. This is low considering that all the genes on the custom-made granulosa microarray were supposed to be present on the Affymetrix U133 array Chip\textsuperscript{x}, as it represents the human genome. Therefore, it is surprising that only one gene, the CYP19A1, was present in the four different hybridizations across the two chips. The vast majority of differentially expressed transcripts between closely related populations of cells with slight physiological differences fall into the low-abundance transcripts. It has been reported that Affymetrix GeneChips detect no more than 30% of the transcriptome of certain tissues and fail to detect at least 30% of expressed transcripts, leaving ~40% of transcripts falling into the range of unreliable detection (Evans et al., 2002). Affymetrix GeneChip underperformance can be due to probe design, distance of the target sequence from the polyA tail, secondary structures within the target sequence and cross-reactivity of the probe with other transcripts, each of which may influence the detection. This suggests that the Affymetrix array can only reveal regulation of mRNA species of medium to high abundance and can under-perform to detect accurately low-abundance transcripts which represent a large percentage of the transcriptome differentially expressed in FCs. Then, because the Affymetrix array is less accurate than cDNA custom-made array to discover differentially expressed transcripts in two similar populations (Cao et al., 2004), it can explain the fact that few transcripts are overlapped between the two Affymetrix hybridizations. Because we wanted to obtain a large screen of potential candidate genes and detect low-, medium- and high-abundance transcripts, we have performed hybridizations with Affymetrix GeneChip and custom-made microarrays. In accordance with previous mentioned studies, we have obtained more differentially expressed genes with the custom-made array which is more sensitive in the detection of low-abundance genes between two similar populations of cells. Since these two platforms can detect different levels of gene expression, the amount of overlap between these arrays is small. In accordance with these studies, our results suggest that custom-made granulosa microarrays have more sensitivity and accuracy to detect minute differences in gene expression in closely related tissue compared with the Affymetrix array. However, Affymetrix GeneChip microarrays combined with SSH methodology and Q-PCR could be used in association with identify differentially expressed genes.

Following candidate genes selection, expression level of 18 genes was more precisely assessed for their robustness as marker for their possible involvement in follicular competence and five (28%) genes were statistically significant between the positive groups and the negative groups. These results for the candidate validation are in accordance with similar microarray studies using library subtraction and further validation with Q-PCR (Fair, 2003; Dode et al., 2006).

In this study, three of the genes (FDX1, 3βHSD and CYP19A1) that are significantly more expressed in the follicles resulting in a pregnancy are involved in steroidogenesis. Both FDX1 and 3βHSD1 are responsible for progesterone synthesis, whereas CYP19A1 metabolizes androgen into estradiol-17β in granulosa cells (Conley and Hinshelwood, 2001; Simpson, 2004). Human adrenodoxin (ADX) is a member of the ferredoxin family and is a component in the electron transfer system for mitochondrial cytochrome P\textsubscript{450}. In mitochondria, pregnenolone is produced from cholesterol by ADX, ADX transferase and also cytochrome P\textsubscript{450}-side chain cleavage (cytochrome P\textsubscript{450} scc) (Niswender et al., 1994; Grinberg et al., 2000). Pregnenolone can then be metabolized to progesterone by the 3βHSD in granulosa cells (Adashi, 1994).

Following hCG administration, the mRNA expression of ADX-type FDX is strongly up-regulated in rat granulosa cells to reach maximum expression at 4 h post-treatment. Thereafter, mRNA expression gradually decreases until ovulation (12 h after hCG) (Espey and Richards, 2002). Similarly, LH or hCG is the major stimulator of 3βHSD mRNA expression in rat granulosa cells (Martel et al., 1990), and like ADX-FDX 1, 3βHSD mRNA expression decreases before ovulation (Bao et al., 1997). The expression of 3βHSD in bovine granulosa cells is higher in the dominant follicle than in other subordinate follicles, suggesting that 3βHSD may be associated in the selection mechanism of the dominant follicle (Bao et al., 1997). Furthermore, other studies showed that dominant follicles require expression of 3βHSD in human granulosa cells (Dupont et al., 1992). The P450 aromatase (CYP19A1) is well known to be stimulated by FSH and expressed in high concentrations in dominant follicles (Sisco et al., 2003). Therefore, higher expression level of these three enzymes appears to be related to hormonal induction (FSH and LH) of the production of steroid hormones (estrogen and progesterone) and possibly to follicular dominance mechanisms.

Level of SERPINE2 ($P = 0.0499$) and cdc42 ($P = 0.0396$) mRNA in human granulosa cells was also correlated with follicles that resulted in a pregnancy. SERPINE2 is a member of a family of protease inhibitors that use a conformational change to inhibit target enzymes. Serpins appear to be ubiquitous and are involved in a multitude of cellular functions, such as apoptosis and chromatin condensation (Silverman et al., 2001). The expression of SERPINE2 is higher in dominant follicles in the cow (Bedard et al., 2003), is increased by FSH (Cao et al., 2006) but decreases after the LH surge (Bedard et al., 2003; Ndiaye et al., 2005). The cdc42 is a member of the Rho family member of GTP-binding proteins involved in many cellular functions (Erickson and Cerione, 2001). cdc42 can delay the rate of apoptotic progression and then influences programmed cell death (Tu and Cerione, 2001).

The majority of genes found to be more expressed ($P < 0.05$), or tending to be more expressed ($0.05 > P < 0.2$), in positive groups are either known to be induced by the LH (hCG) surge [3βHSD, FDX 1, EGR1 (Espey et al., 2000; Yoshino et al., 2002), SERPINE2, PGK1 (Roy and Terada, 1999), RGS-2 (Ujioka et al., 2000), SERPINA3 (Sherwin et al., 2007)], expressed in the dominant follicle (3βHSD, 2006).
SERPINE2 and CYP19), involved in follicular development (NRP-1) (Shimizu et al., 2006) or in anti-apoptotic role (cdc42). However, for some of the genes found to be good indicators of oocyte competence, their expression is reported to be lower at the moment of ovulation: CYP19A1 (Hernandez-Gonzalez et al., 2006), 3βHSD (Bao et al., 1997), ADX (Espey and Richards, 2002), SERPINE2 (Bedard et al., 2003; Ndiaye et al., 2005), EGR1 (Espey et al., 2000; Espey and Richards, 2002), NRP-1 (Shimizu et al., 2006) and RGS-2 (Espey et al., 2000; Espey and Richards, 2002).

It is well known that the luteinization process begins before ovulation. In human IVF, during the ovarian stimulation protocol, the injection of LH/hCG in high concentrations can stimulate an early luteinization of the granulosa cells (Hillier et al., 1994; Maillet et al., 2005). In this study, genes selected in granulosa cells from follicles bearing a competent oocyte are known to be induced by LH/hCG. However, some genes, like ADX (Espey and Richards, 2002), CYP19A1 (Sanders and Stouffer, 1997) and SERPINE2 (Bedard et al., 2003), are known to be expressed in early stage of corpus lutenum formation. Following the LH surge, competent follicles could be those which have rapidly started the luteinization process.

A second possible hypothesis could be that the dominant follicle gradually acquires LH receptors (Bao et al., 1997) before the pre-ovulatory LH surge (Hugues and Cedrin-Durnerin, 2000). The LH receptor mRNA expression increases linearly with the increase of follicular diameter (Jolly et al., 1994). Therefore, in the context of the superovulation protocol, follicles that possess characteristics similar to a dominant follicle would contain more LH receptors. The desensitization of the receptor is achieved by the dissociation of its agonist (Ascoli et al., 2002). However, because hCG/LH has high affinity with the LH receptor, the dissociation of the agonist is considered irreversible (Ascoli et al., 2002). Thus, the association of LH with the receptor, the desensitization, phosphorylation and internalization is an important control of the presence of LH receptors. This process is able to limit the cellular response following activation of the receptor. High concentrations of LH/hCG lead to mass receptor internalization negative feedback and then desensitization of the target cell to LH (Combarnous et al., 2001). Therefore, in the context of the follicular stimulation protocol, the follicle with the highest sensitivity to LH would be the one responding most strongly with increased expression of the LH-inducible genes.

In summary, the microarray approach is a very useful tool for the discovery of new genes and to provide information with respect to oocyte competence. This technology will help to define the transcriptome of granulosa cells associated with a competent oocyte and also improve the selection of healthy oocytes/embryos resulting in good pregnancy rates. The information about genes expressed in competent follicles will also aid the refinement of hormonal treatments in human patients once the mechanism is fully understood.

Supplementary Data
Supplementary data are available at http://molehr.oxfordjournals.org/.

Acknowledgement
The authors acknowledge Dr Susan Novak for the reviewing of the manuscript.

Funding
MH is supported by Fonds de la Recherche en Santé du Québec (FRSQ) fellowship. MAS is a Project Leader of the Program on Oocyte Health funded under the Healthy Gametes and Great Embryos Strategic Initiative of the Canada Research Chair and Canadian Institutes of Health Research Canada (CIHR).

References
Markers of oocyte quality from subtracted libraries


