2009

Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes

Jose B. Cibelli

Amy E. Iager

Hasan H. Otu
University of Nebraska-Lincoln, hotu2@unl.edu

Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes

https://patents.google.com/patent/WO2011060080A2

Abstract

A genetic means of identifying "pregnancy competent" oocytes is provided. The means comprises detecting the level of expression of one or more genes that are expressed at characteristic levels (upregulated or downregulated) in cumulus cells derived from pregnancy competent oocytes. This characteristic gene expression level, or pattern referred to herein as the "pregnancy signature", also can be used to identify subjects with underlying conditions that impair or prevent the development of a viable pregnancy, *e.g.*, pre-menopausal condition, other hormonal dysfunction, ovarian dysfunction, ovarian cyst, cancer or other cell proliferation disorder, autoimmune disease and the like. In preferred embodiments the pregnancy signature will comprise one or more of ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

Classifications

G01N33/689 Chemical analysis of biological material, e.g. blood, urine; Testing involving biospecific ligand binding methods; Immunological testing involving proteins, peptides or amino acids related to pregnancy or the gonads
View 3 more classifications

WO2011060080A2

WO Application

Find Prior Art Similar

Other languages
French

Inventor
Jose B. Cibelli
Amy E. Iager
Hasan H. Otu

Original Assignee
Gema Diagnostics, Inc.

Priority date
2009-11-10

Family: US (1)EP (1)JP (1)CA (1)WO (1)

Date	App/Pub Number	Status
2010-11-10	PCT/US2010/056252	
2011-05-19	WO2011060080A2	Application
2011-11-17	WO2011060080A3	Application

Info

Patent citations (7)
Non-patent citations (1)
Cited by (3)
Legal events

External links

Description

GENES DIFFERENTIALLY EXPRESSED BY CUMULUS CELLS AND ASSAYS USING SAME TO IDENTIFY PREGNANCY COMPETENT

OOCYTES

Cross Reference to Related Applications

[0001] This application claims priority to US provisional application Serial No. 61/388,296 filed September 30, 2010; US provisional application Serial No. 61/387,313 and 61/387,286 both filed September 28, 2010; US provisional application Serial No. 61/360,556 filed on July 1, 2010 and US provisional application Serial No. 61/259,783 filed on November 10, 2009. This application also relates to US Serial No 11/584,580 filed on October 23, 2006 which is a continuation in part of US Serial No. 11/437,797 filed on May 22, 2006, which is in turn a continuation-in-part of US Serial No. 11/091,883 filed on March 29, 2005. and which in turn claims the benefit of provisional application No. 60/556,875 filed March 29, 2004. All of these applications are incorporated herein by reference in their entirety.

Field of the Invention

[0002] The present invention identifies a genus of 227 human genes, as well as a preferred set of 14 genes, the expression of which on cumulus cells correlates to whether an oocyte that is associated with said cumulus cell, or which is obtained from the same donor, are pregnancy competent, i.e., capable of resulting in a viable pregnancy upon in vitro fertilization. In addition the present invention provides gene expression detection methods and statistical analysis methods that resulted in the identification of these 227 genes and the identification of the preferred set of 14 genes the expression of which on cumulus cells correlates to oocyte competency.

[0003] Based on this discovery, the present invention provides methods and test kits for identifying human oocytes which are potentially suitable for use in IVF procedures by detecting the level of expression of one or more of these 227 genes, or one or more of these 14 genes, by a cumulus cell associated with said oocyte or derived from the same donor. In addition, based on this discovery the invention further provides test kits for the identification of human oocytes that when fertilized and when transferred to a suitable uterine environment are more likely, to yield a viable pregnancy. The set of 227 genes, the expression of which on cumulus cells correlates to pregnancy potential are contained in Table 4 infra In addition the preferred set of 14 genes are found in Table 12 and consist of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0004] Based on the foregoing, the present invention further provides genetic methods of identifying female subjects, preferably human females, having impaired fertility function, e.g., as a result of impaired ovarian function because of age (menopause), underlying disease condition or drug therapy by analyzing the expression of one or more of these 227 specific genes contained in Table 4 or the preferred set of 14 genes on cumulus cells obtained from oocytes isolated from said female subject. [0005] Also, the invention provides methods of evaluating the efficacy of a putative fertility or hormonal treatment by

assessing its effect on the expression of one or more of these 227 or 14 specific genes by cumulus cells of a female subject receiving this fertility or hormonal treatment.

[0006] Background of the Invention

[0007] Currently, there is no reliable commercially available genetic or non- genetic procedure for identifying whether a female subject produces oocytes that are "pregnancy competent", i.e., oocytes which when fertilized by natural or artificial means are capable of giving rise to embryos that in turn are capable of yielding viable offspring when transferred to an appropriate uterine environment. Rather, conventional fertility assessment methods assess fertility e.g., based on hormonal levels, visual inspection of numbers and quality of oocytes, surgical or non-invasive (MRI) inspection of the female reproduction system organs, and the like. Often, when a woman has a problem in producing a viable pregnancy after a prolonged duration, e.g., more than a year, the diagnosis may be an "unexplained" fertility problem and the woman advised to simply keep trying or to seek other options, e.g., adoption or surrogacy.

[0008] Perhaps in part of the lack of a means for identifying pregnancy competent oocytes, the success rate for assisted reproductive technology (ART), pregnancy and birth rates following in vitro fertilization (IVF) attempts remain low. Subjective morphological parameters are still a primary criterion to select healthy embryos used for in IVF and ICSI programs. However, such criteria do not truly predict the competence of an embryo. Many studies have shown that a combination of several different morphologic criteria leads to more accurate embryo selection. Morphological criteria for embryo selection are assessed on the day of transfer, and are principally based on early embryonic cleavage (25-27h post insemination), the number and size of blastomeres on day two, day three, or day five, fragmentation percentage and the presence of multi-nucleation in the 4 or 8 cell stage (Fenwick et al., Hum Reprod, 17, 407-12. (2002).

[0009] A recent study has shown that the selection of oocytes for insemination does not improve outcome of ART as compared to the transfer of all available embryos, irrespective of their quality (La Sala et al., Fertil Steril. (2008)).

[0010] There is a need to identify viable embryos with the highest implantation potential to increase IVF success rates, reduce the number of embryos for fresh replacement and lower multiple pregnancy rates. For all these reasons, several biomarkers for embryo selection are currently being investigated (Haouzi et al., Gynecol Obstet Fertil, 36, 730-742. (2008); He et al., Nature, 444, 12-3. (2006)).

[0011] As embryos that result in pregnancy differ in their metabolic profiles compared to embryos that do not, some studies are trying to identify a molecular signature that can be detected by non-invasive evaluation of the embryo culture medium (Brison et al.,. Hum Reprod, 19, 2319-24. (2004); Gardner et al., Fertil Steril, 76, 1175-80. (2001); Sakkas and Gardner, Curr Opin Obstet Gynecol, 17, 283-8 (2005); Seli et al., Fertil Steril, 88, 1350-7. (2007); Zhu et al. Fertil Steril.

(2007).

[0012] Genomics are also providing vital knowledge of genetic and cellular function during embryonic development. McKenzie et al., Hum Reprod, 19, 2869- 74. (2004); Feuerstein et al, Hum Reprod, 22, 3069-77 have reported, that the expression of several genes in cumulus cells, such as cyclooxygenase 2 (COX2), was indicative of oocyte and embryo quality. In addition Gremlin 1 (GREM1), hyaluronic acid synthase 2 (HAS2), steroidogenic acute regulatory protein (STAR), stearoyl-coenzyme A desaturase 1 and 5 (SCD1 and 5), amphiregulin (AREG) and pentraxin 3 (PTX3) have also been reported to be positively correlated with embryo quality (Zhang et al., Fertil Steril, 83 Suppl 1, 1169-79. (2005)). More recently, the expression of glutathione peroxidase 3 (GPX3), chemokine receptor 4 (CXCR4), cyclin D2 (CCND2) and catenin delta 1 (CTNND1) in human cumulus cells have been shown to be inversely

correlated with embryo quality, based on early-cleavage rates during embryonic development (van Montfoort et al., (2008) Mol Hum Reprod, 14, 157-68.(2008)).

[0013] Also Cillo et al., Reprod. 134:645-50 (2007) suggests a correlation between the expression of certain cumulus genes, i.e., HAS2, GREM1 and PTX3 and oocyte quality and embryo development. Still further Assidi et al. Biol. Reprod. 79(2) 209-222 (2008) suggest a correlation as to the expression of certain cumulus genes, i.e., EGFR, CD44, HAS2, PTSG2 and BTC and oocyte quality and development of embryos therefrom. Further, Bettegowda et al., Biol. Reprod. 79(2):301-309 (2008) suggest a correlation as to the expression of certain proteinase cathepsin genes and bovine oocyte quality and development of offspring therefrom.

[0014] In addition, a patent was recently issued to Zhang et al. (August 11, 2009) claims the detection of pentraxin 3 and a BCL-2 member on cumulus cells to assess oocyte quality. Also, US20040058975 published on March 25, 2004 teaches that antagonism of the EP2 receptor and/or cycloxygenase COX-2 promotes cumulus cell proliferation and oocyte development.

[0015] Also, while early cleavage has been shown to be a reliable biomarker for predicting pregnancy (Lundin et al, Hum Reprod, 16, 2652-7. (2001); Van Montfoort et al., Hum Reprod, 19, 2103-8 (2004; Yang et al, Fertil Steril, 88, 1573-8 (2007)), little has been reported correlating gene expression profiles of cumulus cells with respect to pregnancy outcome (but see Assou et al., Mol Hum Reprod. 2008 Dec; 14(12):711-9. Epub 2008 Nov 21).

[0016] Therefore, notwithstanding the foregoing, providing alternative and more predictive methods for identifying oocytes suitable for use in ΓYT procedures and in identifying the genetic bases of fertility problems in women would be highly desirable. In particular an identification of other genes, and biomarkers, the expression of which by cumulus cells correlates to pregnancy competency of oocytes and test kits and assays using same would be highly desirable as this could enhance the outcome of IVF procedures. [0017] These methods and test kits would in addition provide for the identification of women with oocyte related fertility problems, which is desirable as such fertility problems may correlate to other health issues that preclude pregnancy, e.g., cancer, menopausal condition, hormonal dysfunction, ovarian cyst, or other underlying disease or health related problems.

Brief Description and Objects of the Invention

[0018] The present invention relates to a method for selecting a competent oocyte, comprising a step of measuring the expression level of one of 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0019] The present invention also relates to a method for selecting a competent embryo, comprising a step of measuring the expression level of specific genes in a cumulus cell surrounding the embryo, wherein said genes are those in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0020] The present invention also relates to a method for selecting a competent oocyte or a competent embryo, comprising a step of measuring in a cumulus cell surrounding said oocyte or said embryo the expression level of one or more genes selected from the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSPl, GPR137B, IDUA, KCTD5,

KRAS, NCAMl, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0021] Aberrant expression levels of one or more of these genes are predictive of a non competent oocyte or embryo due to early embryo arrest.

[0022] As discussed infra, it has been found that the level of expression of these genes by a cumulus cell of a woman donor correlates to the likelihood that an oocyte associated with said cumulus cell or derived from the same subject are "pregnancy competent" when fertilized by natural or artificial means. These genes and expression levels constitute what Applicants refer to as the "pregnancy signature". In addition the pregnancy signature may further include one or more of the genes disclosed in Applicant's prior applications identified supra.

[0023] It is a related object of the invention to provide a novel method of determining whether an individual has a genetic associated fertility problem which potentially renders the individual's oocytes unsuitable for use in IVF methods based on the detected level of expression of one or more genes or corresponding polypeptides which constitute the "pregnancy signature." The genes and gene products which constitute the pregnancy signature are again preferably selected from those contained in Table 4 and/or are selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP. [0024] It is another object of the invention to provide a method of evaluating the efficacy of a female fertility treatment which comprises:

(i) treating a female subject putatively having a problem that prevents or inhibits her from having a "viable pregnancy" and

(ii) isolating at least one oocyte from said female subject and cells associated therewith after said fertility treatment;

(iii) isolating at least one cumulus cell associated with said isolated oocyte, and detecting the level of expression of at least one gene selected from those in Table 4 or at least one gene selected from ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP that is expressed at a characteristic level of expression in "pregnancy competent" oocytes; and

(iv) determining the putative efficacy of said fertility treatment based on whether said gene is expressed at a level characteristic of "pregnancy competent" oocytes as a result of treatment.

[0025] It is another specific object of the invention to provide novel methods of treating infertility by modulating the expression of one or more genes that constitute the pregnancy signature. These methods include the administration of compounds that agonize or antagonize the expression of one or more of the genes contained in Table 4 or ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, ERAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and

TERF2IP and their splice or allelic variants.

[0026] It is another object of the invention to provide animal models for evaluating the efficacy of putative fertility treatments comprising identifying genes which are expressed at characteristic levels in cumulus cells associated with pregnancy competent oocytes of a non-human animal, e.g., a non-human primate; and assessing the efficacy of a putative fertility treatment in said non- human animal based on its effect on said gene expression levels, i.e., whether said treatment results in said gene expression levels better mimicking gene expression levels observed in cumulus cells associated with pregnancy competent oocytes, ("pregnancy signature"), i.e. one or more of the 227 genes in Table 4 or one or more of the 14 gene genus consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

Detailed Description of the Figures

[0027] In Figure la-c, the inventors separately show the clustering of all F and N samples (65 samples), training set samples (33 samples), and validation set samples (32 samples) using all genes (a: all

samples, b: Training set, c: Validation set). Samples are clustered with hierarchical clustering utilizing average linkage method using Pearson's correlation as the metric of similarity following row normalization Sneath, P. (1973) Numerical taxonomy; the principles and practice of numerical classification. W. H. Freeman, San Francisco, CA USA. Figure 1: Clustering of N and F samples using all genes (a: all samples, b: Training set, c: Validation set).

[0028] Figure 2a-b: shows the clustering of N and F samples using 1180 descriptive genes (a: Training set, b: Validation set). The results therein show differential expression based on t-test (p<0.05 with Bonferroni correction for multiple hypothesis testing) which were identified in the training set (F vs. N)· Resulting 1180 genes, called "descriptive genes", were used to cluster Training and Validation sets separately.

[0029] Figure 3 shows the clustering of N and F samples using 227 predictive genes (a: Training set, b: Validation set). The data reveals that the only sample incorrectly predicted in the training set is misplaced in the clustering as well, however, the mixed behavior of F and N samples in validation set clustering emphasizes the contribution made by the weighted voting approach.

[0030] Figure 4 schematically depicts methods to A) assign significance to the predictor gene set (PG's) in Table 4, B) Refine PG's, and C) further analyze final predictor gene set

[0031] Figure 5. For each gene, number of samples for which the gene has a value of 40 is shown. Results are calculated for the "old" 35 samples. In Figure 5, we show the number of samples with a value of 40 for each gene, separately plotted for our genes (196 genes labeled "Hasan genes") and all 379 genes on TLDA (labeled "All genes"). [0032] Figure 6. Number of genes with a value of 40 is shown for each sample.

Results are calculated for the "old" 35 samples: For each gene, number of samples for which the gene has a value of 40 is shown. Results are calculated for the "new" 14 samples.

[0033] Figure 7. For each gene, number of samples for which the gene has a value of 40 is shown. Results are calculated for the "new" 14 samples.

[0034] Figure 8. Number of genes with a value of 40 is shown for each sample. Results are calculated for the "new" 14 samples.

[0035] Figure 9 :: Distribution of genes based on following factors: Group the gene belongs to (P or A); Agreement of the gene's up/down regulation in TLDA and microarray (10, if the direction is the same and -10, otherwise); Number of samples for which the gene has a value of 40. The analysis is performed separately for scaled and unsealed values with varying number of outliers excluded

[0036] Detailed Description of the Invention

[0037] Prior to discussing the invention in more detail, the following definitions are provided. Otherwise all words and phrases in this application are to be construed by their ordinary meaning, as they would be interpreted by an ordinary skilled artisan within the context of the invention. [0038] "Pregnancy-competent oocyte": refers to a female gamete or egg that when fertilized by natural or artificial means is capable of yielding a viable pregnancy when it is comprised in a suitable uterine environment.

[0039] "The term "competent embryo" similarly refers to an embryo with a high implantation rate leading to pregnancy. The term "high implantation rate" means the potential of the embryo when transferred in uterus, to be implanted in the uterine environment and to give rise to a viable fetus, which in turn develops into a viable offspring absent a procedure or event that terminates said pregnancy.

[0040] "Viable-pregnancy": refers to the development of a fertilized oocyte when contained in a suitable uterine environment and its development into a viable fetus, which in turn develops into a viable offspring absent a procedure or event that terminates said pregnancy.

[0041] "Cumulus cell" refers to a cell comprised in a mass of cells that surrounds an oocyte. This is an example of an "oocyte associated cell". These cells are believed to be involved in providing an oocyte some of its nutritional and or other requirements that are necessary to yield an oocyte which upon fertilization is "pregnancy competent" (Buccione, R., Schroeder, A.C., and Eppig, J.J. (1990). Interactions between somatic cells and germ cells throughout mammalian oogenesis. Biol Reprod 43, 543-547.)

[0042] "Differential gene expression" refers to genes the expression of which varies within a tissue of interest; herein preferably a cell associated with an oocyte, e.g., a cumulus cell.

[0043] "Real Time RT-PCR": refers to a method or device used therein that allows for the simultaneous amplification and quantification of specific RNA transcripts in a sample.

[0044] "Microarray analysis": refers to the quantification of the expression levels of specific genes in a particular sample, e.g., tissue or cell sample.

[0045] "Pregnancy signature": herein refers to the normal level of expression of one or more genes or polypeptides that are selected or encoded by the specific genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP and their orthologs, splice or allelic variants wherein these genes or polypeptides are expressed in normal cumulus cells at levels which correlate to the likelihood that an oocyte that is associated with a cumulus cell which expresses said one or more genes or polypeptides at these characteristic levels are more likely to give rise to a viable pregnancy.

[0046] "Characteristic level of expression of a cumulus gene" herein with respect to a particular detected expressed nucleic acid sequence or polypeptide means that the particular gene or polypeptide is expressed at levels which are substantially similar to the levels observed in cumulus cells that are associated with a normal cumulus cell or one associated with a normal or developmentally competent oocyte.

[0047] By "substantially similar" is meant that the levels of expression of individual genes are preferably within the range of +/- 1-5 fold of the level of expression by a normal cumulus cell, more preferably within the range of +/- 1-3 -fold, still more preferably within the range of +/- 1-1.5 fold and most preferably within the range of +/- 1.0-1.3, 1.0-1.2 or 1.0- 1.2 fold of the detected levels of expression of the gene or polypeptide by a normal cumulus cell.

[0048] According to the invention, the oocyte may result from a natural cycle, a modified natural cycle or a stimulated cycle for cIVF or ICSI. The term "natural cycle" refers to the natural cycle by which the female or woman produces an oocyte. The term "modified natural cycle" refers to the process by which, the female or woman produces an oocyte or two under a mild ovarian stimulation with GnRH antagonists associated with recombinant FSH or hMG. The term "stimulated cycle" refers to the process by which a female or a woman produces one or more oocytes under stimulation with GnRH agonists or antagonists associated with recombinant FSH or hMG.

[0049] "Oocyte or cumulus cell determined to possess suitable pregnancy signature or to be pregnancy competent" refers to an oocyte or a cumulus cell associated with the oocyte or an oocyte derived from the same subject at around the same time (within 0-6 months) as the tested cumulus cell which has been determined to express at least one of the genes or polypeptides encoded by those in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4,

DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP or an ortholog or splice or allelic variant thereof in a manner characteristic of the level of

expression by a normal cumulus cell. Preferably at least 2 or 3 genes are expressed in a characteristic manner, more preferably at least 3-10 genes, 10-50 genes and even up to 100 genes or more of those contained in Table 4 or their allelic or splice variants. It should be understood that if the expression of numerous genes are evaluated in the subject genetic based assays, such as in the order of 10 or more, that a suitable pregnancy signature means that all or substantially all, i.e. at least 70-80% of the detected genes are expressed in a manner characteristic of a normal cumulus cell. For example if the expression of 10 genes is detected at least 7, 8 or 9 of the genes will preferably be expressed at the levels consistent with a normal cumulus cell, i.e. one associated with an oocyte capable of giving rise to a normal embryo and viable pregnancy.

[0050] In general with respect to the pregnancy signature the characteristic levels of expression is observed for at least 3-5, 5-10, 10 to 20, and potentially at least 50 to 100 genes, that are expressed at characteristic levels in cumulus cells, that surround "pregnancy competent" oocytes. This is intended to encompass the level at which the gene is expressed and the distribution of gene expression within cumulus cells analyzed. [0051] "Pregnancy signature gene": refers to a gene which is expressed at characteristic levels by a cumulus cell, which is associated with a normal or

"pregnancy competent" oocyte. These genes are contained in Table 4 and further include the 14 genes selected from the group consisting of ABCA6, DDIT4,

DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA,

SDF4, SLC26A3, and TERF2IP and their orthologs, splice and allelic variants.

In the table the genes are referenced by their name as well as Accession number.

It should be understood that the invention further encompasses detection of allelic and splice variants of these genes and orthologs.

[0052] "Probe suitable for detection of the expression of a pregnancy signature gene or polypeptide" refers to a nucleic acid sequence or sequences or ligand such as an antibody that specifically detects the expression of the transcribed gene or corresponding polypeptide. In a preferred embodiment expression is selected by use of real time PCR detection methods.

[0053] "IVF": refers to in vitro fertilization.

[0054] The term "classical in vitro fertilization" or "cIVF" refers to a process by which oocytes are fertilized by sperm outside of the body, in vitro. ΓvT is a major treatment in infertility when in vivo conception has failed. The term "intracytoplasmic sperm injection" or "ICSI" refers to an in vitro fertilization procedure in which a single sperm is injected directly into an oocyte. This procedure is most commonly used to overcome male infertility factors, although it may also be used where oocytes cannot easily be penetrated by sperm, and occasionally as a method of in vitro fertilization, especially that associated with sperm donation.

[0055] "Zona pellucida" refers to the outermost region of an oocyte.

[0056] "Method for detecting differential expressed genes" encompasses any known method for quantitatively evaluating differential gene expression using a probe that specifically detects for the expressed gene transcript or encoded polypeptide. Examples of such methods include indexing differential display reverse transcription polymerase chain reaction (DDRT-PCR; Mahadeva et al, 1998, J. Mol. Biol. 284:1391-1318; WO 94/01582; subtractive mRNA hybridization (See Advanced Mol. Biol.; R.M. Twyman (1999) Bios Scientific Publishers, Oxford, p. 334, the use of nucleic acid arrays or microarrays (see Nature Genetics, 1999, vol. 21, Suppl. 1061) and the serial analysis of gene expression. (SAGE)

See e.g., Valculesev et al, Science (1995) 270:484-487) and real time PCR (RT-PCR). For example, differential levels of a transcribed gene in an oocyte cell can be detected by use of Northern blotting, and/or RT-PCR.

[0057] A referred method is the CRL amplification protocol refers to the novel total RNA amplification protocol disclosed in Applicant's earlier applications that combines template-switching PCR and T7 based amplification methods. This protocol is well suited for samples wherein only a few cells or limited total RNA is available.

[0058] Preferably, the "pregnancy signature" genes are detected by hybridization of RNA or DNA to DNA chips, e.g., filter arrays comprising cDNA sequences or glass chips containing cDNA or in situ synthesized oligonucleotide sequences. Filtered arrays are typically better for high and medium abundance genes. DNA chips can detect low abundance genes. In the exemplary embodiment the sample may be probed with Affymetrix GeneChips comprising genes from the human genome or a subset thereof.

[0059] Alternatively, polypeptide arrays comprising the polypeptides encoded by pregnancy signature genes or antibodies that bind thereto may be produced and used for detection and diagnosis.

[0060] "EASE" is a gene ontology protocol that from a list of genes forms subgroups based on functional categories assigned to each gene based on the probability of seeing the number of subgroup genes within a category given the frequency of genes from that category appearing on the microarray.

[0061] Based on the foregoing the present invention provides a novel method of detecting whether a female, preferably human or non-human mammal, produces "pregnancy competent" oocytes or whether a particular oocyte is pregnancy competent. The method involves detecting the levels of expression of one or more genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP that are expressed at characteristic levels by cumulus cells associated with (surrounding) oocytes that are "pregnancy competent", i.e., these oocytes when fertilized by natural or artificial means (IVF), and transferred into a suitable uterine environment are capable of yielding a viable pregnancy, i.e., embryo that develops into a viable fetus and eventually an offspring unless the pregnancy is terminated by some event or procedure, e.g., a surgical or hormonal intervention.

[0062] As described herein the inventors have determined as set of genes expressed in cumulus cells that are biomarkers for embryo potential and pregnancy outcome. They demonstrated that genes expression profile of cumulus cells which surrounds oocyte correlated to different pregnancy outcomes, allowing the identification of a specific expression signature of embryos developing toward pregnancy. Their results indicate that analysis of cumulus cells surrounding the oocyte is a non-invasive approach for embryo selection.

[0063] The set of predictive genes in Table 4 and the 14 gene set identified in Table 12 are known human genes. However, the expression of these genes (on cumulus cells) had not heretofore been correlated to oocyte competency or embryo development. Therefore, this invention relates to a method for selecting a competent oocyte, comprising a step of measuring the expression level of specific genes in a cumulus cell surrounding said oocyte, wherein said genes include at least one of the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0064] The methods of the invention may further comprise a step consisting of comparing the expression level of the genes in the sample with a control, wherein detecting differential in the expression level of the genes between the sample and the control is indicative whether the oocyte is competent. The control may consist in sample comprising cumulus cells associated with a competent oocyte or in a sample comprising cumulus cells associated with an unfertilized oocyte.

[0065] The methods of the invention are applicable preferably to human women but may be applicable to other mammals (e.g., primates, dogs, cats, pigs, cows...).

[0066] The methods of the invention are particularly suitable for assessing the efficacy of an in vitro fertilization treatment. Accordingly the invention also relates to a method for assessing the efficacy of a controlled ovarian hyperstimulation (COS) protocol in a female subject comprising: i) providing from said female subject at least one oocyte with its cumulus cells; ii) determining by a method of the invention whether said oocyte is a competent oocyte.

[0067] Then after such a method, the embryologist may select the competent oocytes and in vitro fertilize them, for example using a classical in vitro fertilization (cIVF) protocol or under an intracytoplasmic sperm injection (ICSI) protocol.

[0068] A further object of the invention relates to a method for monitoring the efficacy of a controlled ovarian hyperstimulation (COS) protocol comprising: i) isolating from said woman at least one oocyte with its cumulus cells under natural, modified or stimulated cycles; ii) determining by a method of the invention whether said oocyte is a competent oocyte; iii) and monitoring the efficacy of COS treatment based on whether it results in a competent oocyte.

[0069] The COS treatment may be based on at least one active ingredient selected from the group consisting of GnRH agonists or antagonists associated with recombinant FSH or hMG.

[0070] The present invention also relates to a method for selecting a competent embryo, comprising a step of measuring the expression level of at least one of the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0071] The methods of the invention may further comprise a step consisting of comparing the expression level of the genes in the sample with a control, wherein detecting differential in the expression level of the genes between the sample and the control is indicative whether the embryo is competent. The control may consist in sample comprising cumulus cells associated with an embryo that gives rise to a viable fetus or in a sample comprising cumulus cells associated with an embryo that does not give rise to a viable fetus.

[0072] It is noted that the methods of the invention leads to an independence from morphological considerations of the embryo. Two embryos may have the same morphological aspects but by a method of the invention may present a different implantation rate leading to pregnancy. [0073] The methods of the invention are applicable preferably to human women but may be applicable to other mammals (e.g. primates, dogs, cats, pigs, cows...).

[0074] The present invention also relates to a method for determining whether an embryo is a competent embryo, comprising a step consisting in measuring the expression level of 45 genes in a cumulus cell surrounding the embryo, wherein said genes include at least one of the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0075] The present invention also relates to a method for determining whether an embryo is a competent embryo, comprising: i) providing an oocyte with its cumulus cells; ii) in vitro fertilizing said oocyte; and iii) determining whether the embryo that results from step ii) is competent by determining by a method of the invention whether said oocyte of step i), is a competent oocyte.

[0076] The present invention also relates to a method for selecting a competent oocyte or a competent embryo, comprising a step of measuring in a cumulus cell surrounding said oocyte or said embryo the expression level of one or more genes selected from at least one of the 227 genes in Table 4 or the 14

genes selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP. Aberrant expression of one or more of these genes selected may be predictive of a non competent oocyte or embryo, the inability of the embryo being unable to implant or of a non competent oocyte or embryo due to early embryo arrest.

[0077] The methods of the invention are particularly suitable for enhancing the pregnancy outcome of a female. Accordingly the invention also relates to a method for enhancing the pregnancy outcome of a female comprising: i) selecting a competent embryo by performing a method of the invention ;iii) implanting the embryo selected at step i) in the uterus of said female, wherein said female may or may not be the oocyte donor.

[0078] The method as above described will thus help embryologist to avoid the transfer in uterus of embryos with a poor potential for pregnancy out come. The method as above described is also particularly suitable for avoiding multiple pregnancies by selecting the competent embryo able to lead to an implantation and a pregnancy.

[0079] In all above cases, the methods described the relationship between genes expression profile of cumulus cells and embryo and pregnancy outcomes.

[0080] Methods for determining the expression level of the genes of the invention:

[0081] Determination of the expression level of the genes in the "pregnancy signature" i.e., at least one of the 227 genes in Table 4 or at least one of the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.can be performed by a variety of techniques. Generally, the expression level as determined is a relative expression level.

[0082] More preferably, the determination comprises contacting the sample with selective reagents such as probes, primers or ligands, and thereby detecting the presence, or measuring the amount, of polypeptide or nucleic acids of interest originally in the sample. Contacting may be performed in any suitable device, such as a plate, microtitre dish, test tube, well, glass, column, and so forth. In specific embodiments, the contacting is performed on a substrate coated with the reagent, such as a nucleic acid array or a specific ligand array. The substrate may be a solid or semi-solid substrate such as any suitable support comprising glass, plastic, nylon, paper, metal, polymers and the like. The substrate may be of various forms and sizes, such as a slide, a membrane, a bead, a column, a gel, etc. The contacting may be made under any condition suitable for a detectable complex, such as a nucleic acid hybrid or an antibody- antigen complex, to be formed between the reagent and the nucleic acids or polypeptides of the sample.

[0083] In a preferred embodiment, the expression level may be determined by determining the quantity of mRNA.

[0084] Methods for determining the quantity of mRNA are well known in the art. For example the nucleic acid contained in the samples (e.g., cell or tissue prepared from the patient) is first extracted according to standard methods, for example using lytic enzymes or chemical solutions or extracted by nucleic-acid-binding resins following the manufacturer's instructions. The extracted mRNA is then detected by hybridization (e. g., Northern blot analysis) and/or amplification

(e.g., RT-PCR). Preferably quantitative or semi-quantitative RT-PCR is preferred. Real-time quantitative or semi-quantitative RT-PCR is particularly advantageous. Other methods of amplification include ligase chain reaction

(LCR), transcription- mediated amplification (TMA), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA).

[0085] Nucleic acids having at least 10 nucleotides and exhibiting sequence complementarity or homology to the mRNA of interest herein find utility as hybridization probes or amplification primers. It is understood that such nucleic acids need not be identical, but are typically at least about 80% identical to the homologous region of comparable size, more preferably 85% identical and even more preferably 90-95% identical. In certain embodiments, it is advantageous to use nucleic acids in combination with appropriate means, such as a detectable label, for detecting hybridization. A wide variety of appropriate indicators are known in the art including, fluorescent, radioactive, enzymatic, or other ligands (e. g. avidin/biotin).

[0086] Probes typically comprise single-stranded nucleic acids of between 10 to 1000 nucleotides in length, for instance of between 10 and 800, more preferably of between 15 and 700, typically of between 20 and 500. Primers typically are shorter single- stranded nucleic acids, of between 10 to 25 nucleotides in length, designed to perfectly or almost perfectly match a nucleic acid of interest, to be amplified. The probes and primers are "specific" to the nucleic acids they hybridize to, i.e. they preferably hybridize under high stringency hybridization conditions (corresponding to the highest melting temperature Tm, e.g., 50 % formamide, 5x or 6x SCC. SCC is a 0.15 M NaCl, 0.015 M Na-citrate). The nucleic acid primers or probes used in the above amplification and detection method may be assembled as a kit. Such a kit includes consensus primers and molecular probes. A preferred kit also includes the components necessary to determine if amplification has occurred. The kit may also include, for example, PCR buffers and enzymes; positive control sequences, reaction control primers; and instructions for amplifying and detecting the specific sequences.

[0087] In a particular embodiment, the methods of the invention comprise the steps of providing total RNAs extracted from cumulus cells and subjecting the RNAs to amplification and hybridization to specific probes, more particularly by means of a quantitative or semi quantitative RT-PCR.

[0088] In another preferred embodiment, the expression level is determined by DNA chip analysis. Such DNA chip or nucleic acid microarray consists of different nucleic acid probes that are chemically attached to a substrate, which can be a microchip, a glass slide or a micro sphere- sized bead. A microchip may be constituted of polymers, plastics, resins, polysaccharides, silica or silica-based materials, carbon, metals, inorganic glasses, or nitrocellulose. Probes comprise nucleic acids such as cDNAs or oligonucleotides that may be about 10 to about 60 base pairs. To determine the expression level, a sample from a test subject, optionally first subjected to a reverse transcription, is labeled and contacted with the microarray in hybridization conditions, leading to the formation of complexes between target nucleic acids that are complementary to probe sequences attached to the microarray surface. The labeled hybridized complexes are then detected and can be quantified or semi-quantified. Labeling may be achieved by various methods, e.g. by using radioactive or fluorescent labeling. Many variants of the microarray hybridization technology are available to the man skilled in the art (see e.g. the review by Hoheisel, Nature Reviews, Genetics, 2006, 7:200- 210)

[0089] In this context, the invention further provides a DNA chip comprising a solid support which carries nucleic acids that are specific to at least one of the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, KRAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[0090] Other methods for determining the expression level of said genes include the determination of the quantity of proteins encoded by said genes.

[0091] Such methods comprise contacting the sample with a binding partner capable of selectively interacting with a marker protein present in the sample. The binding partner is generally an antibody that may be polyclonal or monoclonal, preferably monoclonal.

[0092] The presence of the protein can be detected using standard electrophoretic and immunodiagnostic techniques, including immunoassays such as competition, direct reaction, or sandwich type assays. Such

assays include, but are not limited to, Western blots; agglutination tests; enzyme-labeled and mediated immunoassays, such as ELISAs; biotin/avidin type assays; radioimmunoassays; Immunoelectrophoresis; immunoprecipitation, etc. The reactions generally include revealing labels such as fluorescent, chemiluminescent, radioactive, enzymatic labels or dye molecules, or other methods for detecting the formation of a complex between the antigen and the antibody or antibodies reacted therewith.

[0093] The aforementioned assays generally involve separation of unbound protein in a liquid phase from a solid phase support to which antigen-antibody complexes are bound. Solid supports which can be used in the practice of the invention include substrates such as nitrocellulose (e. g., in membrane or microtitre well form); polyvinylchloride (e. g., sheets or microtitre wells); polystyrene latex (e.g., beads or microtitre plates); polyvinylidine fluoride; diazotized paper; nylon membranes; activated beads, magnetically responsive beads, and the like. More particularly, an ELISA method can be used, wherein the wells of a microtiter plate are coated with an antibody against the protein to be tested. A biological sample containing or suspected of containing the marker protein is then added to the coated wells. After a period of incubation sufficient to allow the formation of antibody-antigen complexes, the plate (s) can be washed to remove unbound moieties and a detectably labeled secondary binding molecule added. The secondary binding molecule is allowed to react with any captured sample marker protein, the plate washed and the presence of the secondary binding molecule detected using methods well known in the art.

[0094] Alternatively an immunohistochemistry (1HC) method may be preferred. IHC specifically provides a method of detecting targets in a sample or tissue specimen in situ. The overall cellular integrity of the sample is maintained in 1HC, thus allowing detection of both the presence and location of the targets of interest. Typically a sample is fixed with formalin, embedded in paraffin and cut into sections for staining and subsequent inspection by light microscopy. Current methods of IHC use either direct labeling or secondary antibody-based or hapten- based labeling. Examples of known IHC systems include, for example, EnVision(TM) (DakoCytomation), Powervision(R) (Immunovision, Springdale, AZ), the NBA(TM) kit (Zymed Laboratories Inc., South San Francisco, CA), HistoFine(R) (Nichirei Corp, Tokyo, Japan).

[0095] In particular embodiment, a tissue section (e.g. a sample comprising cumulus cells) may be mounted on a slide or other support after incubation with antibodies directed against the proteins encoded by the genes of interest. Then, microscopic inspections in the sample mounted on a suitable solid support may be performed. For the production of photomicrographs, sections comprising samples may be mounted on a glass slide or other planar support, to highlight by selective staining the presence of the proteins of interest.

[0096] Therefore IHC samples may include, for instance: (a) preparations comprising cumulus cells (b) fixed and embedded said cells and (c) detecting the proteins of interest in said cells samples. In some embodiments, an IHC staining procedure may comprise steps such as: cutting and trimming tissue, fixation, dehydration, paraffin infiltration, cutting in thin sections, mounting onto glass slides, baking, deparaffination, rehydration, antigen retrieval, blocking steps, applying primary antibodies, washing, applying secondary antibodies (optionally coupled to a suitable detectable label), washing, counter staining, and microscopic examination.

[0097] The invention also relates to a kit for performing the methods as above described, wherein said kit comprises means for measuring the expression level the levels of at least one of the 227 genes in Table 4 or the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, KRAS, NCAM1, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP that are indicative whether the oocyte or the embryo is competent.

[0098] The invention is further illustrated by the following description of how the inventors determined that the expression of the 227 genes in Table 4 and the 14 gene set on a cumulus cell correlates to oocyte competency and embryo development upon implantation and working examples. However, these examples and description should not be interpreted in any way as limiting the scope of the present invention.

[0099] An exemplary means by which this was effected is described in detail as follows. [00100] The first aspect of reducing the invention to practice involved identifying genes which constitute the pregnancy signature in women and potentially other mammals and was achieved by identifying and comparing the expression of genes in cumulus cells collected from women donors which are pregnancy competent or not. This was effected by collecting cumulus cells from different human oocytes of donor women and implanting patients with one or two putatively fertilized eggs. These patients were then, based on the results of the implantation, divided into three groups based on full, partial, and no pregnancy. For each oocyte used in the process, the transcriptional profile of at least one cumulus cell surrounding the particular oocyte was determined using

Affymetrix HG 133 Plus 2 arrays containing over 54,000 transcripts. Patients were included in the study only if they did not meet any of the exclusion criteria identified in Table 1.

Table 1. Patient Exclusion Criteria

[00101] More particularly, in order to find gene signatures predictive of an oocyte's ability to produce a healthy baby, the inventors profiled the transcriptome of cumulus cells surrounding the oocyte using Affymetrix HG 133

Plus 2 arrays containing over 54,000 transcripts. Total RNA from individual cumulus samples was isolated using the PicoPure RNA isolation kit (Molecular

Devices, Sunnyvale, CA). Sample RNA was amplified using a protocol developed in-house which ensures faithful and consistent amplification of small amounts of

RNA to levels required for microarray analysis (Kocabas, et al., Proc Natl Acad

Sci U S A, 103, 14027-14032 (2006)).

[00102] Resulting amplified RNA (aRNA) was hybridized to the Affymetrix arrays. Thirty-six samples were used for which none of the embryo transfers led to successful pregnancies (labeled N for No success) and 30 samples for which all of the transfers led to successful pregnancies (labeled F for Full success). There were no known confounding factors to effect pregnancy success and relevant clinical parameters such as age or IVF cycle number did not vary significantly between the F and N groups.

[00103] Quality Control (QC) parameters were calculated for all 65 samples using Expression Console™ (EC) software freely available by the manufacturer (Affymetrix). All QC parameters including scaling factor (coefficient needed to equate the 2% trimmed mean of overall chip intensity), percentage of probe sets called present, 3'-5' ratios for spike and labeling controls and housekeeping genes were within acceptable ranges (as described in manufacturer's guidelines) for all the samples. There were no known confounding factors to affect pregnancy success and relevant clinical parameters such as oocyte age or IVF cycle number did not vary significantly (t-test $p > 0.05$) between F and N groups (see Table 1). Additional criteria for acceptance included absence of Polycystic Ovarian Syndrome (PCOS), no history of chemotherapy or radiation to the abdomen or pelvis, absence of >4 cm intramural or submucosal fibroids, and on the male side, no history of testicular biopsy and sperm count of > 5 million

[00104] In order to prove the soundness of the prediction model, F and N samples were divided randomly into training and validation sets. The goal was to find a predictive set of genes developed on the training set and then test the performance of the predictive genes on the validation set, which has not been used in development of the predictive model. This strategy (as opposed to using all the samples to develop a signature) prevents over-fitting and provides an assessment of predictive signature's robustness (Nevins, J.R. and Potti, A. (2007) Mining gene expression profiles: expression signatures as cancer phenotypes, Nat Rev Genet, 8, 601-609.)

[00105] As shown in Table 2, 33 samples (15F; 18N) were used in the training set and 32 samples (15F; 18N) in the validation set.

[00106] Samples used in training and validation sets are shown in Table 2.

Table 2. Samples used in the training and validation sets for prediction purposes.

[00107] In Figure 1, the inventors separately show the clustering of all F and N samples (65 samples), training set samples (33 samples), and validation set samples (32 samples) using all genes. Samples are clustered with hierarchical clustering utilizing average linkage method using Pearson's correlation as the metric of similarity following row normalization Sneath, P. (1973) Numerical taxonomy; the principles and practice of numerical classification. W. H.

Freeman, San Francisco, CA USA.

[00108] During the clustering process complete transcriptional profiling on the chips, i.e. all 54K+ transcripts are used. The clustering of all three data sets indicates a lack of separation based on pregnancy success. This in turn suggests the need for supervised learning analysis to find phenotype specific gene identification to correlate the expression results with success and to eventually identify a predictive gene set.

[00109] In order to find genes that correlate with success, genes that show differential expression based on t-test (p<0.05 with Bonferroni correction for multiple hypothesis testing) were identified in the training set (F vs. N). Resulting 1180 genes, called "descriptive genes", were used to cluster Training and Validation sets separately (Figure 2). The results show successful separation of N and F samples, especially in the training set as samples in this set have been used to identify the pregnancy signature genes in Table 4 and the 14 genes selected from the group consisting of ABCA6, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, RAS, NCAMI, NDNL2, OLFML3, PTPRA, SDF4, SLC26A3, and TERF2IP.

[00110] Next, using these 1180 genes, leave-one-out-cross-validation (LIOXV) was performed in the training set. In this method, first number of genes in the predictive gene set, say P, is fixed. Then one sample in the training set is left-out and top P genes using the remaining samples that differentiate between N and F are calculated. Using these P genes, the sample that is left out is predicted as N or F. This process is cycled through all 33 samples in the training set (leaving one out at a time). The total number of correct predictions is listed as the accuracy of the predictor on the training set.

[00111] During LIOXV process, different values for P, number of predictor genes, are tried and for ones that show good LIOXV prediction accuracy, these genes are applied on the validation set. The number of samples correctly predicted in the validation set is reported as prediction accuracy in the validation set. The smallest P that yields high training and validation accuracies, i.e. P for which accuracy graph reaches a plateau, are reported as the predictor gene set. Prediction algorithms employed were Weighted Voting (Golub et aL, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science, 286, 531-537(1999), Bayesian Compound Covariate); Wright et 1., Proc Natl Acad Sci U S A, 100, 9991-9996 (1999)); Diagonal Linear Discriminant, Dudoit et aL, citation, 97, 77-87); Nearest Centroid, k-Nearest Neighbors (Golub et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science, 286, 531-537 (1999)); Shrunken Centroids, (Tibshirani et al., Proc Natl Acad Sci U S A, 99, 6567-6572 (2002)); Support Vector Machines (Radmacher, et al., J Comput Biol, 9, 505-511 (2002)) and Compound Covariate (Hedenfalk et aL, N Engl J Med, 344, 539-548 (2001)). [00112] In our analysis the weighted voting approach performed the best prediction with a 227 gene predictor set yielding 97% LIOXV accuracy (32/33 correct predictions) and 88% (28/32 correct predictions) validation set accuracy.

These 227 genes, called "predictive genes" yielded significant (p<0.05) prediction results on both training and validation sets using Fisher's tests. Prediction results are shown in Table 3.

[00113] Table 3. Prediction results for training and validation data sets.

[00114] In Figure 3, we show clustering of training and validation sets using the 227 predictive gene list. Note that the only sample incorrectly predicted in the training set is misplaced in the clustering as well, however, the mixed behavior of F and N samples in validation set clustering emphasizes the contribution made by the weighted voting approach.

[00115] Based on the foregoing the inventors have identified a set of human genes (contained in Table 4) the expression of which may be assessed on cumulus cells and compared to the level of expression of said gene that correlates to that of a normal oocyte, i.e., one associated with an oocyte that is capable of giving rise to a viable pregnancy, in order to assess the pregnancy potential of an oocyte associated with said cumulus cell (or an oocyte from the same donor that was used to isolate the tested cumulus cell). Analysis of additional samples resulted in identification of a preferred set of 14 genes (contained in Table 12) the expression of which may be assessed on cumulus cells and compared to the level of expression of said gene that correlates to that of a normal oocyte, i.e., one associated with an oocyte that is capable of giving rise to a viable pregnancy, in order to assess the pregnancy potential of an oocyte associated with said cumulus cell (or an oocyte from the same donor that was used to isolate the tested cumulus cell).

[00116] While these lists of genes are clinically relevant, we anticipate that these lists of 227 and 14 predictor genes may be further refined, subject to functional analysis, and further validated in order to identify a more optimal pregnancy signature set of genes. [00117] With respect thereto, the following techniques which may be used to refine this list of predictor genes contained in Table 4 and in the 14 gene set contained in Table 12 are described below.

[00118] The inventors will assign significance to these 227 or 14 predictor genes (PG's) using two strategies: i) we will permute class labels, identify optimum 227 gene predictors using the same method applied previously, and calculate their performance on the perturbed data set; ii) we will test the performance of randomly chosen 227 gene predictors using the original data set. Performance comparison of PG'S against results obtained using aforementioned strategies are used to assess PG'S's significance. In order to refine PG'S, we will divide the complete data set into alternative training and validation sets and for each split calculate an optimum predictive gene set. These predictors are compared to each other and PG'S to obtain a final predictive gene set composed of genes consistently coming up as good predictors. This final set will then be evaluated for its significance and accuracy. Finally, we will further analyze this final predictor gene set for i) Gene Ontology (GO) functional classification to find significantly over-represented GO categories, ii) involvement in known biological pathways, gene networks, disease pathways, small molecule and drug target interactions, iii) promoter region analysis to identify common/distinct transcriptional regulatory elements and miRNA target analysis, and iv) their localization, secretion potential, and ligation properties. In what follows we explain in detail this overall workflow, which is summarized in Figure 4. [00119] We have previously applied the approaches defined here on finding predictive biomarkers in diabetes, myelodysplastic syndrome, chronic liver disease and various forms of cancer using both high throughput transcriptional profiling and proteomic data sets (See: Aivadoet al., Proc Natl Acad Sci U S A,

104, 1307-1312 2007).; Jones et al., Clin Cancer Res, 11, 5730-5739 (2007); Jones, et al., J Urol, 179, 730-736 (2007) ; Out et al. Diabetes Care, 30, 638-643 (2007) ;

Out et al., J Biol Chem, 282, 11197-11204 (2007).; Prall et al., Int J Hematol, 89,

173-187(2009); Spentzos et al., J Clin Oncol, 23, 7911-7918 (2005). and Zinkin et al, Clin Cancer Res, 14, 470-477 (2008). In all these studies, the biomarkers were found to be statistically significantly predictive and were independently validated both experimentally and analytically on new data sets that have not been used to define the predictor set.

[00120] Prediction Algorithm

[00121] The "signal to noise ratio" (SNR) is used to assess predictor value of a gene g (Golub et al., Science, 286, 531-537 (1999).) Let pF(g) and pN(g) be the mean value of gene g in F (successful pregnancy) and N (failed pregnancy) sample groups, respectively. Similarly, let oF(g) and oN(g) be the standard deviation of gene g in F and N sample groups, respectively. We define $SNR(g) = [uF(g) - pN(g)] / [oF(g) + oN(g)]$. This metric defines a neighborhood in RM around ideal gene expression vectors for both groups where $M = | F | + | N |$, total number of samples in the data set. SNR punishes genes with an expression highly deviant in either group and provides a signed ranking method for a gene's membership. In this case large positive values indicate a good predictor for the F group and large negative values (in absolute value) indicate a good predictor for the N group. We also define the boundary between the correlation between idealized expression patterns and a given gene g as $B(g) = [uF(g) + pN(g)]/2$.

[00122] In this method we assess the predictor gene set of P genes $G = \{gl, g2, gP\}$, a group of F and N samples and a new sample S to be predicted. The vote of gi, $1 < i < P$, is defined as $Vi = SNR(gi) [S(gi) - B(gi)]$, where S(gi) represents the signal value of gene gi in S. Vi represents how well S(gi) relates to the "behavior" of gi in F and N samples. If Vi is positive, we conclude that based on gi, S is predicted to be F and if Vi is negative gi predicts S as N. Cycling through all genes in the predictor set we obtain P votes and let VF be the sum of all positive votes and VN be the sum of all negative votes. If VF is greater than VN in absolute value, we predict sample S as F; otherwise we predict S as N. In our previous studies we have obtained a robust predictor gene set using a training set, which was tested on an independent validation set.

[00123] Significance of Predictor Gene Set

[00124] In order to obtain a robust predictor gene set, samples are randomly divided into training and validation sets. We have previously left out cross validation (LIOXV) using weighted voting to obtain the 227 predictor gene set. In this analysis method we let T and V be the number of F and N samples used in training and validation sets, respectively. In this context, we left one of the samples in the training group out and found P genes with the highest SNR value in T-l F and N samples in the training set. Using these P genes, the sample that was left out is predicted either as F or N using the procedure described above. Number of correct predictions amount to LIOXV prediction accuracy on the training set and minimum P with highest prediction accuracy is carried on to be tested on the validation set. In our preliminary studies P was found to be 227 and our prediction accuracies were 97% and 88% on training and validation sets, respectively.

[00125] We also assess the statistical significance of these 227 predictive genes. In order to assess this, we employ a two step strategy. In the first phase, we randomly shuffle the class labels of the samples used in the training set, i.e., we call a random subset of F samples as N and similarly call a random subset of N samples as F. Using this perturbed data set, we calculate LIOXV accuracy of 227 genes with the weighted voting method on the training set. In this setting, we choose 227 genes that have the highest SNR values using the permuted class labels. Assume the class label permutation is performed B times and the LIOXV accuracy of bth permutation is Ab. We assess LIOXV accuracy significance of our original 227 gene predictor set as $p = 1/B\sum I(Ab > 97\%)$, where I is an indicator function and assumes a value of 1 if $Ab > 97\%$ and 0 otherwise. Similarly, we let A'b be the prediction accuracy of 227 genes obtained using both permutations on the validation set. We then assess the prediction accuracy significance of our 227 gene predictor set on the validation set as $p = 1/B\sum I(A'b > 88\%)$. [00126] In the second phase to assess significance, we retain the original class labels but pick random 227 gene sets and evaluate their LIOXV accuracy on the training set and prediction accuracy in the validation set. We perform this random selection B times and let Ab and A'b be the LIOXV accuracy on the training set and prediction accuracy in the validation set for the bth selection, respectively. Similarly, the significance for LIOXV accuracy is calculated as $p =$

$1 B X I(Ab > 97\%)$, and the significance for the validation set prediction accuracy are calculated as $p = 1/B\sum I(A'b > 88\%)$. In both phases B=1000.

[00127] Further Refinement of the Predictor Gene Set

[00128] In order to further refine the 227 gene predictor set we also employ different training and validation set splits and look for the overlap in the resulting gene sets. With an attempt to find well defined predictors, in the refinement process, we do not split our whole data set in half into training and validation sets; rather, we use 75% of the samples for the training set and 25% of the samples for the validation set. Furthermore, we adopt a ten-fold cross validation strategy instead of a LIOXV. In this case, when finding a predictor on the training set, we do not leave one sample out at a time; instead, we leave 10% of the total number of samples in the training set out at a time. At each iteration, a predictor set is calculated using the remaining 90% samples and the left out 10% samples is predicted with the predictor set. Total number of correct predictions is then used to calculate prediction accuracy. [00129] For each data split, we evaluate the prediction accuracy on the validation set and the significance of the predictor gene set as described above.

We then intersect the predictor gene sets found at each split with each other and with the original 227 gene predictor set to find genes that consistently come up as good predictors. To this end, we form a more refined predictor gene set and calculate its prediction accuracy and significance on the overall data set using cross validation.

[00130] Clustering and Functional Analysis

[00131] Clustering of samples and genes are performed using Unweighted Pair Group Method with Arithmetic- mean (UPGMA)(Sneath, P.H.A. (1973) Numerical taxonomy; the principles and practice of numerical classification. W. H. Freeman, San Francisco, CA USA.), a hierarchical clustering technique used to construct a similarity tree, and principal components analysis (PCA)(Otu, et al„ J Biol Chem, 282, 11197-11204 (2007)). In hierarchical clustering, expression data matrix is row-normalized for each gene prior to the application of average linkage clustering and Pearson's correlation is used as the distance measure. In PCA, which projects multivariate data objects into a lower dimensional space while retaining as much of the original variance as possible, each sample is normalized to mean zero and standard deviation one.

[00132] Functional analysis is comprised of finding Gene Ontology (GO) categories in the gene lists of interest that warrant further investigation (Ashburner, et al., Nat Genet, 25, 25-29 (2000)). Expression Analysis Systematic Explorer (EASE) identifies biologically- relevant categories that are over- represented in the set and therefore may be of further interest (Hosack et al., Genome Biol, 4, R70) (2003).

[00133] To accomplish this, EASE maps each probe to an Entrez Gene identifier (Maglott, et al., Nucleic Acids Res, 35, D26-31 (2001)) that is associated with a GO category. GO Consortium (Geneontology.org) assigns each gene (where applicable) to one or more classes in the three GO categories: biological function, cellular process, and molecular function. EASE identifies GO categories in the input gene list that are over-represented using jackknife iterative resampling of Fisher exact probabilities, with Bonferroni multiple testing correction. The "EASE score" is the upper bound of the distribution of Jackknife Fisher exact probabilities. Categories containing low numbers of genes are under-weighted so that the EASE score is more robust than the Fisher exact test. EASE analysis will test the predictor gene list against all genes on the chip, and an EASE score are calculated for likelihood of overrepresentation of a GO category in the input list. Overrepresentation describes a group of genes that belong to a certain GO category, e.g. cell cycle, that appear more often in the given input list than would be expected to occur if the distribution were random. The EASE score is a significance level with smaller EASE scores indicating increasing confidence in overrepresentation. We select GO categories that have EASE scores of 0.05 or lower as significantly over-represented.

[00134] Pathway Analysis [00135] Pathway analysis, functional enrichment analysis or gene set analysis focuses on predefined gene sets or classes that are significantly regulated in a microarray study. We use Ingenuity Software Knowledge Base (1KB), (Redwood City, CA) to identify networks and pathways that

best explain underlying transcriptional regulations. 1KB uses interactions between genes and/or gene products based on manual curation of scientific literature providing a robust interaction database. Once a gene list of interest is analyzed the results are viewed in two ways: i) networks formed using input gene list and a limited expansion ii) known biological pathway that significantly host a subset of the input genes. Both results can further be analyzed in terms of drugs, small metabolites, functions, and diseases known to interact partly with the final gene networks. In this way, we identify pathways best explained by our predictive gene set and infer further perturbations to these networks. Furthermore, we are able to generate new interaction networks that do not necessarily fall in predefined canonical pathways.

[00136] Promoter and miRNA target analysis

[00137] We will interrogate 5kb upstream of genes in our predictive signature and analyze these identified promoter regions using promoter analysis and interaction tool, PAINT (Vadigepalli et al, OMICS, 7, 235-252 (2003)) and oPOSSUM (Ho Sui et al., Nucleic Acids Res, 33, 3154-3164 (2005)). These methods model the promoter regions of input gene lists and identify transcription regulatory elements (TRE) in these regions. The results are then analyzed for over-representations of TREs and discover potential transcription factors that are important in gene regulation for the underlying microarray data.

[00138] Another potential mechanism of regulation of predictive genes is via miRNAs that regulate gene expression primarily through post-transcriptional repression or mRNA degradation in a sequence-specific manner. We can identify miRNA targets sites for predictive genes using TargetScanS (http://www.targetscan.org/). TargetScanS is an improved version of the Targetscan that searches the UTRs for segments of perfect Watson-Crick complementarities to base 2-8 of the miRNA, calculates a folding free energy G for each miRNA-target site interaction using RNAeval, and then assigns a Z score to each UTR (Lewis et al., Cell, 115, 787-798 (2003)). This way miRNA targets in the input gene list are identified through computation of miRNA binding sites.

[00139] Prediction of secreted proteins and membrane ligands

[00140] Localization and transport of proteins in the cell are governed by intrinsic signals in their amino acid sequences. We can analyze our predictive gene list in this context via the proteins they code for. In cases where annotation of a protein is sufficient to inform us about its localization, secretion potential, ligation and other properties, a manual filtering will suffice to isolate targets for further study. When substantial annotation is lacking, we are able to predict aforementioned properties of the protein using a highly accurate computational algorithm, TargetP, that uses neural networks and takes in consideration the characteristic of the N terminus of the proteins (Emanuelsson et al. Nat Protoc,

2, 953-971 (2007)). In case of secreted protein predictions, it helps to filter out ones that have transmembrane domains which would lessen the secretion potential of the protein. The transmembrane domain prediction on these proteins is performed using the standalone version of TMHMM, which uses a hidden Markov model for prediction (Sonnhammer et al., Proc Int Conf Intell Syst Mol Biol, 6, 175-182 (1998)). We also can supplement this analysis by the approach defined in PSORT9 Horton et al, Nucleic Acids Res, 35, W585-587 (2007)), which uses k-nearest neighbor classification and weighted matrices resulting from gapless multiple alignment in its prediction strategy.

[00141] Alternative Strategies

[00142] In addition we can optionally enhance our analysis in two ways:

[00143] Signal value calculation and normalization: Although model based methods such as dChip (Li et al., Genome Biol, 2, RESEARCH0032 (2001).) and RMA (Irizarry et al., Nucleic Acids Res, 31, el5. (2001)) may be used to normalize and summarize gene chip data and have been shown to be superior to Affymetrix MAS 5.0 normalization (Hubbell et al., Bioinformatics, 18, 1585- 1592.(2002)), we have avoided using these methods as they depend on a baseline sample and do not adapt to addition of new

samples (Barash et al., Bioinformatics, 20, 839-846 (2004)). In other words, as new samples are added to the data set, previous model based signal analysis becomes invalid and complete analysis workflow is therefore repeated. Because of our constantly growing data set and need to validate our previous findings in newly added samples, the inventors use MAS 5.0, a signal value calculation and normalization method that is robust to addition of new samples. A new method called Probe Logarithmic Intensity Error (PLIER) workflow in the Expression Console from Affymetrix (www.affymetrix.com) includes quintile normalization and produces improved signal values by utilizing the probes affinities, empirical probe performance and by handling the error appropriately across low and high concentrations (Katz et al., BMC Bioinformatics, 7, 464 (2006). Although PLIER is not as robust as MAS 5.0 to addition of new samples, it provides a better performance compared to model based methods and therefore are our alternative strategy for signal value calculation and normalization.

[00144] Prediction algorithm:

[00145] In addition to weighted voting, which has been the best performing prediction algorithm in our previous studies, we plan to try the following prediction algorithms in case our results do not conform to the success criteria set forth in this application: Bayesian Compound Covariate (Dudoit et al., Journal of the American Statistical Association, 97, 77-87 (2002); Wright et al., Proc Natl Acad Sci U S A, 100, 9991-9996 (2003)), Diagonal Linear Discriminant, Nearest Centroid, k-Nearest Neighbors(Golub et al., Science, 286, 531-537 (1999), Shrunken Centroids (Tibshirani et al., Proc Natl Acad Sci U S A, 99, 6567-6572 (2002), Support Vector Machines (Radmacher et al., J Comput Biol, 9, 505-511 (2002)), and Compound Covariate (Hedenfalk et al., N Engl J Med, 344,

539-548 (2001)).

[00146] Validation of predictive gene set using qRT-PCR-based Taqman™ Low Density Arrays (LDAs)

[00147] Taqman LDAs (Applied Biosystems) are microfluidic plates which allow for the simultaneous qRT-PCR analysis of 384 genes, from very small amounts of RNA, and with high fidelity. We use a single custom LDA to validate two sets of genes: 1.) the PG'S identified in the preliminary microarray study, and 2.) a set supplied by our collaborator, Dagan Wells, of Oxford University. In an independent study, Wells and colleagues identified a set of genes expressed in cumulus cells which are associated with aneuploidy in oocytes. Cytogenetic studies have revealed that, in women with an average age of 32, one quarter of oocytes are aneuploid (Fragouli Cytogenet Genome Res; 114:30-38 (2006). By combining both of these gene sets, a more optimal stronger pregnancy signature may be identified.

[00148] Based thereon we randomly select 25 F (Full pregnancy success) and 24 N (No pregnancy success) cumulus samples from the set of 65 that were previously subjected to microarray analysis for processing on our custom LDA. These cDNA (remaining from microarray analysis) are prepared and processed on an ABI 7900HT machine, one sample per LDA, according to Applied Biosystems' Taqman LDA instructions. Absolute quantification of each transcript is performed. Resultant amplification data are analyzed using 7000 System SDS Software (Applied Biosystems) and each gene intensity value are normalized to a control gene.

[00149] Comparison with microarray results

[00150] We apply Spearman's rank correlation and Mann- Whitney U test to compare expression levels for each gene across samples obtained via microarray and LDA. We will adopt these non-parametric and rank-based approaches as distributional assumptions in parametric models may not be valid in the two platforms and the signal values may not be comparable between the two platforms due to the difference in underlying technologies and normalization methods. We also calculate the degree of fold change between the F and N samples used on LDA for each of the genes tested. If the direction of this change is in accord with what we see in microarray results, gene's expression profile across the two platforms remain not-significantly changed (p>0.05) and highly correlated (r>0.9), we will label these genes as

validated. We will assess validated genes' prediction power on samples used in LDA experiment using both LDA and microarray signal values separately. We also assess validated genes' prediction accuracy using the complete data set using microarray signal values. In these prediction processes we divide the samples into training and validation sets (75%-25% split) and calculate prediction accuracy by building a model on the training set and testing it on the validation set. In the prediction strategy we apply the weighted voting method as previously described.

[00151] Alternative Strategies [00152] If the predictor genes are not validated as described above on the LDA platform or if a good prediction result is not obtained using validated genes, the prediction strategy can be modified to obtain higher accuracy and genes well correlated with pregnancy outcome by sacrificing the robustness of the predictive gene signature. For this purpose, we do not split the original data set into training and validation sets to calculate a predictive gene signature. Instead, we use the complete microarray data set to build a predictor, calculate its accuracy, and assess its statistical significance as previously described. During this prediction strategy we will use leave-ten-fold-out cross validation applied on the complete data set using weighted voting.

[00153] Application of pregnancy predictors in other samples

[00154] 15 F (Full pregnancy success) and 15 N (No pregnancy success) new cumulus samples are collected. RNAs are isolated as described above. Reverse transcription are completed using ABI's High Capacity cDNA Reverse Transcription kit and cDNA amplification are completed using custom preamplification pools (ABI). Preamplification and qRT-PCR cycling will occur sequentially in an ABI 7900HT machine, one sample per LDA, according to Applied Biosystems' Taqman LDA instructions. Absolute quantification of each transcript is performed. Resultant amplification data are analyzed using 7000 System SDS Software (Applied Biosystems) and each gene intensity value are normalized to a control gene. [00155] Using the final predictor genes that have been validated on the LDA platform, we build a prediction model on the complete data set consisting of samples in the prospective study. We then apply this model on the new samples using the signal values obtained in the retrospective study. In the prediction strategy we apply weighted voting method as previously described.

[00156] Alternative Strategies

[00157] Optionally, we may repeat our experiments performed in the retrospective study in order to get a more precise signal value, or increase the number of samples used in order to potentially enhance the statistical power of our validation efforts.

[00158] Therefore, based on the foregoing, in preferred embodiments the inventive methods are used to identify women subjects who produce or do not produce pregnancy competent oocytes based on the levels of expression of a set of differentially expressed genes contained in Table 4 or the set of 14 genes contained in Table 12 or the corresponding encoded polypeptides. However, the inventive methods are applicable to non-human animals as well, e.g., other mammals, including dogs, bats, bovines, horses, avians, amphibians, reptiles, et al. For example, the subject invention may be used to derive animal models for the study of putative female fertility treatments., i.e. by screening for compounds that modulate the expression of one or more of the pregnancy signature genes in Table 4 or the 14 gene set contained in Table 12 or a complement thereof. Additionally, the present invention may be used to identify female subjects who have an abnormality that precludes or inhibits their ability to produce pregnancy competent oocytes, e.g., ovarian dysfunction, ovarian cyst, pre-menopausal or menopausal condition, cancer, autoimmune disorder, hormonal dysfunction, cell proliferation disorder, or another health condition that inhibits or precludes the development of pregnancy competent oocytes.

[00159] For example, subjects who do not express specific pregnancy signature genes at characteristic expression levels are screened to assess whether they have an underlying health condition that precludes them from producing pregnancy competent oocytes. Particularly, such subjects are screened to assess whether they are exhibiting signs of menopause, whether they have a cancer, autoimmune

disease or ovarian abnormality, e.g., ovarian cyst, or whether they have another health condition, e.g., hormonal disorder, allergic disorder, etc., that may preclude the development of "pregnancy competent" oocytes.

[00160] Additionally, the subject methods may be used to assess the efficacy of putative female fertility treatments in humans or non-human female subjects. Essentially, such methods will comprise treating a female subject, preferably a woman, with a putative fertility enhancing treatment, isolating at least one oocyte and associated surrounding cumulus cells from said woman after treatment, optionally further isolating at least one oocyte and associated surrounding cells prior to treatment, isolating at least one cumulus cell from each of said isolated oocytes; detecting the levels of expression of at least one gene that is expressed or not expressed at characteristic levels by cumulus cells that are associated with (surround) pregnancy competent oocytes; and assessing the efficacy of said putative fertility treatment based on whether it results in cumulus cells that express at least one pregnancy signature gene at levels more characteristic of cumulus cells that surround pregnancy competent oocytes (than without treatment). As noted, while female human subjects are preferred, the subject methods may be used to assess the efficacy of putative fertility treatments in non-human female animals, e.g., female non-human primates or other suitable animal models for the evaluation of putative human fertility treatments.

[00161] Still further, the present invention may be used to enhance the efficacy of in vitro or in vivo fertility treatments. Particularly, oocytes that are found to be "pregnancy incompetent", or are immature, may be cultured in a medium containing one or more gene products that are encoded by genes identified as being "pregnancy signature" genes, e.g., hormones, growth factors, differentiation factors, and the like, prior to, during, or after in vivo, or in vitro fertilization. Essentially, the presence of these gene products should supplement for a deficiency in nutritional gene products that are ordinarily expressed by cumulus cells that surround "pregnancy competent" oocytes, and which normally nurture oocytes and thereby facilitate the capability of these oocytes to yield viable pregnancies upon fertilization.

[00162] Alternatively, one or more gene products encoded by said pregnancy signature genes or compounds that modulate the expression of such genes may be administered to a subject who is discovered not to produce pregnancy competent oocytes according to the methods of the invention. Such administration may be parenteral, e.g., by intravenous, intramuscular, subcutaneous injection or by oral or transdermal administration. Alternatively, these gene products may be administered locally to a target site, e.g., a female ovarian or uterine environment. For example, a female subject may have her uterus or ovary implanted with a drug delivery device that provides for the sustained delivery of one or more gene products encoded by "pregnancy signature" genes, or modulators of such genes.

[00163] Thus, in general, the present invention involves the identification and characterization, in terms of gene identity and relative abundance, of genes that are expressed by, cumulus cells derived from an egg, preferably human egg, at the time of ovulation, preferably cumulus cells, the expression levels of which correlate to the capability of said egg to give rise to a viable pregnancy upon natural or artificial fertilization and transferral to a suitable uterine environment.

[00164] In one exemplary embodiment, of the invention assays the expression of any combination, i.e., at least 1, 2, 3, 4, 5, 6, 10, 50, ...100 200 ....227 of the genes in Table 4 or any combination of the 14 genes in Table 12 by cumulus cells relative to levels expressed by normal cumulus cells using known nucleic acid or protein detection testing methods as exemplified above as a means of assessing oocyte competency or whether an individual produces competent oocytes. [00165] However, while the invention in an exemplary embodiment will select any combination of the genes in Table 4 and Table 12 genes in a preferred embodiment the invention will assay the expression of at least 2 of the genes in

Table 12, more preferably at least 3 and up to and including all 14 of these genes as a means of assessing oocyte competency. In addition the inventive methods alternatively may be practiced by monitoring the expression levels of one or more of the differentially expressed cumulus cell expressed genes selected from those in Table 4 and 12, the expression of which correlates to oocyte competency, in association with other genes, the expression of which is similarly found to be predictive.

[00166] Table 4: Human Genes Differentially Expressed By Human Cumulus Cells Associated With Pregnancy Competency

TABLE 5:

HUMAN GENES USED TO VALIDATE DIFFERENTIAL EXPRESSION DATA

[00167] As shown in Table 4 we identified 227 genes predictive of IVF success and 4128 differentially expressed genes. On the TLDA chips, we represented 199 of these genes, representing 196 unique ones (B3GNT7, ATP8A1, and DNAJC15 are represented by two probe sets). Out of these 196, 141 belong to the predictive set (represented by P in the "group" column of Excel files) and 55 belong to differentially expressed set (represented by A after "AH" as all samples were used to find differentially expressed genes).

[00168] The output of TLDA experiments are threshold cycle (Ct) values, which are "the fractional cycle number at which the fluorescence passes the threshold". These values are in logarithmic scale (base 2) and are inversely related to expression. For example, let Ct values for genes X and Y be 18 and 19, respectively. This means we "detect" X "well" at 18th cycle, which is, in comparison to Y (detected well at 19th cycle) renders two times greater expression level. Hence, to calculate classical fold change between X and Y (X Y), we need to subtract X from Y and take that to power 2.

[00169] There are no spike-in controls to provide normalization across arrays. To do approximate normalization, Ct values of all genes are considered across samples and one with the most stable expression is chosen. Delta Ct (dCt) value for each gene in a plate (sample) is calculated by subtracting the Ct value of the "stable" gene (in that plate) from each gene's Ct value. Statminer analysis found the "stable" gene to be GAPDH.

[00170] Samples used on TLDA from our "old samples used in prediction analysis" are given in Table 6 below. In the Table the "*" corresponds to an incorrectly predicted sample in previous analysis.

[00171] Table 6. Samples from previous analysis used in TLDA experiments

Therefore, we have 19 F and 16 N samples for a total of 35. We also ran 14 new samples: 11 N 3 F, as shown in Table 7. Table 7. New samples used in TLDA ex eriments.

[00172] Data Quality

[00173] We checked distribution for the "detection" call of genes. If a gene's expression level is undetermined, it is assigned a value of 40. In Figure 5, we show the number of samples with a value of 40 for each gene, separately plotted for our genes (196 genes labeled "Hasan genes") and all 379 genes on TLDA (labeled "All genes"). In this Figure for each gene, number of samples wherein the gene has a value of 40 is shown. In the Figure the results are calculated for all samples.

[00174] In toto for our genes, we have 2237 values equal to 40 out of 6860 (196*35) values indicating that 32.6% of measurements yield undetected genes. In the inventive quantitative analyses it is permissible to have genes with a value of 40 as we expect some genes to be almost non-existent in some samples. By contrast, in our methods it is not permissible to have a gene undetected in all 35 samples. For example, for our genes, we have 62 genes for which there has been no detection for 30 or more samples. When all genes are considered, the percentage of measurements with a value of 40 is about 27.4%: 3631 out of 13230

(378*35).

[00175] In Figure 6, we show number of genes called 40 for each sample. This in a way shows the detection level we get for each sample. In this figure we see some samples with suboptimal overall detection like 9_072407, which has 117 of Hasan Genes and 212 of all genes with a value of 40. The average ± st. dev. value of number of genes called 40 across samples is 63.9 ± 14.5 for Hasan Genes and 103.7 ± 28.6 for all genes.

[00176] In addition, we effected a similar analysis for 14 "new" samples. Therein we see 541 out of 2744 (14*196) values equal to 40 in Hasan Genes (corresponding to 19.7% of measurements) and 896 out of 5292 (14*378) values equal to 40 in All Genes (corresponding to 16.9% of measurements). These are more than 10% better than what we see in "old" samples. We show in Figures 7 and 8 a number of samples and number of genes with a value of 40 when data is analyzed with respect to genes and samples, respectively.

[00177] Figure 7 shows for each gene, the number of samples for which the gene has a value of 40. Results are calculated for the tested 14 samples.

[00178] For our gene list, we have 25 genes that have 12 or more samples for which they have a value of 40. In Figure 8 the number of genes with a value of 40 is shown for each sample and results are calculated for the 14 "new" samples. [00179] Therein, we see one sample with bad overall detection: 17_100709_22, which has 85 of a human gene set that are differentially expressed on cumulus cells which we refer to as "Hasan Genes" and 140 of genes with a value of 40 in the old data set. The average ± st. dev. value of number of genes called 40 across samples is 38.6 ± 13.7 for Hasan Genes and 64.0 ± 22.8 for all genes, in the new data set, which is improved relative to the prior data set.

[00180] These results suggest that TLDA detection for the new samples was an improvement relative to the old samples and some sort of filtering on the genes would be required. This filtering could be based on number of samples for which the gene has an expression value of 40, i.e. undetected. As discussed above, for example, for the prior samples, we have numerous genes where the gene is not detected in more than 85% of the samples, regardless of sample's group.

[00181] Data Analysis

[00182] There are various issues to consider such as handling of data points that have a value of 40, calculating fold change, and whether or not to use logged values. Below, we address such issues providing potential solutions.

[00183] Scaling: We have two sets of output: Ct values (logged expression levels) and dCt values, where for a given sample, each gene's dC value is calculated by subtracting GAPDH's Ct value from the gene's Ct value. Since Ct values are logarithmic, this corresponds to dividing each gene's expression value by GAPDH's expression value. In other words, it is the fold change between a gene and GAPDH. Moving on with these values mean calculating fold change between groups based on genes' fold change with respect to GAPDH. Since

GAPDH is not one of the endogenous controls used on the array, there are no spike-in controls used in TLDA, and small variations in logarithmic scale may imply large differences in real values, we approach this with some caution. Nevertheless, we provide analysis both using scaled and unsealed values. For the remainder of this report unsealed values refer to Ct values as obtained in amplification files and scaled values refer to dCt values obtain by subtracting GAPDH.

[00184] Fold Change:

[00185] Assuming we have two samples A and B, and gene X's expression values in these samples are aX and bX, respectively. What we see in TLDA output (Ct values) are log (aX) and log (bX). If you want to calculate fold change between these two samples, you would subtract Ct values and take that to power of

2. That is, FC = 2 log (aX) - log (bX). The reason for this is the following rules: log p - log q = log (p/q) and 2log2p = p. However, since Ct values are reversed, i.e. a smaller value means larger expression, this FC gives you the fold change B/A. To exemplify, if we see a Ct value of 10.8 in A and 12.3 in B, this means this gene is upregulated in A and fold change for B/A is 2 10.8 - 12.3 = 2 -1.5 = 0.35. In other words, this gene is upregulated in A by 1/0.35 = 2.8 times. Another way to arrive this point is first to unlog Ct values and then calculate FC as we know it, except that the direction is reversed, i.e. in Ct world less means more. Hence, we have the expression level for A = 2 10.8 = 1782, the expression level for B = 2

12.3 = 5042, and FC B/A = 1782/5042 = 0.35.

[00186] FC values less than 1 are hard to interpret so what we do is we reverse them and put a minus sign. For the above example, instead of saying FC for B/A is 0.35, we say FC for B/A is -1/0.35 = -2.8. In all my calculations, we always subtracted F values from N values (if we were using log scale) or divided N values by F values (if we used unlogged values) and calculated FC for F/N. we used negative values to depict FCs less than 1 as explained above.

[00187] As if it has not been complicated enough to calculate a simple FC, we have more to think about. The example above contained only two samples, or, you can view it as having one sample in each group. How about if we have more than one sample in each group, as in our case (16 N, 19F)? If you average Ct values, you indeed get a geometric mean of expression levels. If you then subtract averages of Ct values in two groups and then take that to the power of two, this in turn means calculating FC by dividing geometric means of expressions in two groups. The reason for this is the following rules: alogX = logXa and logp + logq = log (pq).

[00188] To give an example, assume you have expression levels a, b, and c in group N and d, e, f, and g in group F. What we see in TLDA output is loga, logb, etc. In order to calculate FC (F/N), if we subtract the average value in F from the average value in N and then take that to power 2, we get the following:

[00189] Average in N = 1/3 [loga + logb + logc] = l/3log[abc] = log (abc)l/3 [00190] Average in F = 1/4 [logd + loge + logf + logg] = l/4log[defg] = log

(defg)l/4

[00191] FC (F/N) = 2 ^ [log (abc)l/3 - log (defg)l/4] = 2 ^ {log [(abc)l/3 / (defg)l/4]}= (abc)l/3 / (defg)l/4

[00192] Recall that geometric mean of n numbers is nth root of their products. Therefore, we always choose to work with unlogged values. That is, we first took Ct values to the power of 2 and then did my analyses.

[00193] 40: 40 is an arbitrary Ct value considered high enough to represent a gene that has not been detected. However, if you set it to 42 instead of 40, all your results will change. Therefore, we resolved this by first looking at all values that are not 40 and ranked them. For Hasan Genes, this corresponds to ranking 4623 values. We then looked at the bottom 2% of these genes, that is the lowest 92 genes; calculated their average and standard deviation, which turned out to be 37.9 and 0.8. We then replaced each 40 by a number randomly chosen between the interval [37.9-0.8, 37.9+0.8].

[00194] Outliers: When you manually look at the expression levels, you often see samples that behave as outliers for a given gene. In order to overcome this we removed the highest and lowest expression levels in a group (N or F) when calculating FC. We also repeated this procedure by removing highest two and lowest two samples in each group. [00195] In conclusion, using the foregoing statistical methods we found that the level of expression of genes in Table 4 (which set of genes includes ABCA6,

DDIT4, DUSP1, GPR173B, IDUA, KCTD5, NDNL2, SLC26A3 and TERF2IP as well as an additional 6 genes KRAS, NCAMI, OLFML3, PTPRA, and SDF4) by cumulus cells correlates to the capability of an oocyte associated therewith or from the same women donor to result in a viable pregnancy. Therefore, methods which detect the expression of one or more of these genes by a cumulus cell may be used in order to determine whether an oocyte associated therewith or from the same women donor is suitable for use in an IVF procedure.

[00196] To confirm these results we did a 14N vs. 17F comparison and a 12N and 15F comparison as described below.

[00197] Filtering Genes: There are some genes that are undetected in most of the samples. In the extreme, we have 20 genes designated "Hasan Genes" that are given a Ct value of 40 in all 35 samples. The use of these genes is not of predictive value. As a general approach, we eliminated genes that are undetected in 25 or more samples.

[00198] Results

[00199] Using the foregoing methods we have generated three data sets based on removal of outliers, where outlier is the sample(s) with highest or lowest expression in a group given a gene: i) we removed no samples (denoted by "remove none" in Excel file names) ii) we removed samples with highest and lowest expression value in each group (N and F), i.e. FC is calculated using 14N and 17F samples (denoted by "remove 1" in Excel file names) iii) we removed samples with top two highest and top two lowest expression values in each group (N and F), i.e. FC is calculated using 12N and 15F samples (denoted by "remove 2" in Excel file names).

[00200] For each data set, we handled 40s as explained above, i.e. we replaced them with random values between average of the lowest two percent of all detected values plus/minus standard deviation of those values. We then unlogged expression levels and generated two files for each data set: Scaled (for a given sample, each gene's expression level is divided by GAPDH's expression level in that sample) and Unsealed (no GAPDH or any other scaling is applied).

[00201] For each Excel file (totaling 6, 2 for each of 3 data sets), we also included columns "group": P (gene is from the predictive signature) A (gene is from differential expression analysis using combined training and validation samples, i.e. all samples in microarray analysis; "count 40": showing number of samples a gene assumes the value 40; "FC TLDA": FC calculated as ratio of averages; "FC Affy": FC coming from microarray analysis; "Agree": showing if the direction of increase is the same in TLDA and Affy; we used 10 to indicate agreement and -10 otherwise.

[00202] In Figure 9 we show the distribution of genes that are in agreement between TLDA and microarray. A tabulated form of these results is in Table 8, where we break down the level of agreement based on genes' up/down regulation from microarray results. Here, we use all genes, i.e. no filtering is applied based on the number of samples where the gene has been shown to be 40.

[00203] In Table 8 below we show the number of genes that show agreement in direction of upregulation between TLDA and microarray analysis in direction of upregulation between TLDA and microarray analysis. S: scaled, U: unsealed. P: predictive gene list. A: Genes obtained when all samples were analyzed together in microarray data set. T: Total genes. No genes were filtered out.

[00204] Table 8:

[00205] These results suggest removing top one or two outliers in each group and best overall agreement is achieved at 61% using scaled data. In addition, we repeated this analysis by filtering out genes that are "undetected" in 25 or more samples. Results showing the distribution of agreement between Affy and TLDA using genes that are detected in 11 or more samples are tabulated in Table 9.

Table 9.

[00206] These results are further optimized as there is an overall 66% agreement when the top 1 or 2 outliers are removed on either side in both groups and scaled values are used. Also, of interest is that in unsealed values, almost all genes show upregulation in F group.

[00207] Prediction

[00208] In the prediction analysis, we used genes that are in agreement between Affy and TLDA when genes that are "undetected" in 25 or more samples are filtered out (Table 9). We applied leave one out cross validation using weighted voting. Best results were obtained when unsealed data was used (27/35 = 77% prediction accuracy) with 6 genes. In case of scaled data, best prediction accuracy was (22/35 = 63%). These genes along with their fold change value (F N) in TLDA using 35 samples and in Affy is shown in Table 10, along with the fold change value (F/N) of these genes in "new" 14 samples.

Table 10. Six predictive enes and corres ondin fold chan e values.

[00209] Prediction results for the 35 samples are given in Table 11. Incorrectly predicted samples are shaded. In total, 3 N and 5 F samples were predicted incorrectly.

Table 11. LIOXV prediction results for the "old" 35 samples.

[00210] In conclusion, using the foregoing statistical methods we found that the level of expression of one or more genes in Table 4 and more preferably one or more of the 14 genes selected from the group consisting of ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPl, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP by cumulus cells correlates to the capability of an oocyte associated therewith or from the same women donor to result in a viable pregnancy. Therefore, methods which detect the expression of one or more of these genes by a cumulus cell may be used in order to determine whether an oocyte associated therewith or from the same women donor is suitable for use in an IVF procedure., as well as for identifying individuals with conditions that result in oocytes unsuitable for use in IVF procedures, and for monitoring the success of fertility treatments.

[00211] REFERENCES

[00212] Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

Claims

Claims
1. A non-invasive method of identifying oocytes that are capable of giving rise to a viable pregnancy when fertilized comprising the following steps:
(i) obtaining at least one cumulus cell associated with an oocyte that is to be tested for pregnancy competency from a female donor or for other oocytes of said donor;
(ii) assaying the expression of at least one gene by said at least one cumulus cell, the expression of which correlates to the capability of an oocyte associated with said cell to yield a viable pregnancy upon fertilization and transferal into a suitable uterine environment wherein said genes are selected from those in Table 4 and/or ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP or their orthologs, splice or allelic variants ; and
(iii) identifying, based on the level of expression of said at least one gene as compared to the characteristic level of expression by a cumulus cell associated with a pregnancy competent oocyte

whether said oocytes or another oocyte derived from said female donor is potentially capable of yielding a viable pregnancy upon fertilization and transferal into a suitable uterine environment.

2. The method of Claim 1 wherein the at least one gene the expression of which is detected is selected from the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

3. The method of Claim 2 wherein the method detects the expression of 2 or more of said genes.

4. The method of Claim 2 wherein said method detects the expression of 3 or more f said genes.

5. The method of claim 2 wherein said method detects the expression of 4 or more of sad genes.

6. The method of Claim 2 wherein said method detects the expression of 5 or more of said genes.

7. The method of Claim 2 wherein said method detects the expression of 6 or more of said genes.

8. The method of Claim 2 wherein said method detects the expression of 7 or more of said genes.

9. The method of Claim 2 wherein said method detects the expression of 8 or more of said genes.

10. The method of claim 1, wherein said oocyte is a mammalian oocyte.

11. The method of claim 10, wherein said oocytes is a human oocyte.

12. The method of claim 10, wherein said oocyte is a non-human primate oocyte.

13. The method of claim 1, wherein the expression of at least 5 genes are measured, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are selected from ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

14. The method of claim 13 wherein the expression of at least 6 genes are measured, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are selected from ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

15. The method of claim 14 wherein the expression of at least 7 genes are measured, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are selected from ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

16. The method of claim 15 wherein the expression of at least 8 genes are measured, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are selected from ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

17. The method of claim 14 wherein the expression of at least 9 genes are measured, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are selected from ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

18. The method of claim 14, wherein the expression of at least 10 genes, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are measured.

19. The method of claim 1, wherein the expression of at least 15 genes, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are identified.

20. The method of claim 1, wherein the expression of at least 20 genes, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy are identified.

21. The method of claim 1, wherein the expression of at least 20 to 50 genes, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy is assayed selected from those in Table 4.

22. The method of claim 1, wherein the expression of at least 50 to 100 genes, the expression of which correlates to the capability of an oocyte to potentially yield a viable pregnancy selected from those in Table 4 are assayed.

23. The method of claim 1 wherein the method of assaying gene expression uses a method that monitors differential gene expression.

24. The method of claim 14 wherein said method comprises indexing differential display reverse transcriptase polymerase chain reaction (DDRT- PCR)

25. The method of claim 1, wherein the oocyte is obtained from a human female who is at most 35 years old.

26. The method of claim 1, wherein the oocyte is obtained from a human female who is at least 25 years old.

27. The method of claim 16, wherein the oocyte is obtained from a human female who is at least 30 years old.

28. The method of claim 18, wherein the oocyte is obtained from a human female who is at least 35 years old.

29. The method of claim 19, wherein the oocyte is obtained from a human female who is at least 40 years old.

30. The method of claim 1, wherein the aberrant expression of said at least one gene is correlated to a condition selected from menopause, cancer, ovarian dysfunction, ovarian cyst, autoimmune disorder and hormonal dysfunction.

31. A method of assessing the efficacy of a fertility treatment comprising: (i) treating a human female with a putative fertility enhancing treatment;

(ii) obtaining an oocyte and cumulus cells associated therewith from said human female after treatment and measuring the expression of at least one gene selected from those contained in Table 4 and further including ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP. or their orthologs and allelic or splice variants by at least one cumulus cell associated with said oocyte, and

(iii) evaluating whether said treatment is effective based on the level of expression of said at least one gene by said oocyte-associated cell as compared to the characteristic level of expression of said gene by a cumulus cell associated with a normal or pregnancy oocyte.

32. The method of claim 31, wherein said fertility treatment comprises hormonal therapy.

33. The method of claim 31, wherein the subject is menopausal and the treatment comprises hormone replacement therapy.

34. The method of claim 1, wherein gene expression is detected by real-time polymerase chain reaction (RT-PCR).

35. The method of claim 31, wherein gene expression is detected by RT-PCR.

36. The method of claim 1 or 31 wherein gene expression is detected differentially by indexing differential display reverse transcriptase polymerase chain reaction (DDRT-PCR).

37. The method of claim 36 wherein said gene expression results are obtained using RNA from a cumulus cell.

38. A method of evaluating fertility potential in a subject comprising detecting the expression levels of specific pregnancy signature genes selected from those in Table 4 and ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP, or their orthologs, splice or allelic variants by a cumulus cell associated with an oocyte whose pregnancy potential is being evaluated or another oocyte collected from said subject, comparing said levels of expression to the characteristic levels of expression of said genes by cumulus cells which are associated with an oocyte capable of yielding a viable pregnancy; and determining whether said subject is potentially "pregnancy competent" based on whether said cumulus cell expresses one or more pregnancy signature genes at levels characteristic of pregnancy competent oocytes.

39. The method of claim 38 wherein said at least one gene includes at least 2 genes selected from the group consisting of the ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

40. . The method of claim 38 wherein said at least one gene includes at least 3 genes selected from the group consisting of ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

41. The method of claim 38 wherein said at least one gene includes at least 4 genes selected from the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

42. The method of claim 38 wherein said at least one gene includes at least 5 genes selected from the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

43. The method of claim 38 wherein said at least one gene includes at least 6 genes selected from the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

44. The method of claim 38 wherein said at least one gene includes at least 7 genes selected from the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

45. The method of claim 38 wherein said at least one gene includes 8, 9, 10, 11, 12, 13, or all 14 of the genes selected from the group consisting of ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

46. The method of claim 45 wherein said method monitors differential gene expression by indexing differential display reverse transcriptase polymerase chain reaction (DDRT-PCR).

47. The method of claim 1 wherein gene expression is detected using antibodies that specifically bind to "pregnancy signature" polypeptides.

48. The method of claim 1 wherein the expression of a set of genes from Table 4 and/or ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP on a donor cumulus cell is used to identify an oocyte abnormality precluding pregnancy selected from aneuploidy,

49. A method for selecting a competent oocyte or a competent embryo, comprising a step of measuring in a cumulus cell surrounding said oocyte said embryo the expression level of one or more genes selected from the genes in Table 4 and/or ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSPI, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

50. A method for selecting a competent oocyte comprising a step of measuring the expression level of at least 5 of the genes in Table 4 or at least 5 of the genes consisting of ABCA6, NCAM1, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP.

51. A method for selecting a competent embryo, comprising a step of measuring the expression level of at least 10 genes in a cumulus cell surrounding the embryo, wherein said genes are selected from those in Table 4 or the genes in the group consisting of ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, NDNL2, SLC26A3, and TERF2IP

52. The method according to any of claims 49-51 which further comprise a step consisting of comparing the expression level of the genes in the sample with a control, wherein detecting differential in the expression level of the genes between the sample and the control is indicative whether the oocyte or embryo is competent.

53. The method of any foregoing claim, further comprising fertilizing an oocyte associated with said cumulus cell.

54. The method of claim 53, further comprising transferring said oocyte into the uterus of a female host.

55. The method of claim 54, wherein said female host is of a different species than said oocyte.

56. The method of claim 54, wherein said female host is of the same species as said oocyte.

57. The method of claim 56, wherein said female host is the individual from which said oocyte was obtained.

58. The method of claim 56, wherein said female host is a different individual than the individual from which said oocyte was obtained.

59. The method of any foregoing claim, wherein comparison of gene expression of the at least one cumulus cell associated with an oocyte that is to be tested for pregnancy competency to the characteristic level of expression by a cumulus cell associated with a pregnancy competent oocyte is performed using a method selected from the group consisting of: weighted voting, Bayesian compound covariate, diagonal linear discriminant, nearest centroid, k-nearest neighbors, shrunken centroids, support vector machines, compound covariate, and any combination thereof.

60. The method of claim 59, wherein comparison of gene expression of the at least one cumulus cell associated with an oocyte that is to be tested for pregnancy competency to the characteristic level of expression by a cumulus cell associated with a pregnancy competent oocyte is performed using weighted voting.

61. The method of claim 59 or 60, comprising comparison of gene expression of the genes ABCA6, NCAMI, OLFML3, PTPRA, SDF4, GPR137B, DDIT4, DUSP1, GPR137B, IDUA, KCTD5, DNL2, SLC26A3, and TERF2IP or their orthologs, splice or allelic variants of the at least one cumulus cell associated with an oocyte that is to be tested for pregnancy competency to the characteristic level of expression by a cumulus cell associated with a pregnancy competent oocyte.

62. The method of any foregoing claim, further comprising producing an indicator that indicates whether said oocytes derived from said female donor is potentially capable of yielding a viable pregnancy upon fertilization and transferal into a suitable uterine environment.

63. The method of claim 62, wherein said indicator is provided as a report.

64. The method of claim 62, wherein said indicator is displayed on an electronic display.

65. The method of claim 62, wherein said indicator is provided as an electronic communication.

Patent Citations (7)

Publication numberPriority datePublication dateAssigneeTitle
Family To Family Citations
US7858308B2 *2004-03-292010-12-28Michigan State UniversityIdentification of genes or polypeptides the expression of which correlates to fertility, ovarian function and/or fetal/newborn viability
US7572601B2 *2004-10-152009-08-11Northwestern UniversityCompositions and methods for determining oocyte development potential
WO2007130673A3 *2006-05-052007-12-27Beth Israel HospitalMethods for the diagnosis and treatment of female infertility using molecular markers
US20100021898A1 *2006-09-152010-01-28Universite LavalMammalian oocyte development competency granulosa markers and uses thereof
US8026065B2 *2006-11-022011-09-27Yale UniversityAssessment of oocyte competence
WO2010059742A1 *2008-11-182010-05-27Collabrx, Inc.Individualized cancer treatment
EP2419526B1 *2009-04-172016-03-30INSERM (Institut National de la Santé et de la Recherche Médicale)Methods for selecting oocytes and competent embryos with high potential for pregnancy outcome
* Cited by examiner, † Cited by third party

Non-Patent Citations (1)

Title
See references of EP2499261A4 *
* Cited by examiner, † Cited by third party

Cited By (3)

Publication numberPriority datePublication dateAssigneeTitle
WO2013056258A1 *2011-10-142013-04-18Gema Diagnostics, Inc.Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes
WO2016042015A1 *2014-09-162016-03-24Centre National De La Recherche Scientifique (Cnrs)Method for evaluating developmental competence of an oocyte
Family To Family Citations
US20170362654A1 *2014-12-092017-12-21The Trustees Of Princeton UniversityBiomarkers of oocyte quality
* Cited by examiner, † Cited by third party, ‡ Family to family citation

Similar Documents

PublicationPublication DateTitle
Mirnics et al.2001Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse
Dangond et al.2004Molecular signature of late-stage human ALS revealed by expression profiling of postmortem spinal cord gray matter
Romero et al.2006The use of high-dimensional biology (genomics, transcriptomics, proteomics, and metabolomics) to understand the preterm parturition syndrome
US20070105136A12007-05-10Methods for identifying, diagnosing, and predicting survival of lymphomas
US20090258002A12009-10-15Biomarkers for Tissue Status
Gasca et al.2007Identifying new human oocyte marker genes: a microarray approach

[Mullen et al.](#)2006Understanding meat quality through the application of genomic and proteomic approaches

[US20020029113A1](#)2002-03-07Method and system for predicting splice variant from DNA chip expression data

[Ponsuksili et al.](#)2008Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle

[Evian Annual Reproduction (EVAR) Workshop Group 2010 Fauser BCJM Diedrich K. Bouchard P. Domínguez F. Matzuk M. Franks S. Hamamah S. Simón C. Devroey P. Ezcurra D. Howles CM](#)2011Contemporary genetic technologies and female reproduction

[Turan et al.](#)2012DNA methylation differences at growth related genes correlate with birth weight: a molecular signature linked to developmental origins of adult disease?

[US20060292572A1](#)2006-12-28Cell-type-specific patterns of gene expression

[Tuggle et al.](#)2007Advances in swine transcriptomics

[WO2004097051A2](#)2004-11-11Methods for diagnosing aml and mds differential gene expression

[US7767395B2](#)2010-08-03Diagnosis of sepsis

[Rosenbluth et al.](#)2014Human embryos secrete microRNAs into culture media—a potential biomarker for implantation

[Yang et al.](#)2014Selection of competent blastocysts for transfer by combining time-lapse monitoring and array CGH testing for patients undergoing preimplantation genetic screening: a prospective study with sibling oocytes

[WO2011051283A1](#)2011-05-05Means and methods for non-invasive diagnosis of chromosomal aneuploidy

[Ouandaogo et al.](#)2011Human cumulus cells molecular signature in relation to oocyte nuclear maturity stage

[Clark-Langone et al.](#)2007Biomarker discovery for colon cancer using a 761 gene RT-PCR assay

[US20100036192A1](#)2010-02-11Methods and systems for assessment of clinical infertility

[Cui et al.](#)2014Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing

[US20120071343A1](#)2012-03-22Biomarkers for differentiating melanoma from benign nevus in the skin

[Maloyan et al.](#)2013Identification and comparative analyses of myocardial miRNAs involved in the fetal response to maternal obesity

[WO2010118991A1](#)2010-10-21Methods for selecting oocytes and competent embryos with high potential for pregnancy outcome

Priority And Related Applications

Priority Applications (10)

Application	Priority date	Filing date	Title
US259783309	2009-11-10	2009-11-10	*US Provisional Application*
US61/259,783	2009-11-10		
US360556160	2010-07-01	2010-07-01	*US Provisional Application*
US61/360,556	2010-07-01		
US387286160	2010-09-28	2010-09-28	*US Provisional Application*
US387313160	2010-09-28	2010-09-28	*US Provisional Application*
US61/387,313	2010-09-28		
US61/387,286	2010-09-28		
US388296160	2010-09-30	2010-09-30	*US Provisional Application*
US61/388,296	2010-09-30		

Applications Claiming Priority (4)

Application	Filing date	Title

EP20100830666 2010-11-10 Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes
CA 2780430 2010-11-10 Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes
JP2012538952A 2010-11-10 Assays for the identification cumulus cells oocytes with pregnant competent using genes and their expressed differentially
US13509111 2010-11-10 Genes differentially expressed by cumulus cells and assays using same to identify pregnancy competent oocytes

Legal Events

DateCodeTitleDescription
2012-05-09 ENP Entry into the national phase in:
Ref document number: 2780430

Country of ref document: CA

2012-05-10 ENP Entry into the national phase in:
Ref document number: 2012538952

Country of ref document: JP

Kind code of ref document: A

2012-05-10 NENP Non-entry into the national phase in:
Ref country code: DE

2012-05-10 WWE Wipo information: entry into national phase
Ref document number: 2012538952

Country of ref document: JP

2012-11-16 WWE Wipo information: entry into national phase
Ref document number: 13509111

Country of ref document: US

Data provided by IFI CLAIMS Patent Services