Spring 2024 – Systems Biology of Reproduction
Discussion Outline – Ovary Systems Biology
Michael K. Skinner – Biol 475/575
Zoom/CUE 418, 10:35-11:50 am, Tuesday & Thursday
February 29, 2024
Week 8

Ovary Systems Biology

Primary Papers:

1. Pla, et al. (2021) Hum Reprod. 18;36(3):756-770
2. Sagvekar, et al. (2019) Clinical Epigenetics 11:61
3. Nilsson, et al. (2010) PLoS ONE 7:e11637

Discussion

Student 8: Reference 1 above

- What approach and technology was used?
- What gene categories and networks were identified?
- What oocyte maturation and folliculogenesis insights were identified?

Student 9: Reference 2 above

- What are the technology used and objectives?
- What epigenetic regulation and gene network were identified?
- What insights are provided into the development of polycystic ovarian disease?

Student 10: Reference 3 above

- What is the experimental and systems approach?
- What new insights provided on primordial follicle development?
- What gene signaling networks were identified for primordial follicle development?

Human Reproduction, Vol.36, No.3, pp. 756–770, 2021
Advance Access Publication on December 13, 2020 doi:10.1093/humrep/deaa335

human reproduction

ORIGINAL ARTICLE *Reproductive biology*

Proteome of fluid from human ovarian small antral follicles reveals insights in folliculogenesis and oocyte maturation

Indira Pla[1,2,†], Aniel Sanchez[1,2,*,†], Susanne Elisabeth Pors[3,*,†], Krzysztof Pawlowski[2,4], Roger Appelqvist[2], K. Barbara Sahlin[1,2], Liv La Cour Poulsen[5], György Marko-Varga[2,6], Claus Yding Andersen[3], and Johan Malm[1,2]

[1]Section for Clinical Chemistry, Department of Translational Medicine, Lund University, Skåne University Hospital Malmö, 205 02 Malmö, Sweden [2]Clinical Protein Science & Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, BMC D13, 221 84 Lund, Sweden [3]Laboratory of Reproductive Biology, The Juliane Marie Centre for Women, Children and Reproduction, University Hospital of Copenhagen, 2100 Copenhagen, Denmark [4]Department of Experimental Design and Bioinformatics, Faculty of Agriculture and Biology, Warsaw University of Life Sciences SGGW, Warszawa 02-787, Poland [5]Fertility Clinic, Department of Gynaecology and Obstetrics, Zealand University Hospital, Lykkebækvej 14, 4600 Køge, Denmark [6]First Department of Surgery, Tokyo Medical University, Shinjiku-ku, Japan

*Correspondence address: Clinical Protein Science & Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University Klinikgatan 32, D13., BMC D13, 221 84 Lund, Sweden. E-mail: aniel.sanchez@med.lu.se (A.S.); Laboratory of Reproductive Biology, University Hospital of Copenhagen, Rigshospitalet, Section 5701, Henrik Harpestrengsvej 6A, DK-2100 Copenhagen, Denmark. E-mail: susanne.elisabeth.pors@regionh.dk (S.E.P.)

Submitted on March 17, 2020; resubmitted on November 3, 2020; editorial decision on November 10, 2020

STUDY QUESTION: Is it possible to identify by mass spectrometry a wider range of proteins and key proteins involved in folliculogenesis and oocyte growth and development by studying follicular fluid (FF) from human small antral follicles (hSAF)?

SUMMARY ANSWER: The largest number of proteins currently reported in human FF was identified in this study analysing hSAF where several proteins showed a strong relationship with follicular developmental processes.

WHAT IS KNOWN ALREADY: Protein composition of human ovarian FF constitutes the microenvironment for oocyte development. Previous proteomics studies have analysed fluids from pre-ovulatory follicles, where large numbers of plasma constituents are transferred through the follicular basal membrane. This attenuates the detection of low abundant proteins, however, the basal membrane of small antral follicles is less permeable, making it possible to detect a large number of proteins, and thereby offering further insights in folliculogenesis.

STUDY DESIGN, SIZE, DURATION: Proteins in FF from unstimulated hSAF (size 6.1 ± 0.4 mm) were characterised by mass spectrometry, supported by high-throughput and targeted proteomics and bioinformatics. The FF protein profiles from hSAF containing oocytes, capable or not of maturing to metaphase II of the second meiotic division during an IVM (n = 13, from 6 women), were also analysed.

PARTICIPANTS/MATERIALS, SETTING, METHODS: We collected FF from hSAF of ovaries that had been surgically removed from 31 women (~28.5 years old) undergoing unilateral ovariectomy for fertility preservation.

MAIN RESULTS AND THE ROLE OF CHANCE: In total, 2461 proteins were identified, of which 1108 identified for the first time in FF. Of the identified proteins, 24 were related to follicular regulatory processes. A total of 35 and 65 proteins were down- and up-regulated, respectively, in fluid from hSAF surrounding oocytes capable of maturing (to MII). We found that changes at the protein level occur already in FF from small antral follicles related to subsequent oocyte maturation.

LIMITATIONS, REASONS FOR CAUTION: A possible limitation of our study is the uncertainty of the proportion of the sampled follicles that are undergoing atresia. Although the FF samples were carefully aspirated and processed to remove possible contaminants, we

[†]The authors consider that the first three authors should be regarded as joint first authors.

cannot ensure the absence of some proteins derived from cellular lysis provoked by technical reasons.

WIDER IMPLICATIONS OF THE FINDINGS: This study is, to our knowledge, the first proteomics characterisation of FF from hSAF obtained from women in their natural menstrual cycle. We demonstrated that the analysis by mass spectrometry of FF from hSAF allows the identification of a greater number of proteins compared to the results obtained from previous analyses of larger follicles. Significant differences found at the protein level in hSAF fluid could predict the ability of the enclosed oocyte to sustain meiotic resumption. If this can be confirmed in further studies, it demonstrates that the viability of the oocyte is determined early on in follicular development and this may open up new pathways for augmenting or attenuating subsequent oocyte viability in the pre-ovulatory follicle ready to undergo ovulation.

STUDY FUNDING/COMPETING INTEREST(S): The authors thank the financial support from ReproUnion, which is funded by the Interreg V EU programme. No conflict of interest was reported by the authors.

TRIAL REGISTRATION NUMBER: N/A

Key words: follicular fluid / small antral follicles / proteomics / oocyte maturation / mass spectrometry

Introduction

A woman's fertility is based on the pool of resting follicles in the ovaries, i.e. the ovarian reserve. This reserve ensures the generation of menstrual cycles and release of fertilisable oocytes that may result in new offspring. Follicular development is initiated by activation of resting primordial follicles and is completed with ovulation of fully mature oocytes approximately half a year later (Gougeon, 2010). During this lengthy period of follicular development, the diameter of the follicle increases from 45 µm to ~20 mm. This process involves several developmental stages, including the formation of the follicular fluid (FF)-filled antrum that begins to form when human ovarian follicles reach a diameter around 250 µm (Rodgers and Irving-Rodgers, 2010). The FF is comprised of secretions from the oocyte and granulosa cells (GC) within the follicles and from theca cells (TC) surrounding the follicle as well as transudates from circulation filtered through the basal membrane. The basal membrane acts as a molecular filter, which means that proteins with a relatively high molecular weight can only enter the FF to a limited extent (Rodgers and Irving-Rodgers, 2010; Siu and Cheng, 2012). This is the reason why FF contains low concentrations of the high molecular weight protein fibrinogen and, therefore, does not coagulate. Additionally, the basal membrane has been described to be both charge and size selective in mouse ovaries (Hess *et al.*, 1998; Siu and Cheng, 2012). The composition of FF is highly variable and is associated with the developmental stage of follicles. In particular, FF reflects GC activity, which is regulated by gonadotropins, steroids, peptide hormones and growth factors. Anti-Müllerian hormone (AMH), part of the transforming growth factor β (TGF-β) superfamily, is present at very high concentrations in small antral follicles with a peak in follicular content around a diameter of 8 mm (Jeppesen *et al.*, 2013). Conversely, sex-steroids, such as estradiol and progesterone, accumulate at very high concentrations in the pre-ovulatory follicles, in orders of magnitude higher than in small antral follicles (Jeppesen *et al.*, 2013).

The FF constitutes the microenvironment in which oocytes develop and as a consequence, the protein composition of FF has attracted considerable interest and many proteomics studies have been conducted (Spitzer *et al.*, 1996; Anahory *et al.*, 2002; Lee *et al.*, 2005; Schweigert *et al.*, 2006; Angelucci *et al.*, 2006; Hanrieder *et al.*, 2008; Estes *et al.*, 2009; Gougeon, 2010; Jarkovska *et al.*, 2010; Twigt *et al.*, 2012, 2015; Ambekar *et al.*, 2013, 2015; Bianchi *et al.*, 2013; Jeppesen *et al.*, 2013; Zamah *et al.*, 2015; Oh *et al.*, 2017). All of these studies have focused on FF collected just prior to ovulation, in connection with assisted reproduction techniques. The fluid from pre-ovulatory follicles has a high dynamic range in terms of protein concentrations, which reduces the possibility of detecting low abundant proteins. At this follicular stage, the FF contains a high number of plasma constituents transferred through the follicular basal membrane during follicular expansion. For this reason, the number of proteins identified in previous studies has been in the range of several hundred. By using FF from human small antral follicles (hSAF), this limitation would be avoided. Furthermore, the study of the FF proteome by mass spectrometry (MS) allows the simultaneous analysis of hundreds of proteins and the current bioinformatics advances can provide, even from early follicular stages, the functional network of the proteins and their role in the follicular development. Hence, the larger the number of identified proteins, the greater the knowledge acquired. In addition, an increased focus on the earlier stages of follicles (i.e. small antral follicles), has shown that these could be a potential source of immature oocytes for women with a low ovarian reserve (Kristensen *et al.*, 2017). Details on the composition of FF from small antral follicles will provide essential knowledge about factors impacting the developmental capacity of the immature oocyte. By contributing to a better understanding of basic follicular processes, central proteins present in small antral follicles could be valuable in creating new physiological and effective methods for maturing oocytes *in vitro* and advancing culture conditions for human follicles, ultimately advancing fertility treatment by augmenting the number and quality of oocytes available for treatment.

The present study aimed to create a detailed fingerprint of proteins present in FF from hSAF to identify from the early follicular stage, candidate proteins that support follicular growth and development.

Materials and methods

An overview of the study workflow is presented in Supplementary Fig. S1.

Reagents and solutions

Unless otherwise specified, all chemical reagents were purchased from Sigma Aldrich (St. Louis, MO, USA). Modified porcine trypsin was

obtained from Promega (Madison, WI, USA) and the water was from a Milli-Q Ultrapure Water System (Millipore, Billerica, MA, USA). Liquid chromatography-mass spectrometry (LC-MS) grade water and organic solvents were supplied by Merck (Darmstadt, Germany). Prior to use, all LC-MS solutions were degassed by sonication. Synthetic peptides were provided by the University of Victoria-Genome BC Proteomics Centre, British Columbia, Canada.

Strong cation exchange (SCX) chromatography microspin columns (MA SEM HIL-SCX, 10–100 mg capacity) and silica C18 ultra-microspin columns (SUM SS18V, 3–30 mg capacity) were purchased from the Nest Group Inc. (South Borough, MA, USA). MARS 7 immunoaffinity spin columns were purchased from Agilent (Agilent Technologies, Inc., CA, USA) and were used to deplete the top seven most abundant proteins, i.e. albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin and fibrinogen.

FF samples acquisition from small antral follicles

Samples of FF from hSAF were collected from ovaries that had been surgically removed from 25 women (25.6 ± 6.1 years old (mean ± SD)) undergoing unilateral ovariectomy for fertility preservation at the Laboratory of Reproductive Biology, Rigshospitalet, Denmark. Cryopreservation of the ovarian cortex was offered to women facing a potentially gonadotoxic treatment and thereby at risk of becoming sterile (Andersen *et al.*, 2008; Schmidt *et al.*, 2011). Cryopreservation of the ovarian cortical tissue provides these women with the option of later transplantation of the frozen-thawed tissue if they become menopausal due to the gonadotoxic treatment. Only patients with diseases unrelated to the ovary were included, and in all cases, the ovary had a macroscopically normal appearance. The diagnosis of the women included breast cancer (n = 11), Hodgkin's lymphoma (n = 3), sarcoma (n = 4), lymphoma (n = 4), brain cancer (n = 1), cervical cancer (n = 1) and systemic lupus erythematous (n = 1). Even though the women had a concurrent cancer diagnosis, they were eligible for fertility cryopreservation after an evaluation of reproductive parameters and ovarian reserve (AMH levels and antral follicle count); therefore, we assumed that the women and their ovaries were reproductively normal. Small antral follicles exposed on the surface of the ovary or visible during the isolation of ovarian cortex were aspirated with a 1-ml syringe fitted with a 26-gauge needle (Becton Dickinson, Brøndby, Denmark). From each ovary, FF from one small antral follicle was collected and the mean size of the follicles was 6.1 ± 0.4 mm (mean ± SD). The diameter was calculated based on the total volume of fluid drawn from the follicle using the calculation of spherical shape ($V = 4/3 * \pi * r^3$). Aspiration of FF had no effect on the fertility preservation procedure. The FF samples had no visible blood contamination but were immediately centrifuged to remove debris and cells. The FF were collected at random times during the menstrual cycle as the dynamics of the hSAF appears to be similar throughout the menstrual cycle (Mcnatty *et al.*, 1983; Westergaard, 1985; Kuang *et al.*, 2014). From 15 of the follicles, 50-µl FF was pooled whilst the remaining 10 follicles were analysed individually.

Furthermore, six women were selected (women with breast cancer (n = 4, sarcoma (n = 1) and multiple sclerosis (n = 1) and an average age of 31.4 ± 1.8 (mean ± SD)) and from each one, fluid from two or three hSAF (size 6 ± 1.5 mm (mean ± SD)) was obtained. Each follicle contained an oocyte that underwent IVM. From each woman, at least one follicle contained an oocyte that matured to metaphase II of the second meiotic division after IVM and another that contained an oocyte unable to mature. Oocytes were matured during a 48 h culture period using the MediCult IVM medium (Origio A/S, Denmark) supplemented with 75 mIU/ml rFSH (Puregon, MSD, the Netherlands), 100 mIU/ml rLH (Luveris, Serono, Germany) and 10 mg/ml human serum albumin. Oocytes were cultured individually in 25-µl drops and separate data for each oocyte was obtained. At the end of the culture period, all oocytes were denudated, and the developmental stage was classified as either germinal vesicle, metaphase I (MI) or metaphase II (MII).

The study was approved by the ethics committee of the municipalities of Copenhagen and Frederiksberg (H-2-2011-044). Informed consent was obtained from all participants.

Sample preparation to MS

A schematic representation of the general strategy followed for the analysis of the samples is shown in Supplementary Fig. S2. For all experiments, the quantitation of total proteins was performed using the bicinchoninic acid (BCA) assay.

Preparation of samples without depletion

Samples were dissolved in 1.6% sodium deoxycholate (SDC) in 50 mM NH_4HCO_3. The disulphide bonds were reduced by adding dithiothreitol (DTT) to a final concentration of 10 mM and incubated at 37°C for 1 h. The free thiol groups were alkylated by adding iodoacetamide (IAA) to a final concentration of 25 mM and incubated for 30 min at room temperature in the dark. The SDC was diluted to 0.5% before digestion with trypsin at an enzyme-to-substrate ratio of 1:100 (w/w) for 16 h at 37°C. The SDC was precipitated by adding 20% formic acid to a polypropylene filter plate with a hydrophilic polyvinylidene difluoride (PVDF) membrane (mean pore size 0.45 µm, Porvair Filtration Group, Fareham, UK).

Sample depletion

To deplete the top 7 or 14 (Supplementary Fig. S2) most abundant proteins, 10 or 25 µl of FF were used per sample, respectively. The depletions, with MARS 7 (spin columns) or MARS 14 column (Human-14 (4.6 × 100 mm) coupled to an 1260 Infinity LC System) both from Agilent technologies, were carried out according to instructions supplied by the manufacturer. After depletion, the buffer was exchanged to SDC 1.6%, 50 mM of Ambic using Amicon Ultra Centrifugal filter (0.5 ml – 10 kDa, Millipore, Tullagreen, Ireland).

Samples were reduced and alkylated as done with non-depleted samples and it took place in the Amicon filter. The buffer was exchanged to Ambic 50 mM and the samples were re-suspended in 100 µl of 50 mM AmBic (30 µg of proteins after BCA quantification). Then samples were digested with trypsin at an enzyme-to-substrate mass ratio of 1/30 for 16 h at 37 °C. The remaining SDC precipitated by adding 20% of formic acid prior to filtering the samples through a polypropylene filter plate with hydrophilic PVDF membrane (mean pore size 0.45 µm, Porvair Filtration Group, Fareham, United Kingdom).

Fractionation of peptides by strong cation exchange chromatography

For SCX chromatography, three aliquots (10 µl each) from the pool were depleted and separately digested. The digests were combined before fractionation. Mainly, a step-wise gradient of potassium chloride: 20, 40, 60, 100, 500 mM and 1 M KCl in 10 mM potassium phosphate, pH 2.8 (or 0.01% phosphoric acid) containing 20% acetonitrile was used to fractionate the peptides.

According to the instructions from the manufacturer, salt was removed from the samples by silica C18 ultra-microspin columns (SUM SS18V, 3–30 mg capacity, The Nest Group Inc., South Borough, MA, USA). After elution with 50% acetonitrile/0.1% TFA, the fractions were dried in a centrifugal evaporator and re-suspended in 0.1% formic acid prior to analysis by LC-MS/MS.

Fractionation of peptides by basic reversed-phase chromatography

Three aliquots (10 µl each) from the pooled FF were depleted and digested separately. We performed a Filter Aided Sample Preparation for protein digestion (Wiśniewski *et al.*, 2009). The digestions were combined prior to drying the sample. The sample was reconstituted in 100 µl ammonium formate/20% acetonitrile, pH 10, loaded on the column and fractionated using an Agilent 1100 series HPLC instrument by reversed-phase chromatography (RP-HPLC) at a flow rate of 50 µl/min. The mobile phase consisted of 20 mM ammonium formate, pH 10 (buffer A) and 20 mM ammonium formate, 80% acetonitrile, pH 10 (buffer B). The peptides were separated using the following gradient: 15 min isocratic hold at 1% solvent B, 1–20% solvent B in 1 min; 20–60% solvent B in 44 min; 60–80% solvent B in 5 min; finally, 15 min isocratic flow at 80% solvent B. Using 96 × 2 ml well plates, fractions were collected every 2 min for a total of 20 fractions. Finally, the fractions were recombined by pooling, and 12 were analysed by LC-MS.

MS acquisition methods

The samples were analysed by data-dependent acquisition (DDA), parallel reaction monitoring (PRM) and multiple reaction monitoring (MRM). Three different mass spectrometers from Thermo Fisher Scientific (San José, CA, USA) were used for the analyses: The Q-Exactive Plus (DDA and PRM, see Supplementary Fig. S2a and b), the Q-Exactive HF-X (DDA) and TSQ Quantiva (MRM) equipped with an easy-spray NG ion source. The spectrometers were connected to an easy-nLC 1000 pump (Thermo Scientific, San José, CA, USA).

In the Q-Exactive Plus mass spectrometer, a top 10 method was used for DDA. Full MS1 spectra were acquired in the Orbitrap mass analyser from m/z 400–1600 at a resolution of 70 000 (at m/z 200), a target automatic gain control (AGC) value of 1e6 and a maximum injection time (IT) of 100 ms over a 60-min HPLC gradient. The 10 most intense peaks with charge state ≥2 were fragmented in the HCD collision cell with a normalised collision energy of 26%. Tandem MS2 spectra were acquired in the Orbitrap mass analyser at a resolution of 35 000 (at m/z 200), a target AGC value of 5e4 and a maximum IT of 100 ms. The underfill ratio was set to 10% and dynamic exclusion was 45 s.

For the Q-Exactive HF-X mass spectrometer, the full MS scans were set with an acquisition range of m/z 375-1500, resolution of 120 000 (at m/z 200), target AGC value as 3 × 106, maximum injection time of 100 ms and normalised collision energy of 28. The top 20 precursors were selected for fragmentation. For the MS2 acquisition, we used a resolution of 15 000 (at m/z 200), target AGC value of 1 × 106, maximum injection time of 50 ms, isolation window of 1.2 m/z and fixed first mass at 110 m/z. Peptide elution was performed with a gradient of 2% of 80% acetonitrile/0.1% formic acid (Solvent B) and 98% of 0.1% formic acid (Solvent A) with a flow of 0.300 µl/min during the first 3 min. This step was followed by an increase in the percentage of solvent B to 25% in 112 min, to 32% in 10 min, and to 45% in 7 min. After 132 min of the gradient, the solvent B percentage was increased to 90% in 8 min and kept constant for 5 min. Finally, the solvent B content was reduced to 2% in 1 min and kept constant for 14 min.

For targeted proteomics (PRM and MRM), we built a spectral library using the MS/MS spectra from synthetic peptides. In addition, we measured a 'reference sample' by adding the synthetic peptides to a pool of all samples.

For PRM, 1 µg in 2 µl was loaded per sample onto the column. The MS2 resolution was 70 000 with an AGC value of 5e5 and a maximum IT of 200 ms. The normalised collision energy was 26%. Peptides were separated on an easy-spray column (25 cm × 75 µm ID, PepMap C18 2 µm, 100 Å) with the flow rate of 300 nl/min and the column temperature at 35°C. Solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile) were used to create a non-linear gradient to elute the peptides (60 min).

In the MRM method, a scheduled mode was used with 5-min detection windows. Peptides (1 µg) were loaded onto an Acclaim PepMap 100 pre-column (100 µm × 2 cm, Thermo Scientific, San José, CA, USA). Peptides were separated on an easy-spray column (15 cm × 75 µm ID, PepMap C18 3 µm, 100 Å) with the flow rate set to 300 nl/min and the column temperature at 35°C. Solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile) were used to create a non-linear gradient to elute the peptides (60 min). Selected reaction monitoring (SRM) transitions were acquired in Q1 and Q3 operated at unit resolution (0.7 fwhm); the collision gas pressure in Q2 was set to 1.5 mTorr. The cycle time was 2 s, and calibrated RF and S-lens values were used. At least three transitions per precursor were monitored.

MS data analysis

Raw files were analysed with Proteome Discoverer v2.2 or 2.4 (Thermo Scientific, San José, CA, USA). To identify the peptides, the MS data were searched against the UniProtKB human database (Released 20180207, 42213 sequences including isoforms). For the analysis of the data generated from FF that contained oocytes capable of maturing or not, we combined an FF spectral database built from the LC-MS/MS analysis of a top 14 depleted pool (hSAF MS1 spectral library) and the MSPepSearch node plus SEQUEST HT. We used the human spectral library 'ProteomeTools_HCD28_PD' and UniprotKB human database (Date: 28 January 2020), respectively. The search was performed with the following parameters: carbamidomethylation of cysteine residues and oxidation of methionine residues as static and dynamic modifications, respectively. Precursor and fragment ion tolerances were 10 ppm and 0.02 Da, respectively. Up to one missed cleavage site for tryptic peptides was allowed. The filters applied were

high (false discovery rate (FDR) < 1%) and medium (FDR < 5%) confidence at peptide and protein level, respectively. The peptide/protein quantification was based on the MS peptide signals (label-free quantification). For label-free quantification, the 'Minora Feature Detector' node was included in the processing workflow, and the nodes 'Precursor Ions Quantifier' and 'Feature Mapper' were included in the consensus workflow.

Comparison of the identified proteins with previous studies

The total number of proteins identified in the current study was compared to 30 previous proteomics studies performed in human FF. Out of 30, 20 studies were found after performing a PubMed and Embase searching using the term 'human follicular fluid proteome'. Considering that one of the previous papers (Bianchi *et al.*, 2016) summarised proteins identified up to 2014 (including 10 proteomics studies), we looked for proteomic studies published between 2013 and 2020. Only proteins reported as 'reviewed' in the UniProtKB/Swiss-Prot protein sequence database were selected. The studies were: Severino *et al.* (2013), Bayasula *et al.* (2013), Regiani *et al.* (2015), Twigt *et al.* (2015), Zamah *et al.* (2015), Ambekar *et al.* (2015), Chen *et al.* (2016), Bianchi *et al.* (2016), Shen *et al.* (2017), Oh *et al.* (2017), Lim *et al.* (2017), Lewandowska *et al.* (2017), Lewandowska *et al.* (2019), Poulsen *et al.* (2019), Li *et al.* (2019) , Zhang *et al.* (2019), Domingues *et al.* (2019), Liu *et al.* (2020a,b) and Zakerkish *et al.* (2020).

Furthermore, the list of identified proteins was compared with a list of proteins from oocytes reported by Virant-Klun *et al.* (2016) and with transcripts from GC published by Kõks *et al.* (2010).

Bioinformatics and statistical analyses

Gene Ontology (GO) and functional enrichment analyses were performed utilising the bioinformatics tool FunRich (Pathan *et al.*, 2015). The bioinformatics web tool DAVID (https://david.ncifcrf.gov/) (Huang *et al.*, 2009) was employed to perform a functional annotation clustering using as input 226 proteins from the high abundant proteome (163 only identified in hSAF, 28 only identified in pre-ovulatory follicles and 35 non-plasma proteins identified in both hSAF and large follicles). This tool was also employed to perform functional annotation clustering of the top 100 dysregulated proteins in FF that contained oocyte capable of reaching MII. Protein relationship networks were generated and analysed using the Ingenuity Pathway Analysis (IPA) software (QIAGEN, Germany).

To perform proteomics quantitative analyses, the intensities of the protein were normalised by $\log_2$ transformation and then standardised by subtracting the median of the sample. Statistical analyses were performed in RStudio software (R Core Team, 2016; RStudio Team, 2016). Coefficient of correlation between MDK, vimentin (VIM) and other proteins was determined by Pearson correlation test using the $\log_2$ intensities values of the protein quantified across 10 individual samples. The *P*-values were adjusted to control the FDR produced by multiple testing. Correlations with $-0.7 \geq r \geq 0.7$ and adjusted *P*-value <0.10 were considered significant. Proteins that significantly correlated with MDK/VIM were subjected to a biological pathway enrichment analysis in FunRich (background: FunRich database) and pathways with a significant enrichment score (BH method: adj. *P*-value <0.05) were selected. A comparative enrichment analysis based on 'cellular components' annotations was performed between positively and negatively correlated protein and a Q-value (Storey-Tibshirani method) <0.05 was considered significant.

To assess whether there were differences at protein level between FF that surrounded oocytes that matured or remained immature, a sparse partial least squares discriminant analysis (sPLS-DA) (Chung and Keles, 2010) was performed using 'mixOmics' R package. This is a multivariate analysis that classifies the samples by performing a multivariate regression using the protein expression matrix (749 proteins quantified in all samples) as predictors and the sampling origin (FF surrounding oocytes mature or immature) as the response. To select the top 100 most informative predictors (e.g. proteins) for discriminating samples, an LASSO penalisation was applied. With the top 100 proteins, a hierarchical clustering plus heatmap was performed using 'ComplexHeatmap' R library. In addition to the multivariable analysis, a Student *t*-test (two-tails) followed by FDR correction was performed to determine differentially expressed proteins. Proteins with an adjusted *P*-value <0.05 were considered significant. Since FF samples surrounding the immature oocyte and FF samples surrounding the MII oocyte originated in the same woman, the analyses (the multivariable and univariate) were performed considering the paired nature of the samples. This was achieved by subtracting from the protein intensity of a given sample the mean of the two or three samples belonging to the woman from which they were extracted. In this analysis, the expression values of VIM were determined by MRM. Differences between groups in term of follicular size were assessed by performing a paired Student *t*-test (two-tails). Secreted proteins dysregulated in FF that contained oocyte capable of reaching MII were correlated with the remaining proteins using a Pearson correlation test and adjusted *P*-values <0.05 were considered significant.

Selection of proteins potentially more concentrated and accessible by MS in FF from hSAF

The selection of proteins potentially more concentrated and accessible by MS in FF from hSAF is presented in the workflow in Supplementary Fig. S3. Proteins identified in the analysis of pool were filtered by proteins identified in the non-depleted samples and those also identified by another method, e.g. SCX or basic RP. Then 368 resulting proteins were compared against the proteins identified in the two largest previous data sets that used similar MS strategies (Zamah *et al.*, 2015; Oh *et al.*, 2017). From the resulting 35 proteins unique to our study, those known as 'classical plasma proteins' (Anderson and Anderson, 2002) (mainly immunoglobulins and complement components) were excluded. Then the resulting proteins were compared with the list of proteins identified in FF by MS up to 2020.

Results

General protein characterisation

In this study, a total of 2461 proteins were identified in FF from hSAF (Supplementary Table SI). A pool of 15 FF samples was first evaluated

to deepen the protein identification and subsequently, 23 separated samples (10 from different women and 13 from six other women) were analysed. Supplementary Tables SII, SIII and SIV contain the list of proteins identified in each case. Detailed information on the protein identification is described in Supplementary File S1 (Supplementary Figs S4, S5 and S6). The list of identified proteins was compared with previous human FF proteomics studies conducted up to 2020 (Supplementary Table SV). A total of 1108 proteins were detected for the first time in our study (Venn diagram: Fig. 1).

The entire set of proteins identified in this study was analysed to determine GO annotations. Figure 2 shows the protein distribution according to biological process, molecular function, cellular compartment and protein class of the FF proteome in hSAF. The generic terms: metabolic process and biological regulation were the most represented biological presses with 26% and 16%, respectively, whilst almost half of the identified proteins were related to catalytic (36%) and binding (43%) molecular functions. The classification based on cellular compartment indicated that 38% of the proteins were mainly nucleus and cytosolic (cell part), whilst 19% were extracellular.

To detect proteins possibly secreted from GC or oocyte, we compared our results with data obtained by transcriptomics in human GC and by proteomics in human oocytes (Kõks *et al.*, 2010; Virant-Klun *et al.*, 2016). In those studies, samples were obtained from pre-ovulatory follicles from women undergoing IVF. A total of 1940 proteins (94%) were found at the transcript level in GC whilst 793 proteins (39%) were identified in oocytes (Supplementary Table SI).

Comparing FF proteome from hSAF and large follicles: high abundance proteome

To our knowledge, this is the first proteomic study of FF from hSAF (<8 mm). To detect proteins possibly more accessible by MS in hSAF, a comparison in terms of the most abundant proteins identified in FF from hSAF and large follicles was carried out. The 'high abundance proteome' (non-depleted samples) identified in our study (413 proteins, Supplementary Table SVI) was compared to the 'high abundance proteome' (400 proteins) of a recent proteomics study performed by our group in large, pre-ovulatory follicles (>14 mm) (Poulsen *et al.*, 2019). Out of the 413 proteins from hSAF, 231 (56%) were also identified in FF from large pre-ovulatory follicles. From these, 196 are commonly detected in plasma (as compared with the human plasma proteome database) and may, therefore, represent plasma-filtrated proteins. According to the GO analysis, the remaining 35 proteins are

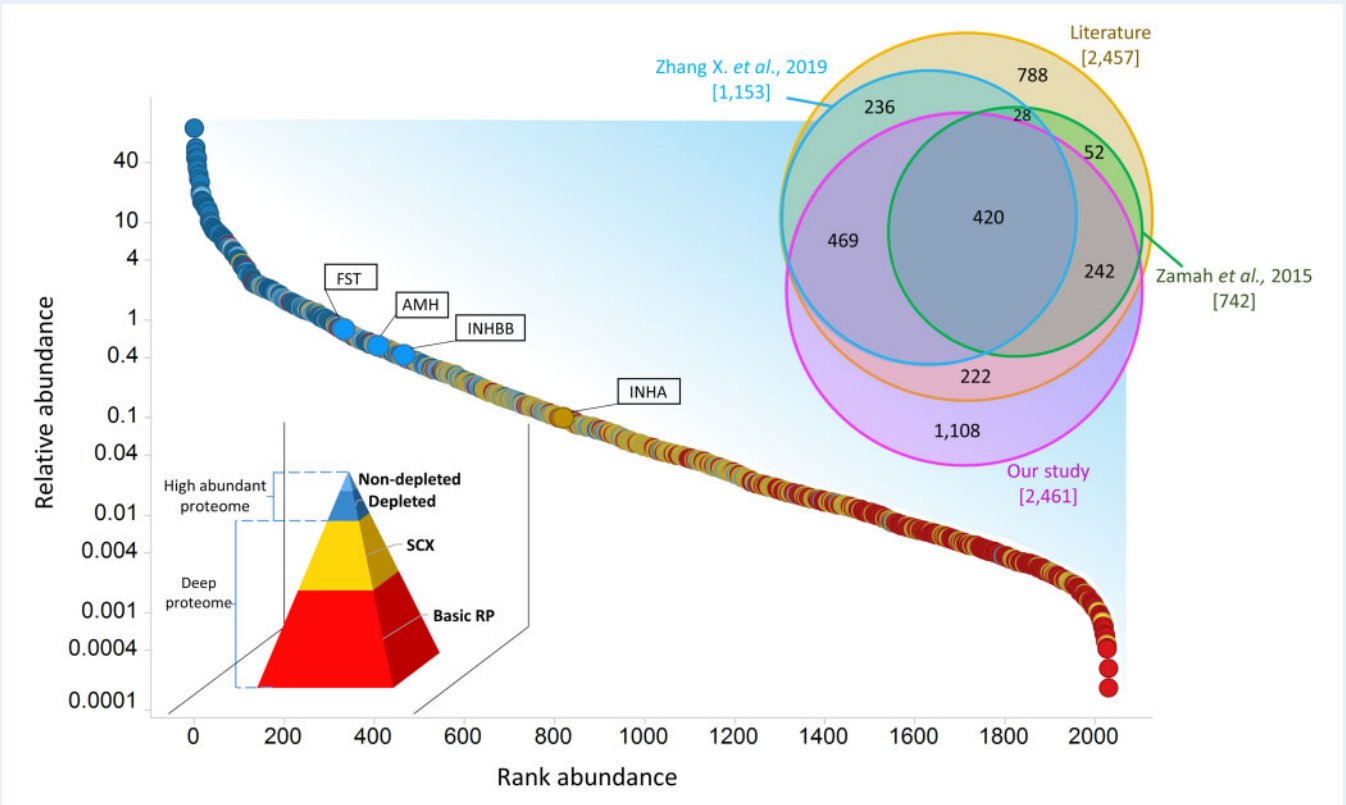


Figure 1. Dynamic rank of proteins identified in our study and Venn diagram of total proteins identified in human follicular fluid (FF) studies. In the Venn diagram, 'literature' (yellow) denotes proteins identified in previous FF studies (up to 2020). The circles in light blue and green colour, represent the two FF studies that currently have identified the highest number of proteins (Zamah *et al.*, 2015; Zhang et al., 2019). In total, 3565 proteins have been identified in FF samples by proteomics. The colours in the pyramid correspond to the colours in the dynamic rank plot. When a more complex method is applied (SCX and basic RP), the results yield a greater number of identifications in total and, in particular, increased identifications of low abundance proteins. Proteins highlighted are known to be relevant in the reproductive system.

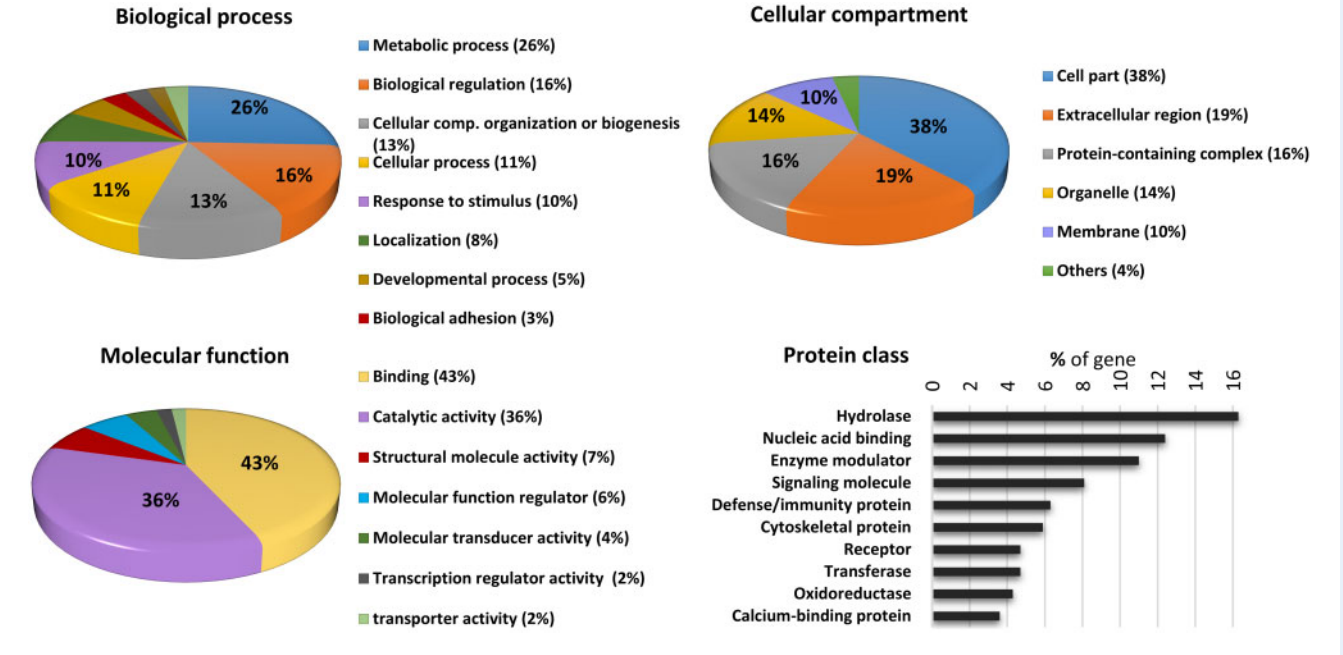


Figure 2. Gene ontology analysis for all follicular fluid (FF) proteins. Distribution of all proteins identified in this study according to three different GO categories and protein class.

distributed across the cell and almost half were secreted proteins (Fig. 3a).

Proteins only identified in hSAF and completely missing in FF from large follicles (163 proteins referred as 'on in hSAF' in Fig. 3a (red colour)) were mostly intracellular proteins (e.g. cytoplasmic and/or nuclear) whilst the extracellular proteins were distributed indistinctly in FF from small follicles and large follicles. However, proteins only identified in large follicles (28 proteins referred as 'off in hSAF' in Fig. 3b) are mostly allocated in the intracellular part. Interestingly, the proportion of secreted proteins was higher in proteins not identified in small follicles, i.e. only identified in large follicles (35/163 versus 13/28).

Furthermore, from the list of proteins identified in hSAF, a group of 24 proteins were highlighted (Supplementary Table SVII). This was carried out following a designed work-flow (see Materials and methods section, Supplementary Fig. S3), which allowed us to access proteins possibly more concentrated or accessible in FF from hSAF. Figure 3c shows the cellular compartments where these proteins are allocated according to the GO analysis. More than 90% of these proteins are intracellular (nucleus and cytosol).

The presence of the 24 proteins in FF from hSAF was verified in 23 samples analysed in this study (10 from different women and 13 from six other women). All of the 24 proteins were identified across the samples (Supplementary Fig. S7, Supplementary Table SVII), with the majority present in more than 70% of the samples.

Interestingly, four of the proteins (AMH, HTRA1, LOXL2 and MDK) are secreted by cells, which could indicate that they play an important role in hSAF, as it is well-known in the case of AMH (Jeppesen *et al.*, 2013). In order to inquire about the function of these proteins, a functional annotation clustering was performed on DAVID (Huang *et al.*, 2009) using an enriched list of proteins from the high abundant proteome that includes the 24 above selected proteins (see Materials and methods section). A cluster of 18 proteins involved in the ovarian follicle development came out (cluster 19 Supplementary Table SVIII, Fig. 3d). AMH and inhibin-B (associated to growth factor activity) are well-known to be present in small follicles at a high concentration (Andersen and Byskov, 2006), which was corroborated in this study by MS. Conversely, among others, the proteins inhibin-A and amphiregulin (AREG) were not identified in the 'high abundance proteome' of hSAF. Two of the four previously mentioned secreted proteins were involved in this cluster: the first one is the well-known AMH and the second one is midkine (MDK).

According to the bioinformatics analyses, specifically, MDK showed relevant evidence that indicated that this protein plays an important role in the follicle. MDK appears to be, out of 24 proteins, the only cell-secreted protein that functionally clusters together with well-known proteins involved in the ovarian follicle development (e.g. inhibin proteins (A and B), AREG, AMH) (Fig. 3d). On the other hand, MDK is involved in biological processes such as growth factor activity and cell differentiation. In a previous study, this protein was suggested to play a role in oocyte maturation in pre-ovulatory follicles (Poulsen *et al.*, 2019). Also, the addition of MDK in the culture medium during IVM seems to improve the maturation rate of oocytes (Nikiforov *et al.*, 2020). Apart from MDK, VIM was another of the 24 proteins more accessible in hSAF that came out in the cluster of ovarian follicle development. This protein is involved in the nuclear reprogramming (Kong *et al.*, 2014; Zhao *et al.*, 2015) and is well-known its function in the cytoskeletal organisation.

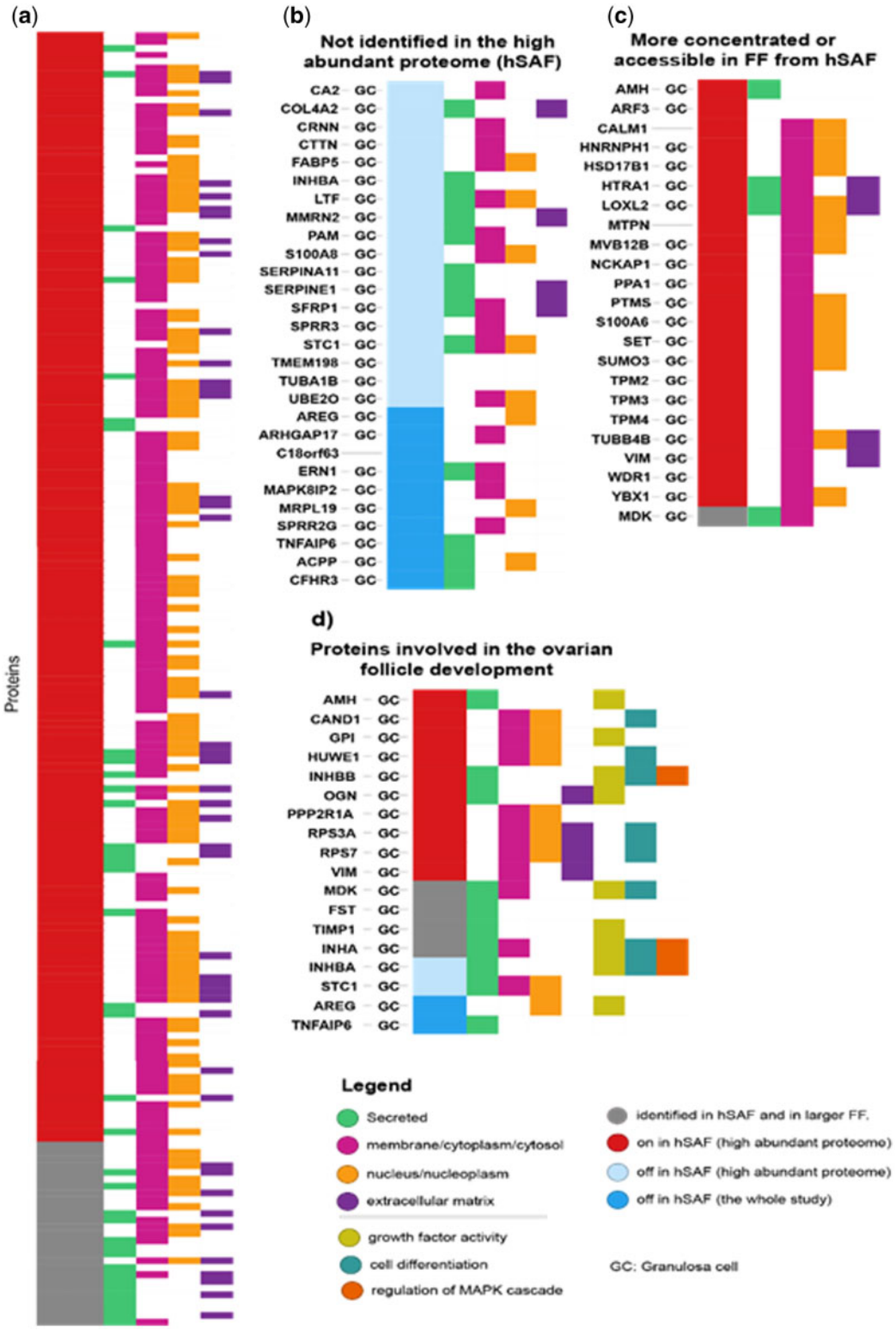


Figure 3. High abundance proteome identified in follicular fluid (FF) from human small antral follicles (hSAF) compared with proteins identified by Poulsen *et al.* (2019) **in large follicles**. In the two studies, protein identification was carried out following the same methodology. (**a**) Red: 163 proteins identified in FF from hSAF and not identified in large follicles. Grey: 35 proteins identified in both studies. (**b**) Proteins identified in FF from large follicles and not identified in hSAF. (**c**) Proteins more concentrated or accessible in hSAF. **(d)** Cluster of 18 proteins grouped by DAVID according to their functional role in the ovarian follicle development.

Functional relationship of MDK and selected proteins

To deepen our understanding of the role that MDK plays in hSAF for follicular development, we focus on proteins strongly correlated with this protein. Expression correlation is known as an indication of a functional association between genes or proteins (Pita-Juárez *et al.*, 2018). Accordingly, we set out to identify proteins and pathways that would likely act together with MDK during follicle development and oocyte maturation using the protein expression quantified in the individual samples. Interestingly, a significant positive strong correlation was found between MDK and VIM ($r = 0.727$, $P = 0.017$). These two proteins, apparently more accessible in hSAF, were also functionally clustered in the previous analysis. Considering the functionality that VIM has in the nuclear reprogramming and the cytoskeletal organisation, we also looked for proteins significantly correlated with VIM. A set of proteins significantly correlated (Pearson correlation: $-0.7 \geq r \geq 0.7$, $P < 0.02$) with MDK and VIM was generated to explore the functional network related to these proteins. The set included 72 and 25 positively and negatively correlated proteins, respectively (Supplementary Table SIX). The proteins that correlated negatively with MDK and VIM were primarily extracellular and cytosolic (Fig. 4b). The proteins that showed a positive correlation were predominantly nuclear proteins.

Proteins that significantly correlated with MDK/VIM (Supplementary Table SIX) were subjected to a biological pathway enrichment analysis. The analysis showed a significant enrichment of proteins positively correlated with MDK and VIM which are associated with gene regulation processes, such as transcription, chromosome maintenance and meiosis (Fig. 4a). These proteins are also associated with processes related to the development and normal function of female reproductive organs (ovaries and uterus), e.g. the epithelial-to-mesenchymal transition (EMT) process (Bilyk *et al.*, 2017). Taken together, these results suggest that MDK and the MDK/VIM protein pair could play a fundamental role in follicle development from early antral follicles and in oocyte maturation at later stages.

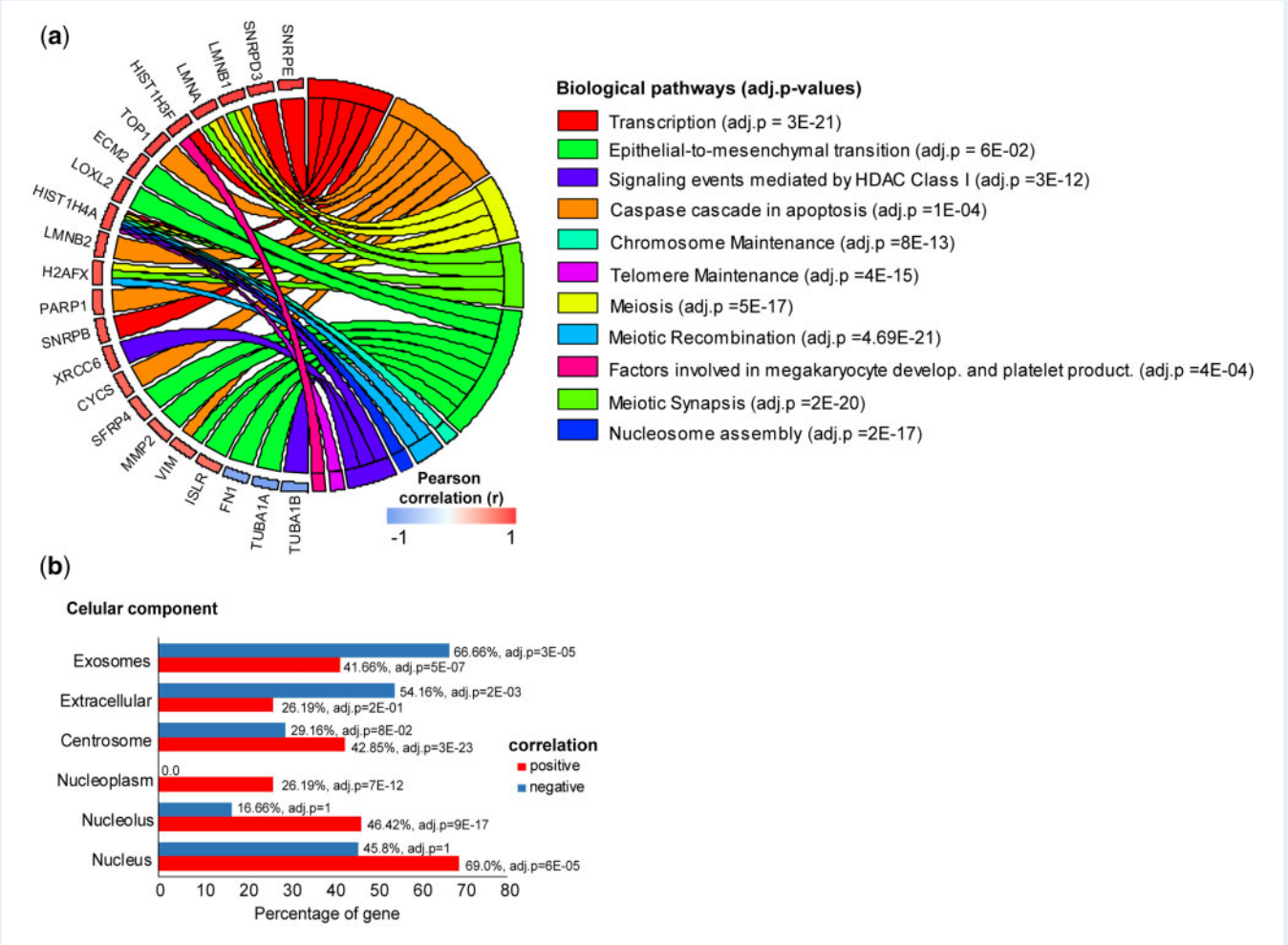


Figure 4. Functional enrichment analysis among proteins correlated positively or negatively with MDK and VIM. (**a**) Biological pathways significantly enriched (BH method: adjusted *P*-value <0.05). (**b**) Comparative enrichment analysis between positively and negatively correlated proteins based on 'cellular components' annotations. Bars represent the percentage of genes in each cellular component. VIM, vimentin.

Proteomic changes in FF from small antral follicles related to subsequent oocyte maturation

We set out to evaluate possible proteomics changes in FF from hSAF related to oocyte maturation. The protein profile of FF that surrounded oocytes capable of reaching MII (n = 7) with the protein profile of FF that surrounded oocytes that remained immature after IVM (n = 6) were compared. The samples were collected from hSAF extracted from six women, from which two or three samples were extracted (see Materials and methods section). The comparison was based on both, a multivariate (sPLS-DA) and univariate analyses (*t*-test). The sPLS-DA method perfectly classified the samples according to their

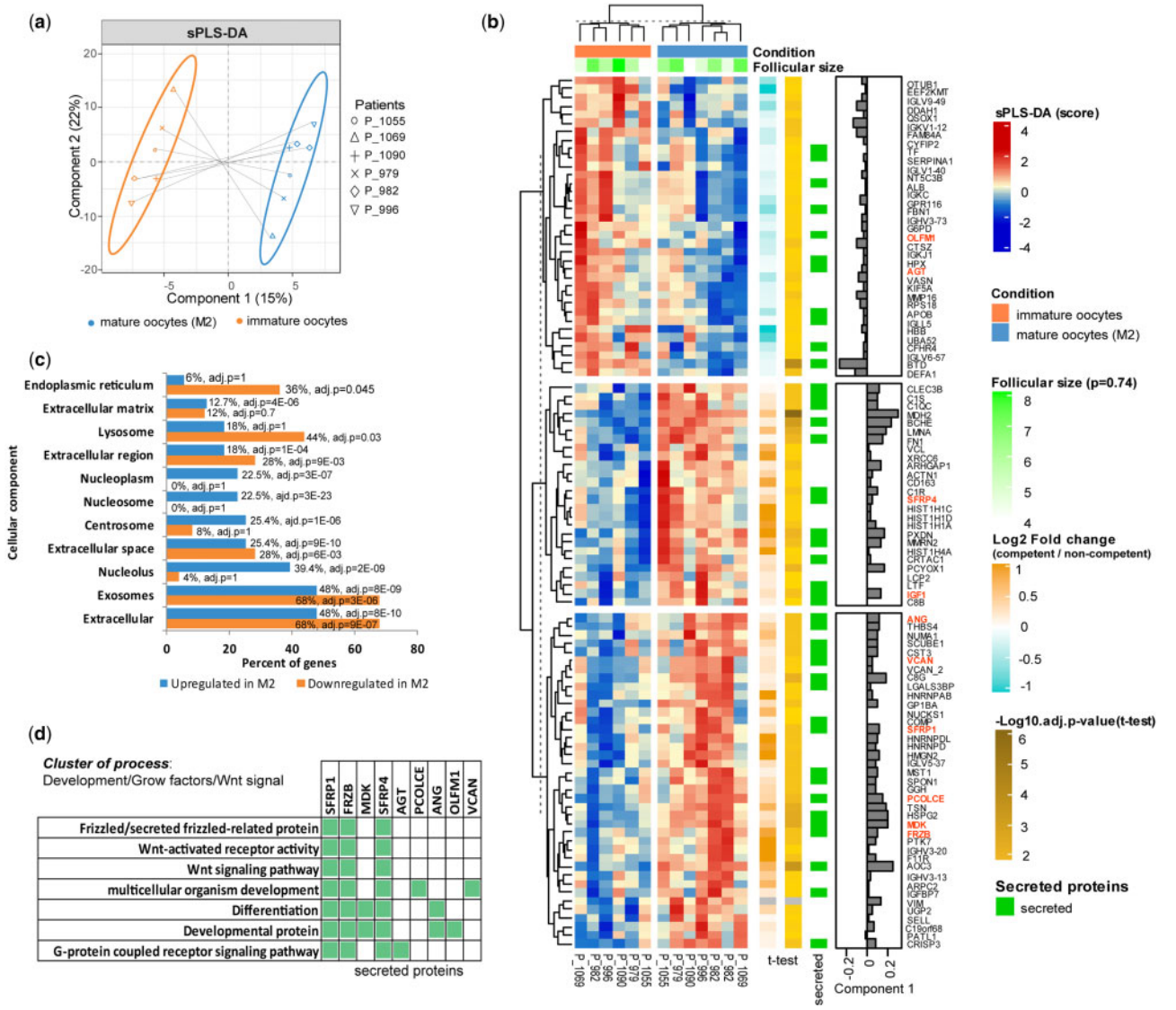


Figure 5. Proteins from human small antral follicles (hSAF) associated with upcoming oocyte maturation. (**a**) Sparse partial squares discriminant analysis performed with 750 proteins quantified in 13 paired FF samples extracted from small antral follicles coming from six women. The analysis discriminated between FF surrounding oocytes capable of achieving metaphase II (MII) after IVM (n = 7, blue) and FF surrounding oocytes unable to mature (n = 6, orange) after IVM. (**b**) Top 100 proteins that contributed to Component 1 of sPLS-DA to discriminate between FF samples. Positive and negative sPLS-DA scores values mean that the protein is up- and down-regulated in FF surrounding oocytes capable to reach M2, respectively. The bar chart indicates the contribution that each protein had in Component 1 (sPLS-DA) to discriminate between groups. (**c**) Comparative enrichment analysis between down- and up-regulated proteins in MII based on 'cellular components' annotations. Bars represent the percentage of genes in each cellular component. (**d**) Functional cluster of secreted proteins involved in development, growth factors and Wnt signal functions (clustered by DAVID software). These proteins were highlighted in red colour in the heat map shown in (b). sPLS-DA, sparse partial least squares discriminant analysis.

sample origin (i.e. FF surrounding oocytes mature or FF surrounding immature) (see Fig. 5a). The top 100 most informative proteins involved in the sample discrimination (shown by Component 1) are displayed in Fig. 5b. Since FF is composed of secretions from granulosa-theca cells and the oocyte, we looked for secreted proteins according to 'The human protein atlas' (https://www.proteinatlas.org/). Out of 100, 42 proteins were secreted (Fig. 5b). The *t*-test also revealed these proteins (Supplementary Table SIV) as significantly dysregulated (adjusted *P*-value <0.05) and the fold changes of the intensities matched with the sPLS-DA scores changes (heat map). Both, down- (35) and up-regulated (65) proteins in FF from small follicles containing oocytes capable of reaching MII were primarily exosome and extracellular proteins (Fig. 5c), however, the per cent of down-regulated proteins was higher. On the other hand, unlike down-regulated proteins, the up-regulated were significantly enriched in several nuclear compartments (Fig. 5c).

As we expected, MDK (adj. $P=0.003$) and VIM (adj. $P=0.016$) were included in the top 100 most dysregulated proteins. Within the groups of up-regulated proteins, there were 11 proteins significantly correlated to MDK and VIM as evaluated above (MDH2, LMNA, FN1, SFRP4, XRCC6, HIST1H4A, ANG, LGALS3BP, HNRNPD, SPON1 and PTK7).

To inquire into the functionality of the top 100 proteins, we performed a functional annotation clustering on the DAVID bioinformatics tool (https://david.ncifcrf.gov/) (Supplementary Table SX). Figure 5d shows one of the functional clusters integrated of secreted proteins (SFRP1, SFRP4, FRZB, MDK, AGT, PCOLCE, ANG, OLFM1, VCAN) involved in development, growth factors and Wnt signal processes. Two and seven of these secreted proteins were down- and up-regulated, respectively.

To further understand the role that these secreted proteins play in the follicle, we performed a biological pathways analysis not only with these proteins, but also with proteins that correlate to them (Supplementary Table SXI). In this case, we wanted to include IGF family proteins since they are stimulators of ovarian follicular development (Mazerbourg and Monget, 2018). IGF1 was included for the correlation since it was significantly up-regulated (adj. $P=0.018$, FC = 0.28) in FF containing oocytes capable of maturation. Interestingly, IGF2 (more commonly related to human folliculogenesis) did not change significantly. The pathways enrichment analysis revealed biological pathways such as transcription, signalling by NOTCH andepitelial-to-mesenchymal transition, among others (see Fig. 6). Proteins involved in the analysis were mostly enriched in the metabolism of protein pathway.

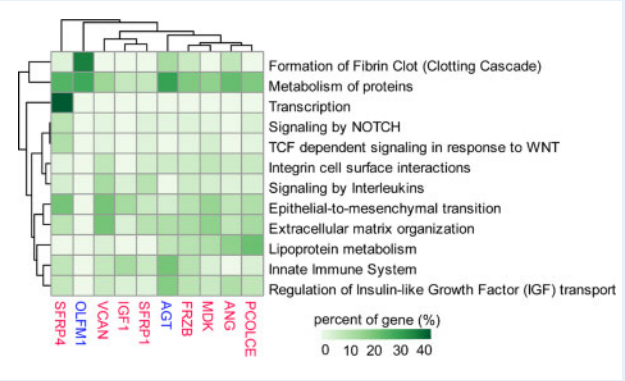


Figure 6. Pathways analysis of proteins that correlate (adjusted $P<0.05$) to secreted proteins involved in development, growth factors and Wnt signal processes. Rows represent the pathways in which the proteins were enriched. The heat map represents the per cent of enriched proteins that correlated to each secreted protein (columns). Proteins that significantly correlate with the secreted proteins are mostly enriched in the metabolism of proteins pathways. Most of the proteins correlated with SFRP4 were enriched in transcription pathways. To see if the correlation is positive or negative, see Supplementary Table SXI. Secreted proteins coloured in red and blue were up- and down-regulated in metaphase II, respectively.

Discussion

This study is to our knowledge, the first proteomics analysis of FF from hSAF obtained from women in their natural menstrual cycle. We were able to identify 2461 proteins of which more than 1108 were new to FF. The data generated constitute the largest number of proteins reported to date in human FF. We present an up-to-date compilation of proteins found in FF (Supplementary Table SV). Furthermore, we identified a signature of FF proteins significantly associated with the ability of the enclosed oocytes to sustain meiotic resumption. This suggests that oocyte viability is affected by the FF already at the early antral stage of follicular development. These results open up for a better understanding of the regulation of human folliculogenesis at earlier stages.

We demonstrated that the analysis by MS of FF from hSAF allows the identification of a greater number of proteins compared to the results obtained from previous analyses of larger follicles. It is well-known that in large follicles, there are more abundant proteins, due to the transfer of plasma constituents through the follicular basal membrane as the follicle expands during the late part of folliculogenesis (Anderson and Anderson, 2002; Zamah *et al.*, 2015). This fact makes the access to low abundance proteins more difficult in large follicles. Most of the new proteins identified in this study come from the cytosol and the nucleus of the cells that surround the antrum.

Follicles used in this study were obtained from ovaries removed surgically from women undergoing fertility preservation. These women did not have diseases related to the ovary and, overall, the ovaries appeared normal at surgery. Thus, we anticipate that the identified proteins could provide relevant physiological information related to folliculogenesis. A possible limitation of our study is the uncertainty of the proportion of the sampled follicles that are undergoing atresia.

On the other hand, we confirmed by MS that AMH and inhibin-B are more concentrated in FF from hSAF compared to large follicles as previously described in an ELISA experiment (Andersen and Byskov, 2006). Conversely, some proteins which were over-represented in the large pre-ovulatory follicles compared to hSAF, may play physiological roles for ovulation rather than for early follicular development. These proteins included inhibin-A and AREG, which are well-known for their determinant functional role in pre-ovulatory follicles (Zamah *et al.*, 2010; Poulsen *et al.*, 2019). Several of these proteins were even up-

regulated during ovulation, including TNFAIP6, AREG, SERPINE1 and ACPP, further highlighting their role in ovulation (Poulsen *et al.*, 2019).

Furthermore, it was possible to identify several signalling pathways and critical components affecting follicular growth and development. Similar biological pathways and processes were identified when we analysed proteins well-correlated to MDK/VIM and proteins well-correlated to those secreted and dysregulated in FF that contain oocytes capable of reaching MII. The biological pathways involved processes linked to follicular development. The chromosome maintenance and nucleosome assembly pathways appeared as part of the chromosome organisation of the GC that supports the progression of follicular growth and maturation. In addition, the EMT process is important for folliculogenesis (Kim *et al.*, 2014) and VIM is a well-known marker of the EMT process (Mirantes *et al.*, 2013). Numerous signalling pathways participate together in EMT up-regulation including PI3K-Akt, EGF, TGF-β, hepatocyte growth factor, MAPK-ERK, NF-kβ, Wnt, Notch, estrogen-receptor-α (ER-α) and HIF-1α. A similar analysis can be performed for the family of frizzled-related proteins (SFRP1, SFRP3, SFRP4) which in addition were up-regulated in FF surrounding oocytes capable of maturing to MII. SFRPs family function as modulators of Wnt signalling through direct interaction with Wnts. This is a targeted pathway that involves secreted glycoproteins that control development in organisms (Mikels and Nusse, 2006) and is also up-regulated in EMT. Thus, the observation of the extracellular matrix protein 2 (ECM2) in the enriched EMT pathway is not surprising. These ECM proteins regulate EMT via interactions with specific integrin receptors (Chen *et al.*, 2013).

Specifically, MDK has been reported to have a pro-survival effect on cumulus-GC (Ikeda and Yamada, 2014). This could explain why proteins significantly correlated (positively) with MDK were enriched in the caspase cascade in the apoptosis pathway. Generally, proteins involved in this pathway are commonly degraded and become markers for apoptosis (VIM, PARP1, LAMNB2 and TOP1). For example, the caspase proteolysis of VIM promotes apoptosis by dismantling intermediate filaments (Byun *et al.*, 2001).

For its part, VIM is a protein responsible for maintaining cell shape and cytoplasm integrity, as well as stabilising cytoskeletal interactions (Eriksson *et al.*, 2009). In addition, VIM was found to be expressed in mouse GC during folliculogenesis and the strongest expression was at earliest stages of follicle growth (Mora *et al.*, 2012). Other studies demonstrated that VIM is required for successful nuclear reprogramming in porcine cloned embryos (Kong *et al.*, 2014).

With the significant dysregulation of 100 proteins, it was demonstrated that changes occur at early follicular stages that could affect the subsequent oocyte maturation process. Interestingly, within the IGF family proteins, IGF1, well-known to be mainly active in non-human species, was up-regulated in this study in FF containing oocytes capable of reaching MII. This finding could be linked with the effect that IGF1 has at early follicular stages. Stubbs *et al.* (2013) demonstrated that IGF-1 stimulates the initiation of follicle growth in cultured ovarian tissue from the normal human ovary (Stubbs *et al.*, 2013). On the other hand, IGF1 has been described to play an important role in mouse oocyte competence and IVM (Toori *et al.*, 2014).

The dysregulated secreted proteins and their correlated network of proteins were involved in pathways such as EMT, extracellular matrix organisation, transcription and metabolism of proteins, among others. In this case, MDK appeared to be up-regulated in FF surrounding oocytes capable of maturing (to MII), which is a piece of valuable information to add to the wide range of studies conducted on this protein in the field of reproductive medicine.

MDK is a member of a family of neurotrophic factors that functions in the central nervous system, and which has been discovered to be expressed in ovarian follicles (Muramatsu *et al.*, 1993; Hirota *et al.*, 2005, 2007; Rauvala, 1989). It has been suggested that MDK is involved in cytoplasmatic maturation of bovine ovarian oocytes and it has been related to the promotion of oocyte developmental competence (Ikeda and Yamada, 2014). This protein acts indirectly via GCs, as no effect is observed in naked oocytes (Ikeda *et al.*, 2006). This is the first time that MDK has been identified in hSAF. Previously, it has been identified in FF from large follicles by western blotting (Hirota *et al.*, 2007) and recently it was identified in large follicles by MS (Poulsen *et al.*, 2019; Zhang *et al.*, 2019; Li *et al.*, 2020). In the study performed by Poulsen *et al.* (2019), MDK was suggested to play a role in oocyte maturation and, recently, the direct impact of MDK in oocyte maturation was assessed by Nikiforov *et al.* (2020). In their study, it was shown that the addition of MDK in the culture medium during IVM significantly improved the maturation rate of oocytes collected from surplus ovarian tissue after fertility preservation (Nikiforov *et al.*, 2020). On the other hand, Özdemir et al. found that the MDK levels in FF and serum may lead to an increase in blastocyst development. They also found that the level of MDK is higher in pregnant than in non-pregnant IVF-ICSI patients (Özdemir *et al.*, 2020).

Collectively, we present in this study, the first characterisation of proteins from hSAF in their natural state by MS. Several proteins observed in this study may have a strong relationship with the follicular developmental process. Furthermore, these data enforce that MDK and VIM are intimately involved in the regulation of oocyte performance in connection with meiotic resumption as observed close to ovulation. Most importantly, however, is that these results demonstrate that profound and significant differences exist in FF from follicles already at a non-selected stage with a diameter of below 9 mm, predicting the ability of the enclosed oocyte to sustain meiotic resumption. If this can be confirmed in further studies, it demonstrates that the viability of the oocyte is determined early on in follicular development and may open up new pathways for augmenting or attenuating subsequent oocyte viability in the pre-ovulatory follicle when ready to undergo ovulation.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE(Perez-Riverol *et al.*, 2019) partner repository with the dataset identifiers PXD013344 and PXD021338.

Acknowledgements

The authors thank the patients who participated in this research project, without whom this research would not be possible. All

personnel, including the clinical activities in fertility preservation, are also thanked for their devoted work.

Authors' roles

C.Y.A. conceived the idea and designed the study together with J.M., I.P., A.S., S.E.P. and G.M.. S.E.P. collected FF and oocytes samples together with L.L.C.P. A.S. designed and executed the mass spectrometry (MS) experiments and acquired the data. I.P. performed bioinformatics and statistical analyses of the proteomics data together with K.P. who performed IPA functional analysis and supervised the bioinformatics analyses. I.P. also interpreted the proteomics results and drafted the manuscript together with A.S. and S.E.P. G.M. and J.M. supervised, contributed and approved the MS experimental design and statistical analyses of proteomics data. K.B.S. participated in the design of MS experiments and did samples management and preparation. C.Y.A. supervised the sample collection and interpreted the results in a biological context together with S.E.P. and L.L.C.P. R.A. managed the collaborative study together with J.M. All the authors revised the manuscript and approved the final version.

Funding

This study is part of the ReproUnion collaborative study, co-financed by Interreg V (European Union programme).

Conflict of interest

The authors have no conflicts of interest to disclose.

References

Ambekar AS, Kelkar DS, Pinto SM, Sharma R, Hinduja I, Zaveri K, Pandey A, Prasad TSK, Gowda H, Mukherjee S. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. *J Clin Endocrinol Metab* 2015;**100**:744–753.

Ambekar AS, Nirujogi RS, Srikanth SM, Chavan S, Kelkar DS, Hinduja I, Zaveri K, Prasad TSK, Harsha HC, Pandey A *et al.* Proteomic analysis of human follicular fluid: a new perspective towards understanding folliculogenesis. *J Proteomics* 2013;**87**:68–77.

Anahory T, Dechaud H, Bennes R, Marin P, Lamb NJ, Laoudj D, Pppelmann M, Becker WM, Petersen A. Identification of new proteins in follicular fluid of mature human follicles. *Electrophoresis* 2002;**23**:1197–1202.

Andersen CY, Byskov AG. Estradiol and regulation of Anti-Müllerian hormone, inhibin-A, and inhibin-B secretion: analysis of small antral and preovulatory human follicles' fluid. *J Clin Endocrinol Metab* 2006;**91**:4064–4069.

Andersen CY, Rosendahl M, Byskov AG, Loft A, Ottosen C, Dueholm M, Schmidt KLTT, Andersen AN, Ernst E. Two successful pregnancies following autotransplantation of frozen/thawed ovarian tissue. *Hum Reprod* 2008;**23**:2266–2272.

Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;**1**:845–867.

Angelucci S, Ciavardelli D, Di GF, Eleuterio E, Sulpizio M, Tiboni GM, Giampietro F, Palumbo P, Di IC. Proteome analysis of human follicular fluid. *Biochim Biophys Acta* 2006;**1764**:1775–1785.

Bayasula IA, Kobayashi H, Goto M, Nakahara T, Nakamura T, Kondo M, Nagatomo Y, Kotani T, Kikkawa F. A proteomic analysis of human follicular fluid: comparison between fertilized oocytes and non-fertilized oocytes in the same patient. *J Assist Reprod Genet* 2013;**30**:1231–1238.

Bianchi L, Gagliardi A, Campanella G, Landi C, Capaldo A, Carleo A, Armini A, De LV, Piomboni P, Focarelli R *et al.* A methodological and functional proteomic approach of human follicular fluid en route for oocyte quality evaluation. *J Proteomics* 2013;**90**:61–76.

Bianchi L, Gagliardi A, Landi C, Focarelli R, Leo V, De Luddi A, Bini L, Piomboni P. Protein pathways working in human follicular fluid: the future for tailored IVF? *Expert Rev Mol Med* 2016;**18**:e9.

Bilyk O, Coatham M, Jewer M, Postovit L-M. Epithelial-to-mesenchymal transition in the female reproductive tract: from normal functioning to disease pathology. *Front Oncol* 2017;**7**:145.

Byun Y, Chen F, Chang R, Trivedi M, Green KJ, Cryns VL. Caspase cleavage of vimentin disrupts intermediate filaments and promotes apoptosis. *Cell Death Differ* 2001;**8**:443–450.

Chen F, Spiessens C, D'Hooghe T, Peeraer K, Carpentier S. Follicular fluid biomarkers for human in vitro fertilization outcome: proof of principle. *Proteome Sci* 2016;**14**:17.

Chen QK, Lee K, Radisky DC, Nelson CM. Extracellular matrix proteins regulate epithelial-mesenchymal transition in mammary epithelial cells. *Differentiation* 2013;**86**:126–132.

Chung D, Keles S. Sparse partial least squares classification for high dimensional data. *Stat Appl Genet Mol Biol* 2010;**9**:17.

Domingues TS, Bonetti TCS, Pimenta DC, Mariano DOC, Barros B, Aquino AP, Motta ELA. Proteomic profile of follicular fluid from patients with polycystic ovary syndrome (PCOS) submitted to in vitro fertilization (IVF) compared to oocyte donors. *J Bras Reprod Assist* 2019;**23**:367–391.

Eriksson JE, Dechat T, Grin B, Helfand B, Mendez M, Pallari HM, Goldman RD. Introducing intermediate filaments: from discovery to disease. *J Clin Invest* 2009;**119**:1763–1771.

Estes SJ, Ye B, Qiu W, Cramer D, Hornstein MD, Missmer SA. A proteomic analysis of IVF follicular fluid in women ≤32 years old. *Fertil Steril* 2009;**92**:1569–1578.

Gougeon A. Human ovarian follicular development: from activation of resting follicles to preovulatory maturation. *Ann Endocrinol (Paris)* 2010;**71**:132–143.

Hanrieder J, Nyakas A, Naessén T, Bergquist J. Proteomic analysis of human follicular fluid using an alternative bottom-up approach. *J Proteome Res* 2008;**7**:443–449.

Hess KA, Chen L, Larsen WJ. The ovarian blood follicle barrier is both charge- and size-selective in mice1. *Biol Reprod* 1998;**58**:705–711.

Hirota Y, Osuga Y, Koga K, Yoshino O, Hirata T, Harada M, Morimoto C, Yano T, Tsutsumi O, Sakuma S *et al.* Possible implication of midkine in the development of endometriosis. *Hum Reprod* 2005;**20**:1084–1089.

Hirota Y, Osuga Y, Nose E, Koga K, Yoshino O, Hirata T, Yano T, Tsutsumi O, Sakuma S, Muramatsu T *et al.* The presence of midkine and its possible implication in human ovarian follicles. *Am J Reprod Immunol* 2007;**58**:367–373.

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44–57.

Ikeda S, Saeki K, Imai H, Yamada M. Abilities of cumulus and granulosa cells to enhance the developmental competence of bovine oocytes during in vitro maturation period are promoted by midkine; a possible implication of its apoptosis suppressing effects. *Reproduction* 2006;**132**:549–557.

Ikeda S, Yamada M. Midkine and cytoplasmic maturation of mammalian oocytes in the context of ovarian follicle physiology. *Br J Pharmacol* 2014;**171**:827–836.

Jarkovska K, Martinkova J, Liskova L, Haiada P, Moos J, Rezabek K, Gadher SJ, Kovarova H, Halada P, Moos J *et al*. Proteome mining of human follicular fluid reveals a crucial role of complement cascade and key biological pathways in women undergoing in vitro fertilization. *J Proteome Res* 2010;**9**:1289–1301.

Jeppesen JV, Anderson RA, Kelsey TW, Christiansen SL, Kristensen SG, Jayaprakasan K, Raine-Fenning N, Campbell BK, Yding Andersen C. Which follicles make the most anti-Müllerian hormone in humans? Evidence for an abrupt decline in AMH production at the time of follicle selection. *Mol Hum Reprod* 2013;**19**:519–527.

Kim YS, Yi BR, Kim NH, Choi KC. Role of the epithelial-mesenchymal transition and its effects on embryonic stem cells. *Exp Mol Med* 2014;**46**:e108.

Kõks S, Velthut A, Sarapik A, Altmäe S, Reinmaa E, Schalkwyk LC, Fernandes C, Lad HV, Soomets U, Jaakma Ü *et al*. The differential transcriptome and ontology profiles of floating and cumulus granulosa cells in stimulated human antral follicles. *MHR Basic Sci Reprod Med* 2010;**16**:229–240.

Kong Q, Xie B, Li J, Huan Y, Huang T, Wei R, Lv J, Liu S, Liu Z. Identification and characterization of an oocyte factor required for porcine nuclear reprogramming. *J Biol Chem* 2014;**289**:6960–6968.

Kristensen SG, Pors SE, Andersen CY. Diving into the oocyte pool. *Curr Opin Obstet Gynecol* 2017;**29**:112–118.

Kuang T-M, Liu CJ-L, Ko Y-C, Lee S-M, Cheng C-Y, Chou P. Distribution and associated factors of optic disc diameter and cup-to-disc ratio in an elderly Chinese population. *J Chinese Med Assoc* 2014;**77**:203–208.

Lee HC, Lee SW, Kyo WL, Lee SW, Cha KY, Kye HK, Lee S. Identification of new proteins in follicular fluid from mature human follicles by direct sample rehydration method of two-dimensional polyacrylamide gel electrophoresis. *J Korean Med Sci* 2005;**20**:456–460.

Lewandowska AE, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S. Qualitative and quantitative analysis of proteome and peptidome of human follicular fluid using multiple samples from single donor with LC−MS and SWATH methodology. *J Proteome Res* 2017;**16**:3053–3067.

Lewandowska AE, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S. Human follicular fluid proteomic and peptidomic composition quantitative studies by SWATH-MS methodology. Applicability of high pH RP-HPLC fractionation. *J Proteomics* 2019;**191**:131–142.

Li H, Huang X, Chang X, Yao J, He Q, Shen Z, Ji Y, Wang K. S100-A9 protein in exosomes derived from follicular fluid promotes inflammation via activation of NF-κB pathway in polycystic ovary syndrome. *J Cell Mol Med* 2019;**24**:114–125.

Lim H-JJ, Seok AE, Han J, Lee J, Lee S, Kang H-GG, Cha BH, Yang Y. N-glycoproteomic analysis of human follicular fluid during natural and stimulated cycles in patients undergoing in vitro fertilization. *Clin Exp Reprod Med* 2017;**44**:63–72.

Liu X, Wang Y, Zhu P, Wang J, Liu J, Li N, Wang W, Zhang W, Zhang C, Wang Y *et al*. Human follicular fluid proteome reveals association between overweight status and oocyte maturation abnormality. *Clin Proteom* 2020a;**17**:22.

Liu Y, Wu Y, Tian M, Luo W, Zhang C, Liu Y, Li K, Cheng W, Liu D. Protein expression profile in IVF follicular fluid and pregnancy outcome analysis in euthyroid women with thyroid autoimmunity. *ACS Omega* 2020b;**5**:11439–11447.

Mazerbourg S, Monget P. Insulin-like growth factor binding proteins and IGFBP proteases: A dynamic system regulating the ovarian folliculogenesis. *Front Endocrinol (Lausanne)* 2018;**9**:134. doi: 10.3389/fendo.2018.0013

Mcnatty KP, Hillier SG, Boogaard AMJ, van den Trimbos-Kemper TCM, Reichert LE, Hall EV. van Follicular development during the luteal phase of the human menstrual cycle. *J Clin Endocrinol Metab* 1983;**56**:1022–1031.

Mikels AJ, Nusse R. Wnts as ligands: processing, secretion and reception. *Oncogene* 2006;**25**:7461–7468.

Mirantes C, Espinosa I, Ferrer I, Dolcet X, Prat J, Matias-Guiu X. Epithelial-to-mesenchymal transition and stem cells in endometrial cancer. *Hum Pathol* 2013;**44**:1973–1981.

Mora JM, Fenwick MA, Castle L, Baithun M, Ryder TA, Mobberley M, Carzaniga R, Franks S, Hardy K. Characterization and significance of adhesion and junction-related proteins in mouse ovarian follicles. *Biol Reprod* 2012;**86**:1–14.

Muramatsu H, Shirahama H, Yonezawa S, Maruta H, Muramatsu T, Midkine A. Retinoic acid-inducible growth/differentiation factor: immunochemical evidence for the function and distribution. *Dev Biol* 1993;**159**:392–402.

Nikiforov D, Junping C, Cadenas J, Shukla V, Blanshard R, Pors SE, Kristensen SG, Macklon KT, Colmorn L, Ernst E *et al*. Improving the maturation rate of human oocytes collected ex vivo during the cryopreservation of ovarian tissue. *J Assist Reprod Genet* 2020;**37**:891–904.

Oh JW, Kim SK, Cho K-C, Kim M-S, Suh CS, Lee JR, Kim KP. Proteomic analysis of human follicular fluid in poor ovarian responders during in vitro fertilization. *Proteomics* 2017;**17**:1600333.

Özdemir A, Karli P, Avci B. Do midkine levels in serum and follicular fluid affect IVF-ICSI outcome? *J Heal Sci Med* 2020;**3**:301–306.

Pathan M, Keerthikumar S, Ang C-S, Gangoda L, Quek CYJ, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A *et al*. FunRich: an open access standalone functional enrichment and interaction network analysis tool. *Proteomics* 2015;**15**:2597–2601.

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M *et al*. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 2019;**47**:D442–D450.

Pita-Juárez Y, Altschuler G, Kariotis S, Wei W, Koler K, Green C, Tanzi RE, Hide W. The pathway coexpression network: revealing pathway relationships. *PLoS Comput Biol* 2018;**14**:e1006042.

Poulsen L, la C, Pla I, Sanchez A, Grøndahl ML, Marko-Varga G, Yding Andersen C, Englund ALM, Malm J. Progressive changes in human follicular fluid composition over the course of ovulation: quantitative proteomic analyses. *Mol Cell Endocrinol* 2019;**495**:110522.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna. https://www.r-project.org/. version 3.6.3 (22 February 2020, date accessed) 2016.

Rauvala H. An 18-kd heparin-binding protein of developing brain that is distinct from fibroblast growth factors. *EMBO J* 1989;**8**:2933–2941.

Regiani T, Cordeiro FB, Costa L, do V. D, Salgueiro J, Cardozo K, Carvalho VM, Perkel KJ, Zylbersztejn DS, Cedenho AP, Turco EG. Lo Follicular fluid alterations in endometriosis: label-free proteomics by MS ^{E} as a functional tool for endometriosis. *Syst Biol Reprod Med* 2015;**61**:263–276.

Rodgers RJ, Irving-Rodgers HF. Formation of the ovarian follicular antrum and follicular fluid. *Biol Reprod* 2010;**82**:1021–1029.

RStudio Team. *RStudio: Integrated Development for R*. RStudio.Boston. http://www.rstudio.com/. version 1.1.463. (13 September 2019, date accessed) 2016.

Schmidt KT, Rosendahl M, Ernst E, Loft A, Andersen AN, Dueholm M, Ottosen C, Andersen CY. Autotransplantation of cryopreserved ovarian tissue in 12 women with chemotherapy-induced premature ovarian failure: the Danish experience. *Fertil Steril* 2011; **95**:695–701.

Schweigert FJ, Gericke B, Wolfram W, Kaisers U, Dudenhausen JW. Peptide and protein profiles in serum and follicular fluid of women undergoing IVF. *Hum Reprod* 2006;**21**:2960–2968.

Severino V, Malorni L, Cicatiello AE, D'Esposito V, Longobardi S, Colacurci N, Miraglia N, Sannolo N, Farina A, Chambery A. Array and iTRAQ labeling for in-depth identification of pathways associated to IVF outcome. *PLoS One* 2013;**8**:e77303.

Shen X, Liu X, Zhu P, Zhang Y, Wang J, Wang Y, Wang W, Liu J, Li N, Liu F. Proteomic analysis of human follicular fluid associated with successful in vitro fertilization. *Reprod Biol Endocrinol* 2017;**15**:58.

Siu MKY, Cheng CY. The Blood-Follicle Barrier (BFB) in disease and in ovarian function. *Adv Exp Med Biol* 2012;**763**:186–192.

Spitzer D, Murach KF, Lottspeich F, Staudach A, Illmensee K. Different protein patterns derived from follicular fluid of mature and immature human follicles. *Hum Reprod* 1996;**11**: 798–807.

Stubbs SA, Webber LJ, Stark J, Rice S, Margara R, Lavery S, Trew GH, Hardy K, Franks S. Role of Insulin-like growth factors in initiation of follicle growth in normal and polycystic human ovaries. *J Clin Endocrinol Metab* 2013;**98**:3298–3305.

Toori MA, Mosavi E, Nikseresht M, Barmak MJ, Mahmoudi R. Influence of insulin-like growth factor-i on maturation and fertilization rate of immature oocyte and embryo development in NMRI mouse with TCM199 and α-MEM medium. *J Clin Diagnostic Res* 2014;**8**:AC05–AC08.

Twigt JM, Bezstarosti K, Demmers J, Lindemans J, Laven JSE, Steegers-Theunissen RP. Preconception folic acid use influences the follicle fluid proteome. *Eur J Clin Invest* 2015;**45**:833–841.

Twigt JM, Bolhuis MEC, Steegers EAP, Hammiche F, van Inzen WG, Laven JSE, Steegers-Theunissen RPM. The preconception diet is associated with the chance of ongoing pregnancy in women undergoing IVF/ICSI treatment. *Hum Reprod* 2012;**27**:2526–2531.

Virant-Klun I, Leicht S, Hughes C, Krijgsveld J. Identification of maturation-specific proteins by single-cell proteomics of human oocytes. *Mol Cell Proteomics* 2016;**15**:2616–2627.

Westergaard H. The sprue syndromes. *Am J Med Sci* 1985;**290**: 249–262.

Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods* 2009;**6**:359–362.

Zakerkish F, Brännström M, Carlsohn E, Sihlbom C, Post S, Thoroddsen A. Proteomic analysis of follicular fluid during human ovulation. *Acta Obstet Gynecol Scand* 2020;**99**:917–924.

Zamah AM, Hassis ME, Albertolle ME, Williams KE. Proteomic analysis of human follicular fluid from fertile women. *Clin Proteom* 2015; **12**:1–12.

Zamah AM, Hsieh M, Chen J, Vigne JL, Rosen MP, Cedars MI, Conti M. Human oocyte maturation is dependent on LH-stimulated accumulation of the epidermal growth factor-like growth factor, amphiregulin. *Hum Reprod* 2010;**25**:2569–2578.

Zhang X, Xu X, Li P, Zhou F, Kong L, Qiu J, Yuan Z, Tan J. TMT based proteomic analysis of human follicular fluid from overweight/obese and normal-weight patients with polycystic ovary syndrome. *Front Endocrinol (Lausanne)* 2019;**10**:821.

Zhao Q, Guo Z, Piao S, Wang C, An T. Discovery of porcine maternal factors related to nuclear reprogramming and early embryo development by proteomic analysis. *Proteome Sci* 2015;**13**:18.

Sagvekar et al. Clinical Epigenetics (2019) 11:61
https://doi.org/10.1186/s13148-019-0657-6

Clinical Epigenetics

RESEARCH Open Access

DNA methylome profiling of granulosa cells reveals altered methylation in genes regulating vital ovarian functions in polycystic ovary syndrome

Pooja Sagvekar[1], Pankaj Kumar[2], Vijay Mangoli[3], Sadhana Desai[3] and Srabani Mukherjee[1*]

Abstract

Background: Women with polycystic ovary syndrome (PCOS) manifest a host of ovarian defects like impaired folliculogenesis, anovulation, and poor oocyte quality, which grossly affect their reproductive health. Addressing the putative epigenetic anomalies that tightly regulate these events is of foremost importance in this disorder. We therefore aimed to carry out DNA methylome profiling of cumulus granulosa cells and assess the methylation and transcript expression profiles of a few differentially methylated genes contributing to ovarian defects in PCOS. A total of 20 controls and 20 women with PCOS were selected from a larger cohort of women undergoing IVF, after carefully screening their sera and follicular fluids for hormonal and biochemical parameters. DNA extracted from cumulus granulosa cells of three women each, from control and PCOS groups was subjected to high-throughput, next generation bisulfite sequencing, using the Illumina HiSeq 2500® platform. Remaining samples were used for the validation of methylation status of some identified genes by pyrosequencing, and the transcript expression profiles of these genes were assessed by quantitative real-time PCR.

Results: In all, 6486 CpG sites representing 3840 genes associated with Wnt signaling, G protein receptor, endothelin/integrin signaling, angiogenesis, chemokine/cytokine-mediated inflammation, etc., showed differential methylation in PCOS. Hypomethylation was noted in 2977 CpGs representing 2063 genes while 2509 CpGs within 1777 genes showed hypermethylation. Methylation differences were also noted in noncoding RNAs regulating several ovarian functions that are dysregulated in PCOS. Few differentially methylated genes such as aldo-keto reductase family 1 member C3, calcium-sensing receptor, resistin, mastermind-like domain 1, growth hormone-releasing hormone receptor and tumor necrosis factor, which predominantly contribute to hyperandrogenism, premature luteolysis, and oocyte development defects, were explored as novel epigenetic candidates in mediating ovarian dysfunction. Methylation profiles of these genes matched with our NGS findings, and their transcript expression patterns correlated with the gene hypo- or hypermethylation status.

Conclusion: Our findings suggest that the epigenetic dysregulation of genes involved in important processes associated with follicular development may contribute to ovarian defects observed in women with PCOS.

Keywords: PCOS, Methylome profiling, DNA methylation, IVF, Granulosa cells, NGS

* Correspondence: mukherjees@nirrh.res.in; srabanimuk@yahoo.com
[1]Department of Molecular Endocrinology, ICMR-National Institute for Research in Reproductive Health, J.M. Street, Parel, Mumbai, Maharashtra 400012, India
Full list of author information is available at the end of the article

Background

Polycystic ovary syndrome (PCOS), one of the leading causes of anovulatory infertility, is characterized by ovarian, neuroendocrine, and metabolic perturbations in women of reproductive age group. With a global prevalence of 6–15% [1], it generally features irregularity or absence of menses and presence of multiple ovarian cysts on ultrasonography, in addition to hyperandrogenemia and systemic insulin excess. Increased pulsatility of GnRH neurons at the hypothalamo-hypophyseal interface elevates gonadotropic secretion of LH over FSH, which alongside co-gonadotropic actions of insulin, promotes increased production of androgens from ovarian folliculo-thecal cells [2, 3]. Adding to the pool of free androgens in circulation is the insulin-mediated suppression of hepatic sex hormone binding globulin (SHBG), which increases the bioavailability of testosterone, thereby affecting its clearance [4]. The cumulative actions of all these events trigger a series of physiological defects including ovarian cyst formation, amenorrhoea, anovulation, infertility, hyperandrogenism, insulin resistance, hyperinsulinemia, obesity, glucose intolerance, lipid abnormalities, type 2 diabetes mellitus (T2DM), hypertension, and cardiovascular disease. In parallel, there is an intrinsic elevation of anti-Mullerian hormone (AMH) due to the presence of cystic follicles arrested in preantral to antral stages [5]. These key events dictate the principal dogma behind the pathophysiology of PCOS understood so far.

Although genetic factors impacting the development and progression of PCOS have been amply investigated, identification and experimental corroboration of cognate epigenetic factors that may contribute to the pathophysiology of this multifaceted disorder remain enigmatic. Both environmental and physiological factors serve as strong determinants for epigenetic alterations. Factors such as intrinsic hormonal aberrations, dysregulation of intrauterine milieu by endocrine disruptors (EDCs) during gestational periods, and lifestyle modifications in subsequent phases of growth and development have been recently implicated in epigenetic predisposition to this disease. Pioneering investigations on epigenetic alterations in PCOS began with studies on peripheral blood leukocytes (PBLs) [6]. However, tissue specificity of epigenetic modifications renders it difficult to extrapolate epigenetic data derived from circulating cells like PBLs, to organs such as ovaries or adipose tissues that are highly affected in PCOS [7]. This necessitates the undertaking of clinical epigenetic studies at a tissue-specific level. Ovary withstands most of the hormonal assaults triggered by systemic aberrations in the neuroendocrine-ovarian axis. It is therefore a primary hotspot for epigenetic perturbations, which may contribute to the multiple follicular and oocyte defects observed in women with PCOS. Defects related to steroidogenesis, follicular growth and dominance, ovulation, oocyte developmental competence, cumulus-oophorus complex (COC) expansion, luteal maintenance, etc., which are under stringent control of gonadotropins and other hormones, have been well documented so far. Therefore, identification of locus/gene-specific epigenetic alterations in the ovaries of these women is of prime significance to understand the pathophysiology of this disorder.

So far, few studies conducted to identify global DNA methylation differences in PBLs and mural as well as cumulus granulosa cells (CGCs) of women with PCOS have yielded ambiguous findings [6, 8, 9]. Also, promoter methylation profiles of a few established candidate genes of PCOS including yes-associated protein (*YAP1*), follistatin (*FST*), aromatase (*CYP19A1*) and luteinizing hormone chorionic gonadotropin receptor (LHCGR) have been investigated in these cells, and ovarian tissues by some groups till date [10–14]. Among these genes, *LHCGR* has been consistently reported to be hypomethylated in women with PCOS and in animal models of PCO [14, 15]. Subsequently, a few high-throughput attempts were made to identify some differentially methylated genes (DMGs) in women with PCOS. These included the use of diverse approaches such as methylated DNA immunoprecipitation (MeDIP) [16] and Illumina platform-based methylation microarray [17] in peripheral blood leukocytes; MeDIP coupled with methyl promoter enrichment microarray [18], as well as methylation microarray combined with microarray-based transcriptome analysis [19] in ovarian tissue biopsies and adipose tissue samples [20]; total RNA sequencing (RNA-seq) coupled with methylation measured by base cleavage and mass spectrometry (EpiTYPER) [21]; and lastly, methylation microarray in mural granulosa cells (MGCs) collected from women undergoing controlled ovarian hyperstimulation (COH) [22]. However, next generation sequencing (NGS)-based methylome studies spanning individual CpG sites, specifically in CGCs which participate in extensive cross talk between the developing oocyte and surrounding follicular milieu while facilitating meiotic maturation and ovulation of competent oocytes, are yet to be reported in women with PCOS. In this study, we have carried out comparative genome-wide bisulfite sequencing of CGCs obtained from women with PCOS and normovulatory healthy controls, using a NGS-based, multiplexed Methyl-Capture Sequencing (MC-Seq) approach. MC-Seq is advantageous over reduced representative bisulfite sequencing (RRBS) and MeDIP-Seq, in that it avoids the over-representation of recurring reads and also over Infinium 450 K methyl-microarray wherein it enables the user to opt for much greater coverage of the epigenome (3.4 million CpG sites and 20.8 million non-CpG sites, as opposed to 4,50,000 CpG sites and 3091 non-CpG loci, respectively) [23]. Also, Infinium microarrays target only those genes

that are known to be differentially methylated in some cancers while MC-Seq differentiates between any genomic region bearing altered methylation marks. Additionally, compared to whole-genome bisulfite sequencing (WGBS), MC-Seq is cost effective and can identify novel genomic loci while reducing the processing time associated with WGBS. This study provides insights on genome-wide data on CpG sites that show altered methylation at a single base resolution in CGCs of women with PCOS.

Results

Hormonal characterization and oocyte quality parameters of the study participants

All study participants were subjected to stringent anthropometric, hormonal, and biochemical characterization prior to further investigation (Table 1). Baseline hormonal and biochemical profiling of 20 controls and 20 women with PCOS (days 3–5 of the follicular phase of menstrual cycle) showed that the baseline levels (day 3 follicular phase estimates in serum) of luteinizing hormone (LH), LH/follicle-stimulating hormone (FSH) ratio, and of AMH were high, while the levels of FSH were low in women with PCOS, compared to controls. Prolactin and thyroid-stimulating hormone (TSH) levels were similar in both groups. Serum levels of estradiol (E_2) and progesterone (P_4) measured before the administration of recombinant human chorionic gonadotropin (rhCG) were comparable between controls and PCOS. Serum E_2 levels on the day of rhCG administration were high in women with PCOS while P_4 levels were similar between these groups. Analysis of oocyte parameters revealed that the total numbers of follicles, oocytes (mature + immature), mature MII oocytes, %MII oocytes, and number of fertilized oocytes were unchanged between controls and women with PCOS. However, the rates of fertilization of MII oocytes were low in women with PCOS compared to controls. In serum and follicular fluid (FF) samples collected on the day of oocyte pick up (d-OPU), total testosterone (TT) levels were high and sex hormone binding globulin (SHBG) levels were low in PCOS. Androgen-excess indices such as free testosterone (Free-T), bioavailable testosterone (Bio-T), and free androgen index (FAI) were high in FF, while in serum, only Bio-T and FAI were high in PCOS.

Identification of differentially methylated targets and their gene ontology analysis

MC-Seq of CGCs of women with PCOS and controls identified a total of 6486 differentially methylated CpG sites associated with 3403 unique genes across the genome, of which 2977 CpG sites were hypomethylated and 2509 CpG sites were hypermethylated. Hypomethylated CpG sites were representative of 2063 (Additional file 1) genes in all, while the hypermethylated sites were linked to a total of 1777 genes (Additional file 2). Of the total DMGs, 438 genes harbored both hyper- and hypomethylated CpG sites. Additionally, many noncoding RNAs including 44 microRNAs (miRs) and 121 pseudogenes also showed differential methylation in women with PCOS (Additional files 1 and 2). Of the 44 differentially methylated miRs, several miRs (miR23A, miR127, miR10B, miR193A mir200B, miR182, and miR140) have been reported to be implicated in impaired follicle growth and steroidogenesis, anovulation, obesity, glucose metabolism, and so on [24]. In pathway enrichment and gene ontology (GO) analyses of the hypomethylated, hypermethylated, and combined gene lists, 227 (11%), 210 (11.82%), and 396 (11.64%) genes from the three respective categories could not be annotated. Among the identified pathways, those for Wnt signaling (Panther-GO-ID, P00057), integrin signaling (P00034), endothelin signaling (P00019), and cadherin signaling (P00012) were enriched in both the hypo- and hypermethylated gene sets (Fig. 1). Other prominent pathways included the platelet-derived growth factor (PDGF) signaling (P00047), inflammation mediated by chemokine and cytokine signaling (P00031), angiogenesis (P00005) and vascular endothelial growth factor (VEGF) signaling (P00056), fibroblast growth factor (FGF) signaling (P00021), G protein signaling (P00026, P00027), T cell activation (P00053), and nicotinic acetylcholine receptor signaling pathways (P00044).

Validation of genes showing differential methylation in regions upstream to transcription start sites

Among the DMGs selected for validation, the upstream CpG sites of five genes, namely aldo-keto reductase 1 family C3 (*AKR1C3*), calcium-sensing receptor (*CASR*), growth hormone-releasing hormone receptor (*GHRHR*), resistin (*RETN*), and mastermind-like domain 1 (*MAMLD1*) were hypomethylated while those of transferrin (*TF*) and tumor necrosis factor (*TNF*) were hypermethylated in PCOS in our methylome analysis. We first investigated whether these seven genes are expressed in CGCs, and whether they are differentially expressed in women with PCOS using qPCR (Fig. 2). Apart from *GHRHR*, which was expressed only at baseline levels, the remaining six genes were abundantly expressed in CGCs. Transcripts of *AKR1C3*, *CASR*, *GHRHR*, *RETN*, and *MAMLD1* were upregulated while those of *TF* and *TNF* were downregulated in CGCs of PCOS women (Fig. 2). To verify whether our NGS findings were replicative in a larger study cohort, we performed pyrosequencing to assess the average percent methylation of these genes in 17 controls and 17 women with PCOS. Hypomethylation of *AKR1C3*, *CASR*, *GHRHR*, *RETN*, and *MAMLD1* genes at the indicated CpG sites was confirmed by pyrosequencing (Fig. 3). Decreased methylation in these genes was consistent with the upregulation of their respective transcripts in women with PCOS (Fig. 2), and the two variables showed an inverse correlation with one another

Table 1 Clinical characteristics of study participants undergoing controlled ovarian hyperstimulation (COH) assessed before and after oocyte retrieval

Parameters assessed	Control ($n = 20$) median (IQR)	PCOS ($n = 20$) median (IQR)	P value
Age in years	30.0 (27.0–31.0)	31.5 (28.0–33.0)	0.124
BMI (kg m^{-2})	23.03 (20.97–25.06)	24.65 (22.24–27.76)	0.173
#Basal FSH (μU/mL)	7.64 (5.52–9.39)	5.71 (4.39–6.61)	0.037*
#Basal LH (μU/mL)	6.19 (3.0–7.66)	9.13 (5.58–12.24)	0.02*
#LH:FSH	0.75 (0.55–1.04)	1.74 (1.15–2.04)	0.0001***
#Prolactin (ng/mL)	14.2 (12.0–18.27)	14.48 (11.72–24.3)	0.43
#TSH (mIU/mL)	1.8 (1.14–2.75)	2.36 (1.55–4.08)	0.0577
#AMH (ng/mL)	3.4 (1.94–7.86)	7.99 (5.21–9.74)	0.0094**
Regular cycle	20 (100%)	3 (15%)	
Oligomenorrhea	0 (0%)	15 (80%)	<0.0001***
Secondary amenorrhea	0 (0%)	2 (5%)	
rFSH (IU)	1600 (1528–2313)	1706 (1350–2025)	0.788
E_2 (ng/mL) before hCG administration	1.49 (1.247–2.24)	2.03 (1.22–2.47)	0.367
E_2 (ng/mL) on hCG administration day	1.71 (1.436–2.32)	2.56 (1.82–3.72)	0.038*
P_4 (ng/mL) before hCG administration	0.3 (0.2–0.55)	0.25 (0.17–0.525)	0.522
P_4 (ng/mL) on hCG administration day	2.65 (2.07–4.05)	4.6 (1.95–6.32)	0.200
Total follicles (*n*)	15.0 (12.0–20.5)	18.5 (15.0–29.25)	0.099
Total oocytes (*n*)	13.5 (10.0–17.25)	16.5 (14.5–26.75)	0.059
MII oocytes (*n*)	10.5 (8.0–13.5)	15.0 (9.0–19.75)	0.057
MII oocytes (%)	82.84 (70.24–90.71)	86.85 (74.2–95.31)	0.366
Total fertilized oocytes (*n*)	6.0 (4.5–9.25)	6.5 (3.75–14.5)	0.464
Rate of MII oocyte fertilization (ROF)	65.63 (51.92–88.13)	52.27 (34.8–67.)	0.039*
E_2 (ng/mL) Serum	0.2 (0.15–0.6)	0.43 (0.15–0.76)	0.418
E_2 (ng/mL) FF	374 (178.6–541.7)	280.3 (125–819)	0.586
P_4 (ng/mL) Serum	1 (3–9)	3.9 (0.3–7.)	0.903
P_4 (ng/mL) FF	8170 (4760–15,880)	6000 (3280–10,140)	0.325
TT (ng/dL) Serum	126 (89–178.5)	172.6 (140–311.3)	0.020*
TT (ng/dL) FF	300 (140.2–372.6)	405.8 (222.6–639.3)	0.030*
SHBG (nmol/L) Serum	130 (103.0–150.0)	95 (79–134.5)	0.041*
SHBG (nmol/L) FF	158 (127.8–205.5)	127.7 (103.3–148.8)	0.024*
Free T (pmol/L) Serum	3.53 (2.07–4.85)	6.03 (3.93–9.73)	0.066
Free T (pmol/L) FF	4.24 (1.55–7.87)	7.5 (5.28–11.74)	0.050*
Bio-T (nmol/L) Serum	0.82 (0.48–1.14)	1.09 (0.8–1.9)	0.033*
Bio-T (nmol/L) FF	0.99 (0.58–1.84)	1.75 (1.24–2.74)	0.050*
FAI serum	3.03 (2.38–6.48)	6.03 (3.93–9.73)	0.035*
FAI FF	3.7 (2.32–7.17)	6.93 (5.48–13.42)	0.050*

Data are represented as median (inter-quartile range) for anthropometric and hormonal characteristics compared been controls and women with PCOS using Mann-Whitney *U* tests. Parameters marked with asterisk (#) denote those measured between days 3–5 of the menstrual cycle (early follicular phase) before initiating the controlled ovarian hyperstimulation (COH) procedure. Parameters marked by "$" were measured in sera and follicular fluids obtained on the day of ovum pick up (d-OPU). Menstrual characteristics were assessed using the chi-square analysis. P values < 0.05 are considered significant for all statistical tests. *P < 0.05, **P < 0.01, ***P </= 0.0001 have been indicated. *BMI* body-mass index, E_2 estradiol, P_4 progesterone, *TT* total testosteronem, *SHBG* sex hormone binding globulin, *Bio-T* bioavailable testosterone, *Free-T* free testosterone, *FAI* free androgen index.

(Table 2). Also, hypermethylation of *TNF* was correlated with downregulation of its transcript expression in women with PCOS (Table 2). However, *TF* which showed hypermethylation in NGS data was found to be hypomethylated upon pyrosequencing, though its transcript was downregulated in PCOS (Figs. 2 and 3). Also, there was no

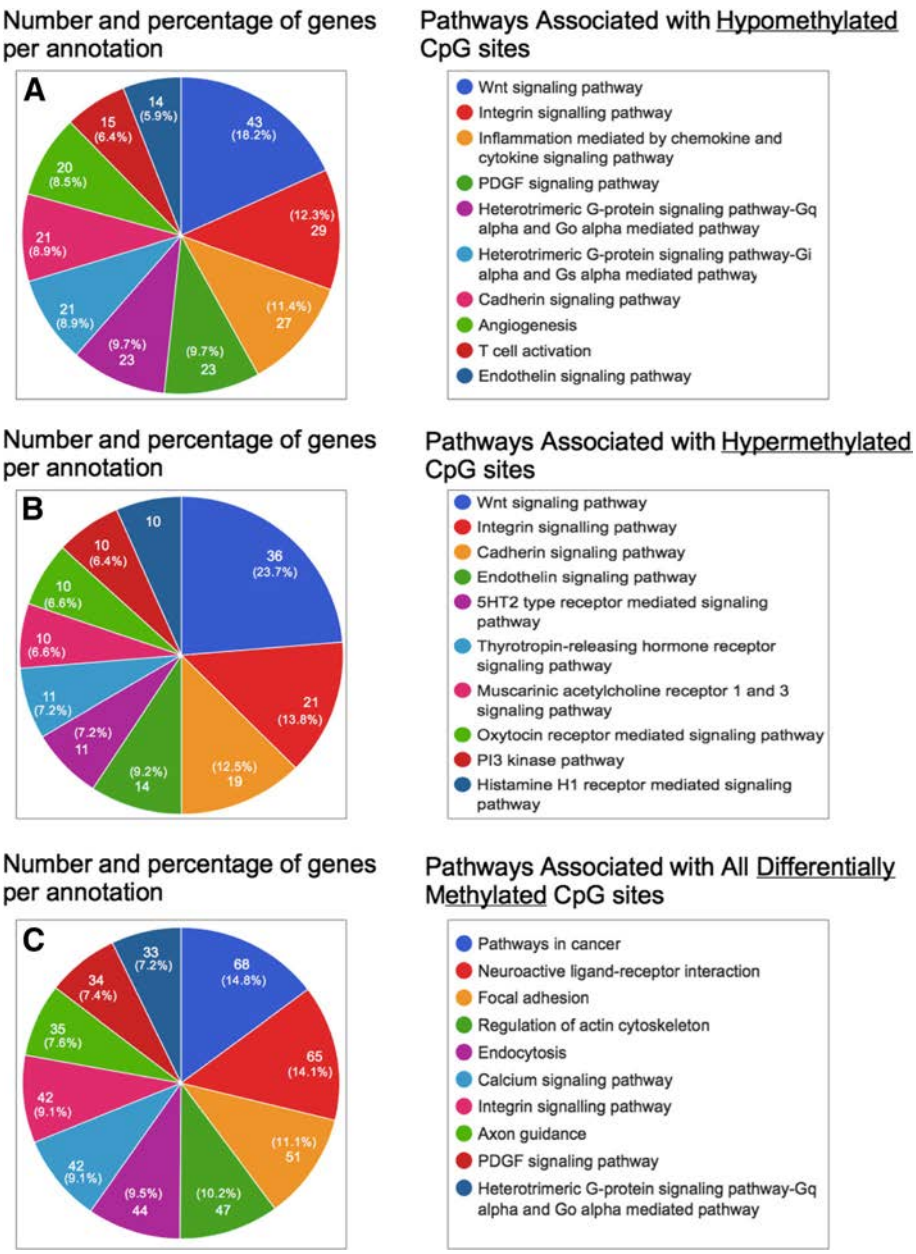


Fig. 1 Pathway analysis of differentially methylated genes identified by MC-Seq of cumulus granulosa cells (CGCs). Pie charts displaying gene ontology (GO) and pathway analyses for the lists of hypomethylated genes (**a**), hypermethylated genes (**b**), and all differentially methylated genes (DMGs) (**c**) between CGCs of controls (C-CC, $n = 3$) and women with PCOS (P-CC, $n = 3$) have been shown in the figure. Analysis was carried out using the GeneCodis3 web tool. Each chart represents the top ten pathways enriched in each of the three datasets

correlation between *TF* transcript levels and its methylation status (Table 2). As supporting evidence, we also evaluated the transcript expression of additional genes such as prostaglandin E receptor (*PTGER1*), leukemia inhibitory factor (*LIF*), and hyaluronan and proteoglycan link protein 1 (*HAPLN1*) which showed hypermethylation in NGS analysis. Transcripts of these genes were downregulated in PCOS (Fig. 2).

Discussion

Tissue-specific DNA methylation changes sired by alterations in the environmental or physiological milieu of an individual can bring about significant changes in gene and protein expression, and therefore predispose them to disease development. Alterations in both transcriptome and proteome profiles of ovarian cells/tissues and FF have been previously reported in PCOS [25–28]. With this

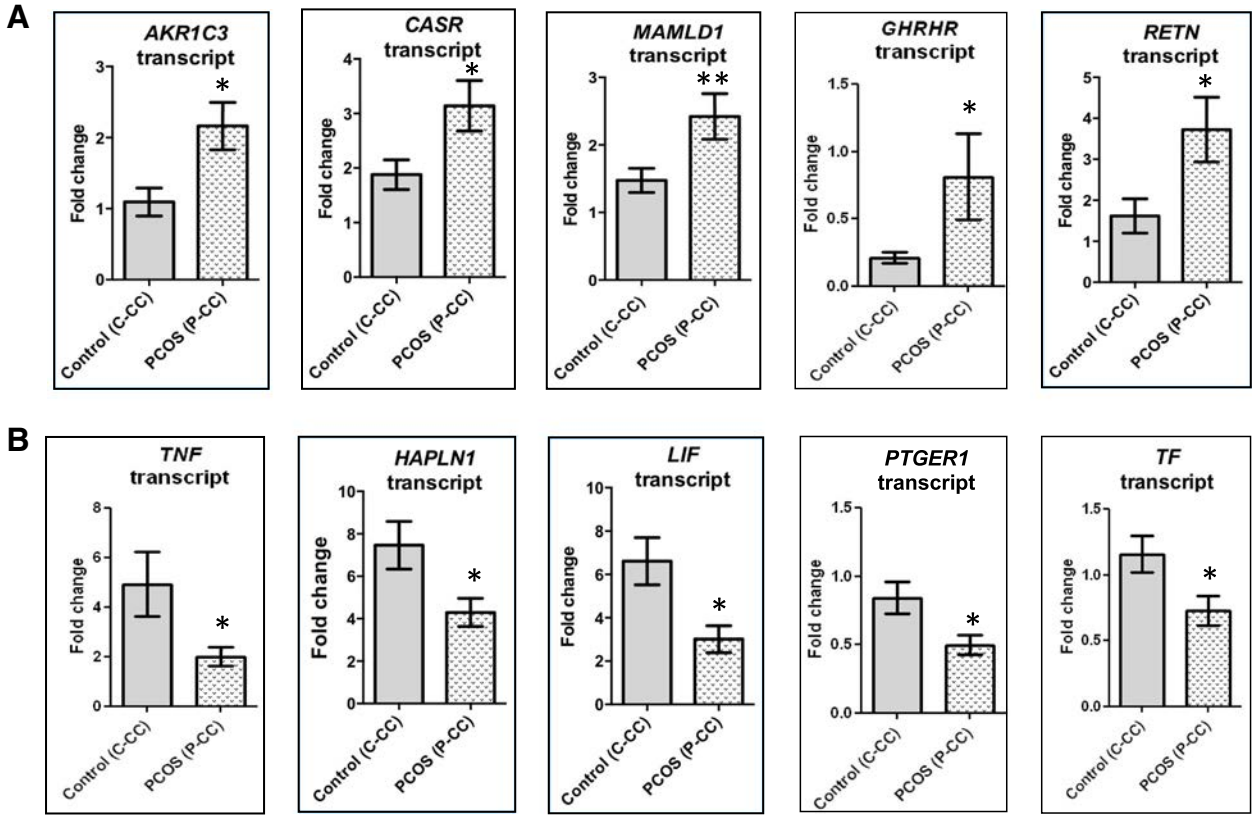


Fig. 2 Transcript expression profiling of few differentially methylated genes identified by MC-Seq. Bar graphs representing transcript expression profiles of 10 differentially methylated genes (DMGs) between CGCs (C-CCs) of controls ($n = 17$) and women with PCOS (P-CCs, $n = 17$) are shown in the figure. Panel (**a**) represents transcript expression data of genes identified as hypomethylated in PCOS by NGS analysis while (**b**) shows transcript profiling of hypermethylated genes. Fold change was evaluated using the $2^{-\Delta\Delta Ct}$ method, where the expression was normalized to 18s levels, using a CGC calibrator sample. Data are presented as "mean ± SEM." $^{*}P < 0.05$, $^{**}P < 0.01$, 'ns' denotes no significant change. Data are analyzed using the Mann-Whitney *U* test.

background information, we had initially conducted a pilot study to screen for the tissue-specific global DNA methylation changes in PBLs and CGCs of controls and women with PCOS. Here, subtle alterations were detected in CpG methylation profiles of long interspersed nucleotide element 1 (LINE1) in PCOS, and these changes were found to be more prominent in CGCs of women with PCOS, compared to PBLs [8]. Since LINE1s are self-replicating transposons and occupy ~ 17% of the human genome, even slight changes in their methylation patterns can be reflective of genomic dysregulation. Therefore, the primary goal of this study was to identify the genome-wide methylation differences in CGCs of women with PCOS, at a single base resolution. GO analysis of the current methylome data revealed that genes regulating cell growth, adhesion, differentiation, proliferation, cell polarity and fate determination, apoptosis, signal transduction, transcription, post-translational modifications, protein binding, metal and nonmetal ion binding, ATP binding, vesicular transport, etc., were differentially methylated in PCOS (Additional file 3). The implications of differential methylation observed in few of these identified genes, which may contribute to hyperandrogenism, defects in COC expansion, oocyte maturation/and ovulation, premature luteolysis, and oxidative stress observed in PCOS, have been discussed here.

Androgen overproduction

CpG hypomethylation in genes such as *AKR1C3*, *GHRHR*, *MAMLD1* and *RETN*, and hypermethylation in *TNF*, which can indirectly contribute to androgen excess, were consistent with increased and decreased levels of the respective gene transcripts (Figs. 2 and 3). AKR1C3 is a steroidogenic enzyme that converts androstenedione (A4) to biologically active testosterone in non-testicular tissues [29]. In PCOS, the increased expression of AKR1C3 and AKR1C3-mediated androgen production, have been reported in the adrenal cortex and visceral adipose tissues [29, 30]. This enzyme is also expressed by GCs (both mural and cumulus granulosa cells) of periovulatory follicles [31]. Since *AKR1C3* expression was found to be high in CGCs of women with PCOS, it may contribute to the high androgen production in

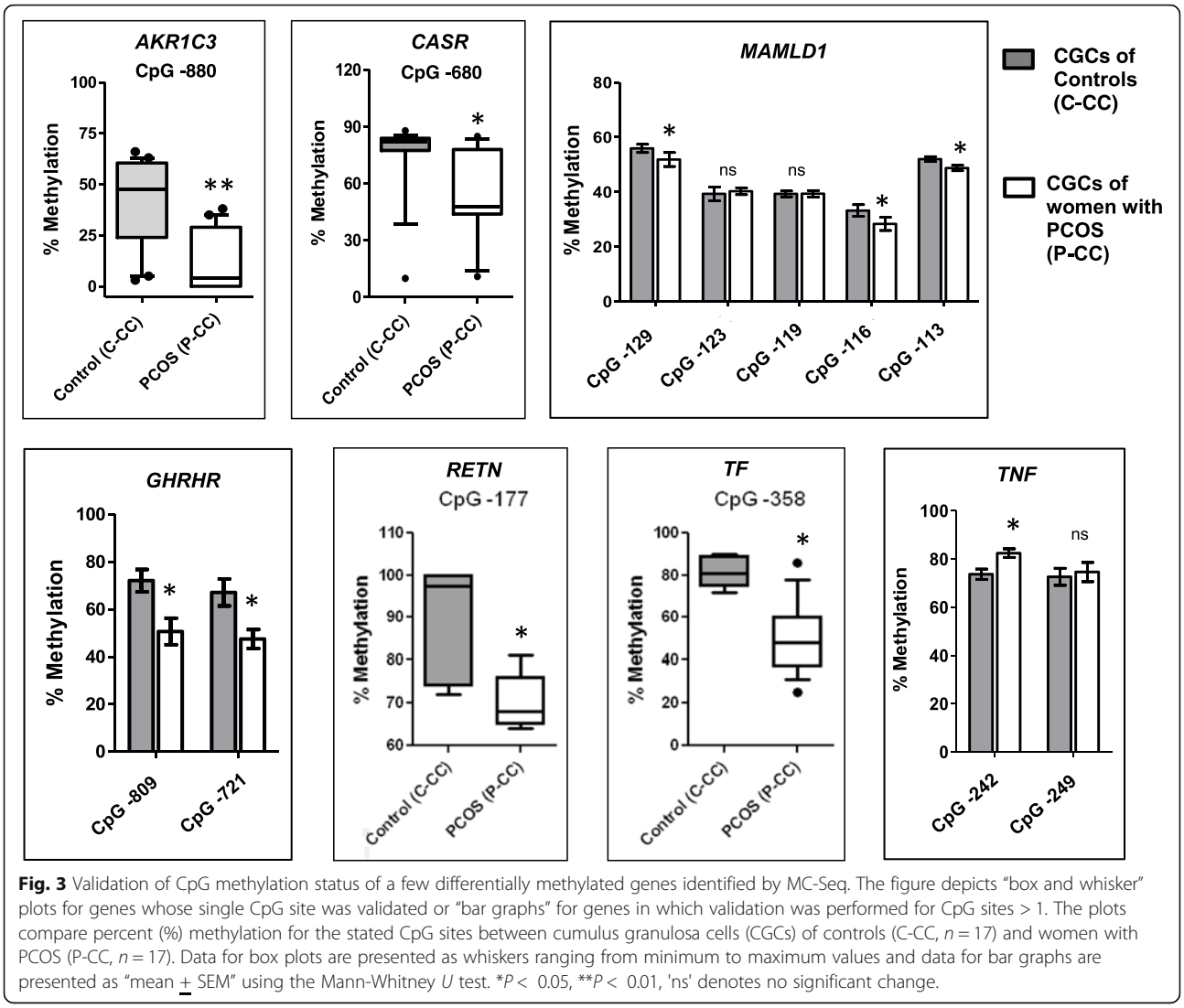


Fig. 3 Validation of CpG methylation status of a few differentially methylated genes identified by MC-Seq. The figure depicts "box and whisker" plots for genes whose single CpG site was validated or "bar graphs" for genes in which validation was performed for CpG sites > 1. The plots compare percent (%) methylation for the stated CpG sites between cumulus granulosa cells (CGCs) of controls (C-CC, $n = 17$) and women with PCOS (P-CC, $n = 17$). Data for box plots are presented as whiskers ranging from minimum to maximum values and data for bar graphs are presented as "mean ± SEM" using the Mann-Whitney *U* test. *$P < 0.05$, **$P < 0.01$, 'ns' denotes no significant change.

Table 2 Correlation analysis of CpG methylation levels of selected genes with their transcript expression profiles

Gene name	CpG site/s	*P* value/FDR cutoff	Methylation and transcript status by pyrosequencing and qPCR	Correlation coefficient (R^2); *P* value
AKR1C3	CpG-880	< 0.05	Hypomethylated upregulated	− 0.53; 0.02*
CASR	CpG-680	< 0.05	Hypomethylated upregulated	− 0.74; < 0.0001*
GHRHR	CpG-809	< 0.025	Hypomethylated upregulated	− 0.669; 0.002*
	CpG-721			− 0.501; 0.034
RETN	CpG-177	< 0.05	Hypomethylated upregulated	− 0.603; 0.029*
MAMLD1	CpG-129	< 0.016	Hypomethylated upregulated	− 0.759; < 0.0001*
	CpG-116			− 0.596; 0.007*
	CpG-113			− 0.612; 0.005*
TNF	CpG-242	< 0.025	Hypermethylated downregulated	− 0.582; 0.011*
TF	CpG-358	< 0.05	Hypomethylated downregulated	− 0.162; 0.507

CpG sites that showed significant difference between controls and PCOS in Fig. 3 were analyzed. *P* values < FDR cutoff values are marked as significant (*)

their ovaries. Next, *GHRHR* is a gene encoding the class B GPCR subfamily receptor, which regulates the release of somatotropin (GH) in the brain and other tissues including the ovary, via binding to the hypothalamic neuropeptide GHRH [32, 33]. Although the classical ovarian function of GH is to enhance sexual maturation at puberty via binding to its receptors (GHRs) [34], the GH-GHR interaction also stimulates the release of insulin-like growth factor (IGF1) via transcriptional activation [34]. IGF1, like insulin, can increase LH production from the pituitary and augment ovarian androgen synthesis in PCOS [35, 36]. Additionally, hyperinsulinemia in PCOS is known to increase the bioavailability of IGF1 via the downregulation of its carrier protein, i.e., IGFBP1 [25, 34]. High levels of GH and IGF1 also increase the sensitivity of developing follicles to gonadotropins [37, 38], and PCOS follicles have been reported to exhibit increased sensitivity and responsiveness to FSH [39]. As a result, the expression of LHCGR which is under the direct control of FSH is found to be elevated in the follicles of women with PCOS [13]. This can further augment LH-mediated ovarian androgen production in PCOS. Thus, hypomethylation and overexpression of *GHRHR* observed in CGCs can be an indirect mediator of androgen excess in PCOS (Fig. 4). Further, the pro-inflammatory cytokine TNF, which suppresses FSH-induced *LHCGR* promoter activation via NF-κB p65, was hypermethylated and low in PCOS [40], thus also making it an additional factor contributing to hyperandrogenemia.

Mutations in MAMLD1, a transcriptional coactivator, have been reported to result in compromised androgen synthesis during male fetal sexual development [41, 42]. However, information regarding a definite role of MAMLD1, its regulation and mechanisms of action in the context of other reproductive functions, is limited. In a murine study, *MAMLD1* knockout male mice showed a reduced testosterone production in Leydig tumor cells while the activation of *MAMLD1* promoter by the transcription factor, SF1, augmented testosterone production via transactivation of the hairy/enhancer of split 3 *HES3* promoter [43]. We therefore propose that *MAMLD1* hypomethylation and upregulation of its transcript in CGCs may be important in contributing to androgen excess in PCOS ovaries (Fig. 4). Next, increased circulatory levels of the adipokine, RETN, which modulates glucose tolerance and insulin action, have been linked to a higher incidence of insulin resistance and PCOS [44, 45]. However, there is some ambiguity regarding the role of RETN in PCOS, since its serum and FF levels have been either found to be high or unchanged in PCOS [46–48]. In theca cells, the dose-dependent increase in RETN showed augmented androgen production [49], while a RETN-like molecule β impaired the glucose tolerance and insulin actions in HEK293T cells and adipocytes [50]. Thus, the overexpression of *RETN* due to the hypomethylation of its promoter may be an important factor contributing to the androgen excess in PCOS (Fig. 4).

Oocyte development, ovulation, and COC matrix expansion defects

Enrichment of calcium (Ca^{2+}) signaling pathway and pathways for regulation of cytoskeletal and focal adhesion elements in our NGS analysis indicated impairment of calcium homeostasis and cellular architecture in CGCs of women with PCOS (Fig. 1, Additional file 3). Ca^{2+} signaling pathways are crucial for the development and maturation of healthy oocytes and CASR, which is an important mediator of this pathway, responds to subtle changes in extracellular Ca^{2+} concentrations and activates or ameliorates the mobilization of intracellular Ca^{2+} stored in tissues [51]. Expression and localization of CASR have been reported in human oocytes and CGCs, wherein it supposedly facilitates bidirectional communication between these cells to either keep oocytes arrested in MI phase, or assist in the full resumption of their cytoplasmic and nuclear maturation upon entering the MII phase [52]. Therefore, alterations in CASR expression may affect oocyte maturation and yield poor quality oocytes as seen in PCOS (Fig. 4). In PCOS, so far, a single report exists on the association of a CASR polymorphism (Hin1I) with altered global calcium homeostasis [53]. Altered methylation has been previously reported in human *CASR* promoter in a few cancer conditions [54–56]. Our results indicate that altered *CASR* expression in CGCs of PCOS women can be also influenced by altered methylation.

TNF expression in CGCs has been reported to be either unchanged or reduced in CGCs of women with PCOS [57, 58]; however, its circulating levels have been found to be high in their serum and FF [59, 60]. Few studies demonstrated that treatment with high levels of TNF increased GC apoptosis, impaired P_4 production from GCs, and caused other steroidogenic defects in these cells [61–63]. However, these studies utilized TNF at 10–20-fold higher doses relative to its physiological levels. In alternate studies, optimal TNF levels have been reported to impart a protective function in the maintenance of bovine GCs and oocytes [64], facilitate ovulation [65], and increase the GC proliferation in animal models [66]. Decreased levels of endogenous TNF in GCs has been attributed to diminished oocyte competence due to a reduction in its downstream effector, i.e. tumor necrosis factor-inducible gene 6 (TNFAIP6) [25], as well as compromised ovulation, and GC proliferation in ovarian follicles [67]. Thus, lowered *TNF* in CGCs of women with PCOS owing to hypermethylation, may hamper

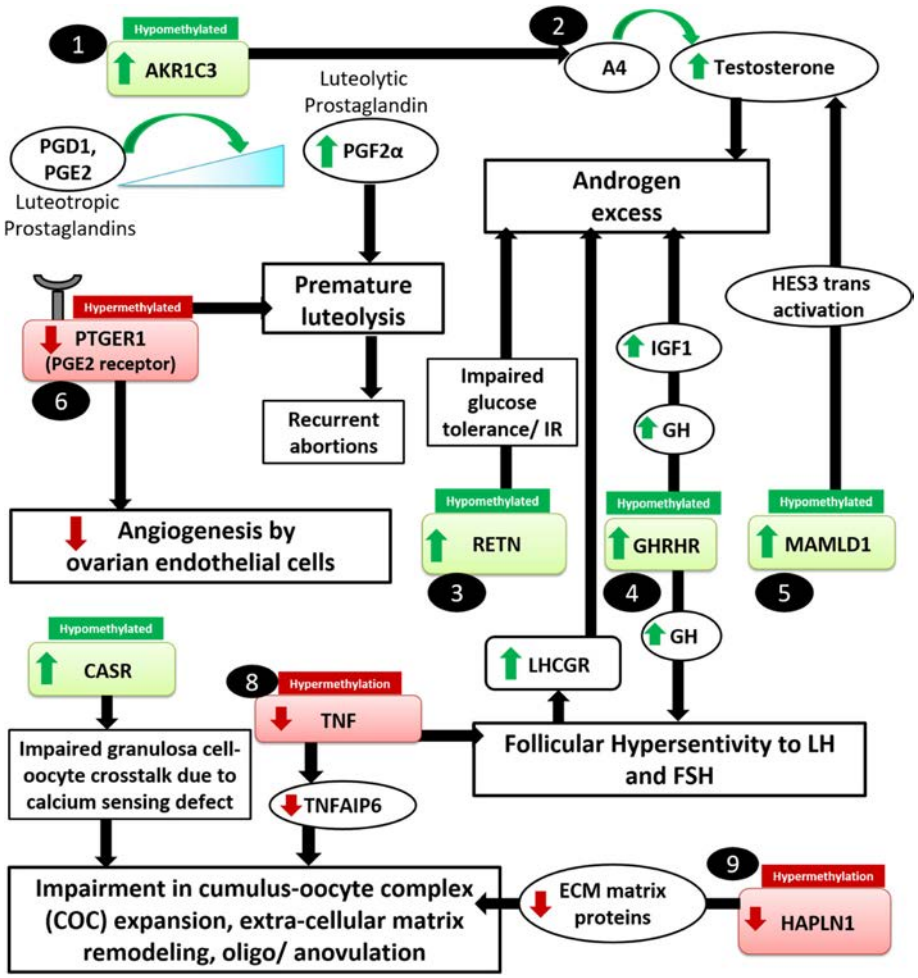


Fig. 4 Altered gene methylation can contribute to ovarian dysfuncion. The figure summarizes some of the processes that could be dysregulated in the follicular compartment of women, due to hypomethylation (indicated by green boxes) or hypermethylation (indicated by red boxes) of genes in cumulus granulosa cells (CGCs). AKR1C3, aldo-keto reductase family 1 member C3; AR, androgen receptor; CASR, calcium-sensing receptor; COC, cumulus-oophorus complex; FF, follicular fluid; FSH, follicle-stimulating hormone; FSHR, follicle-stimulating hormone receptor; GC, granulosa cells; IGF1; insulin-like growth factor 1; INS, insulin; MGC, mural granulosa cells; GHRHR, growth hormone-releasing hormone receptor; HAPLN1, hyaluronan and proteoglycan link protein 1; LH, luteinizing hormone; LHCGR, luteinizing hormone chorionic gonadotropin receptor; LIF, leukemia inhibitory factor; MAMLD1, mastermind-like domain containing 1; PTGER1, prostaglandin receptor E1; RETN, resistin; TC, theca cells; TF, transferrin; TNF, tumor necrosis factor alpha

COC expansion and compromise ovulation. LIF, which was hypermethylated and downregulated in our study, has been reported to be low in FF and serum of women with PCOS [68, 69]. LIF has demonstrated embryotrophic effects in mice and humans [70, 71], and decrease in its levels has shown a positive association with low implantation rates and poor IVF outcome [68, 69]. Since dose-dependent administration of LIF also showed induction of COC expansion and improvement in oocyte competence in humans, it is imperative to investigate the role of epigenetics in the regulation of *LIF* expression in women with PCOS. Our earlier data on proteomics of FF in women with PCOS demonstrated the downregulation of COC matrix proteins including amphiregulin, TNFAIP6, and bikunin, whose diminished expression is implicated in COC matrix expansion defects [25]. Supporting these findings, we also observed hypermethylation and downregulation of *HAPLN1*, which is also a COC matrix-associated protein. HAPLN1 facilitates the expansion of COC matrix and imparts stability to the COC complex by binding to other matrix proteins and proteoglycans like hyaluronic acid, versican, aggrecan, and IαI [72]. Exogenous treatment with HAPLN1 increased the CGC viability in vitro and enabled their transformation into granulosa lutein cells, while knockdown of HAPLN1 decreased the cell viability [72]. Therefore, altered *HAPLN1* methylation may contribute to changes in COC expansion dynamics in PCOS (Fig. 4).

Luteal insufficiency/premature luteolysis

Apart from being an androgen synthesizing enzyme, AKR1C3 also acts as a prostaglandin F synthase (PGFS), which catalyzes the conversion of the luteotrophic prostaglandins, PGD and PGE_2, to a luteolytic form, i.e., PGF2α to facilitate luteolysis [29]. Luteal insufficiency and premature luteolysis are frequent occurrences in PCOS [73], and these have been linked to aberrations in angiogenic mechanisms at the ovarian level [74]. Angiogenesis is largely under the control of prostaglandins [75, 76], and AKR1C family of enzymes are some major regulators of prostaglandins [77]. Disparities in levels of pro and anti-angiogenic factors in the ovary are largely responsible for defects in follicle development, premature degeneration of oocytes, and regression of CL due to an inefficient supply of oxygen and nutrients to the growing follicles [74]. Both, our present data and FF proteome study provide compelling evidence supporting angiogenic dysregulation in the ovarian compartment [25]. Since AKR1C3 has been implicated as a trigger for premature luteolysis [78] and was found to be hypomethylated and high in PCOS follicles, it may serve as an important epigenetic target to investigate this phenomenon in PCOS, Till now, no clear evidence exists on whether PGE_2 or PGF2α levels are altered in PCOS. However, our data shows that the receptor for PGE_2 gene, i.e., *PTGER1*, was hypermethylated in NGS and its transcript was low in CGCs of women with PCOS. Since activation of PTGER1 by specific agonists has demonstrated increased sprout formation in capillaries, both *in vivo* and *in vitro* [79, 80], it can be an important factor for the restoration of follicular angiogenesis. Also, PTGER1 has been shown to stimulate progesterone biosynthesis in human GCs [81]. Since the follicles of PCOS women lack optimum levels of progesterone required for maintenance of CL, low levels of *PTGER1* in PCOS caused by promoter hypermethylation may explain this shortcoming.

Oxidative stress

Lastly, *TF*, which maintains the oxido-reductive homeostasis in proliferating cells and showed hypermethylation in our NGS data, was found to be hypomethylated by pyrosequencing, though its mRNA was downregulated in PCOS. Therefore, methylation changes at other CpG sites in *TF* promoter need to be analyzed. TF primarily transports iron released from hepatic, intestinal, and reticuloendothelial stores to the target tissues via its receptor endocytosis while also alleviating local oxidative stress, acting as a growth factor, and promoting follicle and oocyte maturation [82, 83]. Low levels of TF in follicular fluid of PCOS women have been previously reported [84], which may lead to a higher prevalence of unbound Fe^{2+}; trigger oxidative damage to DNA, lipids, and proteins; and contribute to oxidative stress, which is reported in PCOS ovaries [85]. Figure 5 depicts a summary of biological processes which may be affected in PCOS due to the altered methylation of genes in follicles of women with PCOS.

Limitations

The primary limitation of this study was the low sample size. A number of controls and women with PCOS originally recruited for the study had to be excluded due to the presence of confounding factors such as hyperprolactinemia or thyroid dysfunction, and contamination of FF with blood, or due to recent treatment with metformin or thyroid medications. Further, women having low CGC counts were also excluded as the quantities of nucleic acids were insufficient for methylation and expression analyses. Also, since PCOS is a heterogeneous disorder, investigation of epigenetic changes in a large population based on phenotypic subgrouping of women with PCOS (A, B, C, and D phenotypes) as per the Rotterdam consensus may provide more accurate information on the effect of epigenetic components on PCOS development. Due to our limited sample size, it was not possible to carry out such subgrouping and subsequent analyses.

Our findings highlight that MC-Seq could identify several functionally important loci in CGCs, many of which were either known to be functionally dysregulated in PCOS ovary or have a compelling potential to be established as novel candidates influenced by epigenetic changes. The molecular sequelae of these alterations leading to ovarian dysfunction need to be further addressed via robust functional studies in women with PCOS.

Materials and methods

Study design, participants, sample collection, and estimated parameters

This study was carried out at the ICMR-National Institute for Research in Reproductive Health (NIRRH) as per ethical norms. All participants were recruited from the "Fertility Clinic and IVF Center" (Mumbai) after obtaining written informed consents and underwent IVF using a long, GnRH agonist protocol as reported earlier [8]. We initially recruited 35 women with PCOS and 38 age-BMI-matched, healthy, and regularly menstruating controls as per the Rotterdam consensus criteria [86]. Women showing normal ovarian morphology on ultrasound with no signs of hyperandrogenism or insulin resistance, and undergoing COH strictly owing to indications of male factor infertility in their spouses were recruited as controls. Women recruited as PCOS had at least 2 of the following 3 features, i.e., polycystic ovaries (PCO) on ultrasound, irregularity/absence of menses, and/or signs of hyperandrogenism during clinical

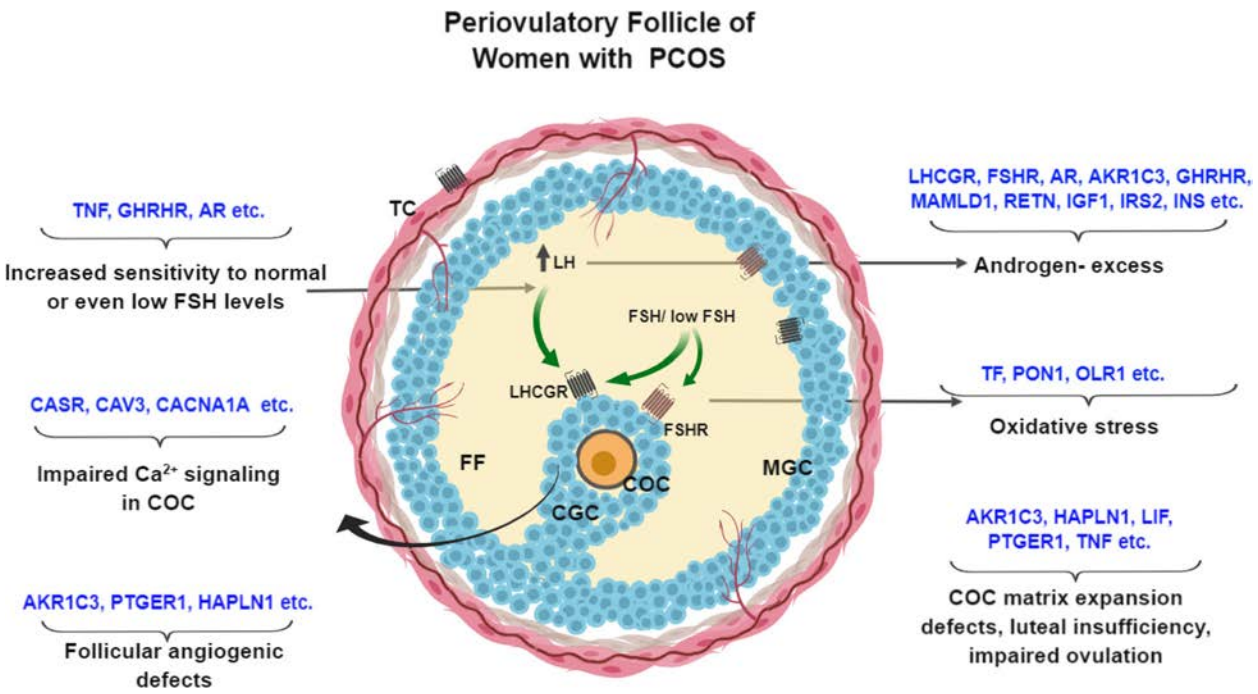


Fig. 5 Summary of the genes and their associated processes found to be affected in PCOS due to altered DNA methylation. The figure represents sa preovulatory follicle of women with PCOS, showing a few differentially methylated genes involved in the perpetuation of androgen excess, premature luteolysis, impaired calcium signaling, COC defects, oxidative stress and angiogenic defects in the follicles of women with PCOS as observed in our study. AKR1C3, aldo-keto reductase family 1 member C3; CASR, calcium-sensing receptor; COC, cumulus-oophorus complex; FF, follicular fluid; FSH, follicle-stimulating hormone; FSHR, follicle-stimulating hormone receptor; GC, granulosa cells; IGF1; insulin-like growth factor 1; INS, insulin; MGC, mural granulosa cells; GHRHR, growth hormone-releasing hormone receptor; HAPLN1, hyaluronan and proteoglycan link protein 1; LH, luteinizing hormone; LHCGR, luteinizing hormone chorionic gonadotropin receptor; LIF, leukemia inhibitory factor; MAMLD1, mastermind-like domain containing 1; PTGER1, prostaglandin receptor E1; RETN, resistin; TC, theca cells; TF, transferrin; TNF, tumor necrosis factor alpha

screening, however the presence of PCO morphology was used as a mandate for all women recruited as PCOS. These women were carefully screened for their baseline hormonal estimates and biochemical characteristics using serum and follicular fluid samples (Table 1). Further, we had to exclude samples of women who were on metformin or thyroid medications, whose follicular fluid had blood contamination and whose CGC counts were low, since substantial amounts of DNA were required for high-throughput methylome sequencing and bisulfite PCRs. After this, 20 controls and 20 women with PCOS were selected for the study (Table 1). Baseline hormonal estimates (between days 3–7 of menstrual cycle) for LH, FSH, prolactin, TSH, and AMH could be obtained from IVF clinical records, while fasting serum and macroscopically clear FF collected on d-OPU were assayed for E_2, P_4, TT, and SHBG using commercial ELISA kits (Diagnostics Biochem Canada Inc., Dorchester, Ontario, Canada). Androgen excess indices were calculated using TT and SHBG values [8]. Upon follicle maturation, the levels of E_2 and P_4, which are routinely measured 1 day prior to and 1 day after rhCG administration (10,000 IU) to monitor ovarian response, were also recorded. COCs suspended in FF aspirates were separated from FF and manually stripped off to dissociate the cumulus granulosa cells (CGCs) from their oocytes. CGCs were washed, resuspended in ovum buffer, and transported from IVF center to the lab at 37 °C for further processing. The numbers of total retrieved oocytes and mature oocytes (in the MII phase of meiosis) were obtained from clinical records, and rates of fertilization of MII oocytes were calculated.

Methyl-capture sequencing (MC-Seq) by NGS approach and analysis of DMRs

From among the 20 women with PCOS and 20 controls selected for the study, individual DNA samples of CGCs from 3 PCOS women and 3 age-BMI-matched controls having a total yield of > 2 μg were subjected to MC-Seq using the NGS approach. Whole genomic DNA was extracted from CGCs of these women using QIAamp DNA mini kit (Qiagen, Hilden, Germany) and processed for library preparation. 1 μg of genomic DNA per sample was sheared in fragments of approximately 150 bp using the Covaris S220 ultrasonicator and further processed for DNA end repair, 3′-adenylation, methylated adapter

ligation, and hybridization to the Methyl-Seq library probes using the SureSelect Methyl-Seq Library Prep Kit (Agilent Technologies, CA, USA) as per the manufacturer's instructions. The hybrids were captured using streptavidin beads (Dynabeads) and subjected to bisulfite conversion using the EZ-DNA Methylation Gold Kit (Zymo Research, CA, USA). The library was PCR amplified, indexed using SureSelectXT Methyl-Seq indexing primers (Agilent Technologies, CA, USA), and pooled for 100 bp paired-end multiplexed sequencing on Illumina HiSeq 2500 platform.

Processing and alignment of reads

Whole-genome targeted methylation sequencing reads of the above six samples were aligned to human reference genome hg19 assembly. The hg19 reference genome was converted to a DNA methylation reference genome, and genome indexing was performed using Bismark genome preparation utility (v.14.3). The adapter trimmed sequencing reads were aligned to the converted methylation reference genome using Bismark tool with two allowed mismatches. The remaining parameters from Bismark were used as default. Further, deduplication of aligned reads was performed using "deduplicate_bismark" utility of Bismark tool.

Methylation analysis and functional annotation

Methylation extractor utility of Bismark tool was used to extract the methylation call for every methylated cytosine (C) in all three contexts: CpG, CHG, and CHH (where H is A or C or T). Differential methylation analysis was performed using "calculateDiffMeth" function of R-based methylKit package at q value cutoff threshold of ≤ 0.01 and methylation difference ≥ 25%. Hierarchical clustering of samples based on the similarity of their methylation profile was performed using the euclidean distance metric, and ward method clustering approaches of methylKit package. Further, enrichment and annotation of both hypo- and hypermethylated sites within up 10 kb of annotated transcription start site (TSS) to transcription end site (UCSC hg19) were performed using in-house perl scripts. GeneCodis3 web-based tool [87] was used for gene ontology (GO) and pathway enrichment analysis of the DMGs.

Selection of genes for validation studies

Genes showing differential methylation at CpG sites upstream up to 1000 bases relative to their TSS in our NGS data were selected for validation. These included a total of 354 hypermethylated and 397 hypomethylated genes (total $n = 735$ unique genes). The lists of these 735 genes, and the 4354 genes enlisted in Ovarian Kaleidoscope (OKdb), were compared using a Venn analysis, and a total of 132 genes that were common to both datasets could be identified. OKdb is an online search tool based on microarray-based transcriptome profiling and independent study reports on genes and proteins identified in ovarian tissues and cells [88]. Upon identification of genes common to our NGS dataset and OKdb, and a careful review of literature, 7 genes participating in the perpetration of androgen excess, impaired angiogenesis, luteal insufficiency, COC matrix defects, and oocyte defects, namely *AKR1C3, CASR, GHRHR, MAMLD1, RETN, TF,* and *TNF,* were selected for validation by both pyrosequencing and qPCR. Additionally, the transcript expression profiles of prostaglandin E receptor (*PTGER1)*, leukemia inhibitory factor *(LIF)*, and hyaluronan and proteoglycan link protein 1 (*HAPLN1*), were evaluated in CGCs of controls and PCOS women as supporting evidence for pathways associated with the above genes.

Extraction of nucleic acids from CGCs for qPCR and pyrosequencing

Total DNA and RNA were extracted from CGCs of 17 controls and 17 women with PCOS using the NucleoSpin TriPrep kit (Macherey-Nagel, Düren, Germany) for validation of the NGS data. The quality and yield of nucleic acids were assessed by agarose gel electrophoresis and by evaluating their spectrophotometric ratios at 260 and 280 nm. DNA samples were stored at − 20 °C while RNA was stored at − 80 °C until further use.

cDNA synthesis and quantitative real-time PCR

cDNA was synthesized from 500 ng of RNA ($n = 34$) using a first strand cDNA synthesis kit (Takara Bio USA Inc.). Transcript levels of DMGs selected for validation (Additional file 4) were assayed by TaqMan chemistry using the TaqMan™ Universal Master Mix II with UNG, and FAM-labeled probes (ThermoFisher Scientific, MA, USA). Assay containing VIC-labeled 18s rRNA probe was used as the housekeeping control. qPCR was carried out using cDNA dilutions ranging between neat to 1:100. Fold change in gene expression between controls and PCOS was evaluated using the $2^{-\Delta\Delta Ct}$ method, where the expression was normalized to 18s levels, using a CGC calibrator sample.

Bisulfite primer design and pyrosequencing

Primers for the validation of selected genes were designed on the Pyromark Q96 ID machine using the PyroMark® Assay Design SW 2.0 software (Qiagen), the list of which has been provided in Additional file 4. Primers were procured from Sigma-Aldrich. The reverse pyrosequencing primers were tagged with biotin at the 5′-end and HPLC purified. Approximately 300–500 ng DNA from CGCs of 17 controls and 17 women with PCOS was bisulfite converted using the MethylCode bisulfite conversion kit (Invitrogen-ThermoFisher

Scientific, MA, USA). The region of interest was amplified by two rounds of PCR. PCR product (1–3 μL from total volume of 20 μL) from the first round was used as template for the 2nd PCR round, which was scaled up to 45 μL. For both PCRs, initial denaturation was performed at 95 °C for 15 min followed by 40 rounds of amplification at 94 °C for 30s, annealing at the respective optimized temperatures for 10s, and 72 °C for 60s with a final extension at 72 °C for 10 min. Product amplification after both rounds was confirmed by agarose gel electrophoresis. For pyrosequencing, 40 μL of the PCR product was subjected to clean up on the Pyromark Q96ID sequencing workstation as per the manufacturer's instructions and sequenced using 1.6 μL (16 picomoles) of each of the gene-specific sequencing primers.

Statistical analysis

Mann-Whitney *U* tests were employed for all univariate assessments of continuous variables including hormonal and biochemical parameters, DNA methylation, and transcript expression levels assessed between CGCs of controls and women with PCOS. Correlation between the CpG methylation status of genes and their respective transcripts was determined using a two-tailed Spearman's correlation coefficient. FDR cutoffs were applied for genes showing multiple CpG sites as significant after Mann-Whitney testing.

Additional files

Additional file 1: Hypomethylated CpG sites were representative of 2063 genes. (XLSX 408 kb)

Additional file 2: Hypermethylated sites were linked to a total of 1777 genes (XLSX 354 kb)

Additional File 3: Figure includes horizontal bar charts showing components such as A) molecular functions, B) biological processes, and C) cellular components that were most highly enriched in datasets obtained for all differentially methylated, hypomethylated, and hypermethylated genes identified in the NGS analysis. *X*-axis represents the number of genes present within each annotated category. (TIF 901 kb)

Additional File 4: The table enlists pyrosequencing primer sets designed to evaluate upstream/promoter CpG methylation in differentially methylated genes (DMGs) selected for validation in controls (C-CC, $n = 17$) and women with PCOS (P-CC, $n = 17$) and commercial TaqMan Assay IDs used for validating the transcript expression levels of selected DMGs in controls and women with PCOS. (DOCX 13 kb)

Abbreviations

A4: Androstenedione; AKR1C3: Aldo-keto reductase family 1 member C3; AMH: Anti-Mullerian hormone; AMHR2: Anti-Mullerian hormone receptor 2; AR: Androgen receptor; CASR: Calcium-sensing receptor; C-CC: Cumulus granulosa cells of controls; CGC: Cumulus granulosa cell; CL: Corpus luteum; COC: Cumulus-oophorus complex; COH: Controlled ovarian hyperstimulation; DMG: Differentially methylated gene; E_2: Estradiol; EDC: Endocrine disruptor; FAI: Free androgen index; FF: Follicular fluid; FGF: Fibroblast growth factor; FSH: Follicle-stimulating hormone; FSHR: Follicle-stimulating hormone receptor; FST: Follistatin; GC: Mural and cumulus granulosa cells; GH: Growth hormone; GHR: Growth hormone receptor; GHRHR: Growth hormone-releasing hormone receptor; GnRH: Gonadotropin-releasing hormone; GO: Gene ontology; H6PD: Hexose-6-phosphate dehydrogenase/glucose 1-dehydrogenase; HAPLN1: Hyaluronan and proteoglycan link protein 1; hCG: Human chorionic gonadotropin; IGF1R: Insulin-like growth factor receptor; IGFBP: Insulin-like growth factor binding proteins; INS: Insulin; IRS-2: Insulin receptor substrate 2; IVF: In vitro fertilization; LH: Luteinizing hormone; LHCGR: Luteinizing hormone chorionic gonadotropin receptor; LIF: Leukemia inhibitory factor; LINE1: Long interspersed nucleotide element 1; MAMLD1: Mastermind-like domain containing 1; MC-Seq: Methyl-capture sequencing; MeDIP: Methylated DNA immunoprecipitation; MGC: Mural granulosa cells; NGS: Next generation sequencing; OKdb: Ovarian Kaleidoscope; P_4: Progesterone; PBL: Peripheral blood leukocytes; P-CC: Cumulus granulosa cells of PCOS; PCOS: Polycystic ovary syndrome; PDGF: Platelet-derived growth factor; PGD: Prostaglandin D; PGE2: Prostaglandin E2; PGF: Prostaglandin F; PGFS: Prostaglandin F synthase; PRKCB: Protein kinase C Beta; PTGER1: Prostaglandin receptor E1; RETN: Resistin; SHBG: Sex hormone binding globulin; T2DM: Type 2 diabetes mellitus; TES: Transcription end site; TF: Transferrin; TNF: Tumor necrosis factor alpha; TNFAIP6: Tumor necrosis factor-inducible gene 6; TSH: Thyroid stimulating hormone (TSH); TSS: Transcription start site; TT: Total testosterone; VEGF: Vascular endothelial growth factor; WGBS: Whole genome bisulfite sequencing; YAP1: Yes-associated protein 1

Acknowledgements

The authors thank all the consenting study participants and collaborators who became a part of this study and Ms. Kavita Suri (Fertility Clinic and IVF Center, Mumbai) who collected the clinical samples of our study participants and provided us with their IVF data records.

Disclosure summary

The authors have nothing to disclose.

Funding

This work is partially supported by Department of Science and Technology (DST) (EMR/2015/001014), NIRRH (NIRRH/RA/623/04–2018) and Indian Council of Medical Research (ICMR), New Delhi, India. Financial assistance provided by DST and Department of Atomic Energy-Board of Research in Nuclear Sciences (DAE-BRNS), Government of India; to PS for pursuing her doctoral studies is acknowledged.

Availability of data and materials

Datasets analyzed during the current study will be made available from the corresponding author upon reasonable request. Supplementary data related to the methylome analysis of cumulus granulosa cells is available on Figshare, an online digital data repository https://doi.org/10.6084/m9.figshare.7370438

Authors' contributions

The study design was conceived by SM and PS. PS collected the patient samples, processed the samples, carried out all the experiments, performed data interpretation, and wrote the paper. PK carried out the bioinformatic analysis and interpretation of the NGS data. VM and SD are clinical collaborators on this manuscript who supervised the recruitment of study participants and collection of their samples at the Fertility Clinic and IVF Center, Mumbai. SM critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Written informed consent was obtained from all study participants who gave blood, follicular fluid and cumulus granulosa cell samples, in accordance with current edition of the NIRRH Ethics Committee for Clinical Studies (Ref: D/ICEC/Sci-54/56/2018 dated May 29, 2018). This committee is recognized by Strategic Initiative for Developing Capacity in Ethical Review (SIDCER) and Forum for Ethics Review Committees in Asia and the Western Pacific Region (FERCAP).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Molecular Endocrinology, ICMR-National Institute for Research in Reproductive Health, J.M. Street, Parel, Mumbai, Maharashtra 400012, India. [2]Colin Jamura Lab, Institute for Stem Cell Biology and Regenerative Medicine (inStem), National Centre for Biological Sciences (NCBS), GKVK, Bellary Road, Bangalore 560065, India. [3]Fertility Clinic and IVF Center, 12-Springfield, 19-Vachha Gandhi Road, Gamdevi, Mumbai, Maharashtra 400007, India.

Received: 14 January 2019 Accepted: 27 March 2019
Published online: 11 April 2019

References

1. Fauser BCJM, Tarlatzis BC, Rebar RW, Legro RS, Balen AH, Lobo R, et al. Consensus on women's health aspects of polycystic ovary syndrome (PCOS): the Amsterdam ESHRE/ASRM-Sponsored 3rd PCOS Consensus Workshop Group. Fertil Steril. 2012;97(1):28–38 e25.
2. Norman RJ, Dewailly D, Legro RS, Hickey TE. Polycystic ovary syndrome. Lancet. 2007;370(9588):685–97.
3. Diamanti-Kandarakis E. Polycystic ovarian syndrome: pathophysiology, molecular aspects and clinical implications. Expert Rev Mol Med. 2008;10:e3.
4. Baptiste CG, Battista M-C, Trottier A, Baillargeon J-P. Insulin and hyperandrogenism in women with polycystic ovary syndrome. J Steroid Biochem Mol Biol. 2010;122(1–3):42–52.
5. Wiweko B, Maidarti M, Priangga MD, Shafira N, Fernando D, Sumapraja K, et al. Anti-mullerian hormone as a diagnostic and prognostic tool for PCOS patients. J Assist Reprod Genet. 2014;31(10):1311–6.
6. Xu N, Azziz R, Goodarzi MO. Epigenetics in polycystic ovary syndrome: a pilot study of global DNA methylation. Fertil Steril. 2010;94(2):781–3 e1.
7. Sandovici I, Naumova AK, Greenwood CM, editors. Epigenetics and establishment of tissue-specific epigenetic states during development traits. Epigenetics complex trait; 2013. p. 35–62.
8. Sagvekar P, Mangoli V, Desai S, Patil A, Mukherjee S. LINE1 CpG-DNA hypomethylation in granulosa cells and blood leukocytes is associated with PCOS and related traits. J Clin Endocrinol Metab. 2017;102(4):1396–405.
9. Pruksananonda K, Wasinarom A, Sereepapong W, Sirayapiwat P, Rattanatanyong P, Mutirangura A. Epigenetic modification of long interspersed elements in cumulus cells of mature and immature oocytes from patients with polycystic ovary syndrome. Clin Exp Reprod Med. 2016; 43(2):82–9.
10. Le Jiang L, Xie JK, Cui JQ, Wei D, Yin BL, Zhang YN, et al. Promoter methylation of yes-associated protein (YAP1) gene in polycystic ovary syndrome. Med. 2017;96(2):e5768.
11. Sang Q, Zhang S, Zou S, Wang H, Feng R, Li Q, et al. Quantitative analysis of follistatin (FST) promoter methylation in peripheral blood of patients with polycystic ovary syndrome. Reprod BioMed Online. 2013;26(2):157–63.
12. Yu Y-Y, Sun C-X, Liu Y-K, Li Y, Wang L, Zhang W. Promoter methylation of *CYP19A1* gene in Chinese polycystic ovary syndrome patients. Gynecol Obstet Investig. 2013;76(4):209–13.
13. Wang P, Zhao H, Li T, Zhang W, Wu K, Li M, et al. Hypomethylation of the LH/choriogonadotropin receptor promoter region is a potential mechanism underlying susceptibility to polycystic ovary syndrome. Endocrinology. 2014; 155(4):1445–52.
14. Qu F, Wang F-F, Yin R, Ding G-L, El-Prince M, Gao Q, et al. A molecular mechanism underlying ovarian dysfunction of polycystic ovary syndrome: hyperandrogenism induces epigenetic alterations in the granulosa cells. J Mol Med. 2012;90(8):911–23.
15. Zhu JQ, Zhu L, Liang XW, Xing FQ, Schatten H, Sun QY. Demethylation of LHR in dehydroepiandrosterone-induced mouse model of polycystic ovary syndrome. Mol Hum Reprod. 2009;16(4):260–6.
16. Shen H-R, Qiu L-H, Zhang Z-Q, Qin Y-Y, Cao C, Di W. Genome-wide methylated DNA immunoprecipitation analysis of patients with polycystic ovary syndrome. PLoS One. 2013;8(5):e64801.
17. Li S, Zhu D, Duan H, Ren A, Glintborg D, Andersen M, et al. Differential DNA methylation patterns of polycystic ovarian syndrome in whole blood of Chinese women. Oncotarget. 2016;8(13):20656–66.
18. Yu Y-Y, Sun C-X, Liu Y-K, Li Y, Wang L, Zhang W. Genome-wide screen of ovary-specific DNA methylation in polycystic ovary syndrome. Fertil Steril. 2015;104(1):145–53 e6.
19. Wang X, Wei J, Jiao J, Jiang S, Yu D, Li D. Genome-wide DNA methylation and gene expression patterns provide insight into polycystic ovary syndrome development. Oncotarget. 2014;5(16):6603–10.
20. Kokosar M, Benrick A, Perfilyev A, Fornes R, Nilsson E, Maliqueo M, Johan Behre C, Sazonova A, Ohlsson C, Ling C, Stener-Victorina E. Epigenetic and transcriptional alterations in human adipose tissue of polycystic ovary syndrome. Sci Rep. 2016;6(22883).
21. Pan JX, Tan YJ, Wang FF, Hou NN, Xiang YQ, Zhang JY, et al. Aberrant expression and DNA methylation of lipid metabolism genes in PCOS: a new insight into its pathogenesis. Clin Epigenetics. 2018;10(1):6.
22. Xu J, Bao X, Peng Z, Wang L, Du L, Niu W, et al. Comprehensive analysis of genome-wide DNA methylation across human polycystic ovary syndrome ovary granulosa cell. Oncotarget. 2016;7(19):27899–909.
23. Ling Teh A, Pan H, Lin X, Lim YI, Pawan C, Patro K, et al. Comparison of methyl-capture sequencing vs. Infinium 450K methylation array for methylome analysis in clinical samples. Epigenetics. 2016;11(1):36–48.
24. Sagvekar P, Dadachanji R, Patil K, Mukherjee S. Pathomechanisms of polycystic ovary syndrome: multidimensional approaches. Front Biosci (Elite Ed). 2018;10:384–422.
25. Ambekar AS, Kelkar DS, Pinto SM, Sharma R, Hinduja I, Zaveri K, et al. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. J Clin Endocrinol Metab. 2015;100(2):744–53.
26. Choi DH, Lee WS, Won M, Park M, Park HO, Kim E, et al. The apolipoprotein A-I level is downregulated in the granulosa cells of patients with polycystic ovary syndrome and affects steroidogenesis. J Proteome Res. 2010;9(9): 4329–36.
27. Escobar-Morreale HF, Peral B, Corton M, Camafeita E, Lopez JA, San Millan JL, et al. Proteomic analysis of human omental adipose tissue in the polycystic ovary syndrome using two-dimensional difference gel electrophoresis and mass spectrometry. Hum Reprod. 2007;23(3):651–61.
28. Ma X, Fan L, Meng Y, Hou Z, Mao YD, Wang W, et al. Proteomic analysis of human ovaries from normal and polycystic ovarian syndrome. Mol Hum Reprod. 2007;13(8):527–35.
29. Nakamura Y, Hornsby PJ, Casson P, Morimoto R, Satoh F, Xing Y, et al. Type 5 17beta-hydroxysteroid dehydrogenase (AKR1C3) contributes to testosterone production in the adrenal reticularis. J Clin Endocrinol Metab. The Endocrine Society. 2009;94(6):2192–8.
30. O'Reilly MW, Kempegowda P, Walsh M, Taylor AE, Manolopoulos KN, Allwood JW, et al. AKR1C3-mediated adipose androgen generation drives lipotoxicity in women with polycystic ovary syndrome. J Clin Endocrinol Metab. The Endocrine Society. 2017;102(9):3327–39.
31. Dozier BL, Watanabe K, Duffy DM. Two pathways for prostaglandin F2 alpha synthesis by the primate periovulatory follicle. Reproduction. 2008;136(1): 53–63.
32. Matsoukas M-T, Spyroulias GA. Dynamic properties of the growth hormone releasing hormone receptor (GHRHR) and molecular determinants of GHRH binding. Mol BioSyst. The Royal Society of Chemistry. 2017;13(7):1313–22.
33. Barabutis N, Schally AV. Growth hormone-releasing hormone: extrapituitary effects in physiology and pathology. Cell Cycle. 2010;9(20):4110–6.
34. Sharara FI, Giudice LC. Role of growth hormone in ovarian physiology and onset of puberty. J Soc Gynecol Investig. 1997;4(1):2–7.
35. Bergh C, Carlsson B, Olsson J-H, Selleskog U, Hillensjö T. Regulation of androgen production in cultured human thecal cells by insulin-like growth factor I and insulin. Fertil Steril. 1993;59(2):323–31.
36. Poretsky L. The insulin-related ovarian regulatory system in health and disease. Endocr Rev Oxford University Press. 1999;20(4):535–82.
37. van Dessel HJHMT, Lee PDK, Faessen G, Fauser BCJM, Giudice LC. Elevated serum levels of free insulin-like growth factor I in polycystic ovary syndrome 1. J Clin Endocrinol Metab. 1999;84(9):3030–5.
38. Bachelot A, Monget P, Imbert-BOLLORÉ P, Coshigano K, Kopchick JJ, Kelly PA, et al. Growth hormone is required for ovarian follicular growth. Endocrinology. 2002;143(10):4104–12.
39. Franks S, Stark J, Hardy K. Follicle dynamics and anovulation in polycystic ovary syndrome. Hum Reprod Update. 2008;14(4):367–78.
40. Nakao K, Kishi H, Imai F, Suwa H, Hirakawa T, Minegishi T. TNF-α suppressed FSH-induced LH receptor expression through transcriptional regulation in rat granulosa cells. Endocrinology. 2015;156(9):3192–202.
41. Kalfa N, Fukami M, Philibert P, Audran F, Pienkowski C, Weill J, et al. Screening of mamld1 mutations in 70 children with 46,xy dsd: identification and functional analysis of two new mutations. PLoS One. 2012;7(3):e32505.

42. Kalfa N, Cassorla F, Audran F, Oulad Abdennabi I, Philibert P, Béroud C, et al. Polymorphisms of MAMLD1 gene in hypospadias. J Pediatr Urol. 2011;7(6):585–91.
43. Fukami M, Wada Y, Okada M, Kato F, Katsumata N, Baba T, et al. Mastermind-like domain-containing 1 (MAMLD1 or CXorf6) transactivates the Hes3 promoter, augments testosterone production, and contains the SF1 target sequence. J Biol Chem. 2008;283(9):5525–32.
44. Çapoğlu İ, Erdem F, Uyanık A, Turhan H. Serum levels of resistin and hsCRP in women with PCOS. Open Med SP Versita. 2009;4(4):428–32.
45. Benomar Y, Gertler A, De Lacy P, Crépin D, Hamouda HO, Riffault L, et al. Central resistin overexposure induces insulin resistance through toll-like receptor 4. Diabetes. 2013;62(1):102–44.
46. Seow KM, Juan CC, Hsu YP, Ho LT, Wang YY, Hwang JL. Serum and follicular resistin levels in women with polycystic ovarian syndrome during IVF-stimulated cycles. Hum Reprod. 2005;20(1):117–21.
47. Panidis D, Koliakos G, Kourtis A, Farmakiotis D, Mouslech T, Rousso D. Serum resistin levels in women with polycystic ovary syndrome. Fertil Steril. 2004; 81(2):361–6.
48. Dinka P, Baldani L, Skrgatic M, Kasum G, Zlopasa S, Kralik O, et al. Altered leptin, adiponectin, resistin and ghrelin secretion may represent an intrinsic polycystic ovary syndrome abnormality. Gynecol Endocrinol. 2019:1–5. In press
49. Munir I, Yen HW, Baruth T, Tarkowski R, Azziz R, Magoffin DA, et al. Resistin stimulation of 17α-hydroxylase activity in ovarian theca cells in vitro: relevance to polycystic ovary syndrome. J Clin Endocrinol Metab. 2005;90(8):4852–7.
50. Kushiyama A, Shojima N, Ogihara T, Inukai K, Sakoda H, Fujishiro M, et al. Resistin-like molecule β activates MAPKs, suppresses insulin signaling in hepatocytes, and induces diabetes, hyperlipidemia, and fatty liver in transgenic mice on a high fat diet. J Biol Chem. 2005;280(51):42016–25.
51. Hofer AM, Brown EM. Extracellular calcium sensing and signalling. Nat Rev Mol Cell Biol. 2003;4(7):530–8.
52. Elena Dell'Aquila M, De Santis T, Cho YS, Reshkin SJ, Caroli AM, Maritato F, et al. Localization and quantitative expression of the calcium-sensing receptor protein in human oocytes. Fertil Steril. 2006;85:1240–7.
53. Ranjzad F, Mahban A, Shemirani AI, Mahmoudi T, Vahedi M, Nikzamir A, et al. Influence of gene variants related to calcium homeostasis on biochemical parameters of women with polycystic ovary syndrome. J Assist Reprod Genet. 2011;28(3):225–32.
54. Hizaki K, Yamamoto H, Taniguchi H, Adachi Y, Nakazawa M, Tanuma T, et al. Epigenetic inactivation of calcium-sensing receptor in colorectal carcinogenesis. Mod Pathol. 2011;24(6):876–84.
55. Fetahu IS, Höbaus J, Aggarwal A, Hummel DM, Tennakoon S, Mesteri I, et al. Calcium-sensing receptor silencing in colorectal cancer is associated with promoter hypermethylation and loss of acetylation on histone 3. Int J Cancer. 2014;135(9):2014–23.
56. Casalà C, Gil-Guiñón E, Luis Ordóñez J, Miguel-Queralt S, Rodríguez E, Galván P, et al. The calcium-sensing receptor is silenced by genetic and epigenetic mechanisms in unfavorable neuroblastomas and its reactivation induces ERK1/2-dependent apoptosis. Carcinogenesis. 2013;34(2):268–76.
57. Lee JY, Tae JC, Kim CH, Hwang D, Kim KC, Suh CS, et al. Expression of the genes for peroxisome proliferator-activated receptor-γ, cyclooxygenase-2, and proinflammatory cytokines in granulosa cells from women with polycystic ovary syndrome. Clin Exp Reprod Med. 2017;44(3):146.
58. Schmidt J, Weijdegård B, Mikkelsen AL, Lindenberg S, Nilsson L, Brännström M. Differential expression of inflammation-related genes in the ovarian stroma and granulosa cells of PCOS women. Mol Hum Reprod. 2014;20(1):49–58.
59. Amato G, Conte M, Mazziotti G, Lalli E, Vitolo G, Tucker AT, et al. Serum and follicular fluid cytokines in polycystic ovary syndrome during stimulated cycles. Obstet Gynecol. 2003;101(6):1177–82.
60. Gao L, Gu Y, Yin X. High serum tumor necrosis factor-alpha levels in women with polycystic ovary syndrome: a meta-analysis. Atkin SL, editor PLoS One. 2016;11(10):e0164021.
61. Yamamoto Y, Kuwahara A, Taniguchi Y, Yamasaki M, Tanaka Y, Mukai Y, et al. Tumor necrosis factor alpha inhibits ovulation and induces granulosa cell death in rat ovaries. Reprod Med Biol. 2015;14(3):107–15.
62. Sakumoto R, Shibaya M, Okuda K. Tumor necrosis factor-alpha (TNF alpha) inhibits progesterone and estradiol-17beta production from cultured granulosa cells: presence of TNFalpha receptors in bovine granulosa and theca cells. J Reprod Dev. 2003;49(6):441–9.
63. Morrison LJ, Marcinkiewicz JL. Tumor necrosis factor α enhances oocyte/follicle apoptosis in the neonatal rat ovary1. Biol Reprod. 2002;66(2):450–7.
64. Paulino LRFM, Cunha EV, Barbalho Silva AW, Souza GB, Lopes EPF, Donato MAM, et al. Effects of tumour necrosis factor-alpha and interleukin-1 beta on in vitro development of bovine secondary follicles. Reprod Domest Anim. 2018;53(4):997–1005.
65. Crespo D, Goetz FW, Planas JV. Luteinizing hormone induces ovulation via tumor necrosis factor α-dependent increases in prostaglandin F2α in a nonmammalian vertebrate. Sci Rep. 2015;5:14210.
66. Son D-S, Arai KY, Roby KF, Terranova PF. Tumor necrosis factor α (TNF) increases granulosa cell proliferation: dependence on c-Jun and TNF receptor type 1. Endocrinology. 2004;145(3):1218–26.
67. Yuan Y, Ida JM, Paczkowski M, Krisher RL. Identification of developmental competence-related genes in mature porcine oocytes. Mol Reprod Dev. 2011;78(8):565–75.
68. Li Z, Zhu Y, Li H, Jiang W, Liu H, Yan J, et al. Leukaemia inhibitory factor in serum and follicular fluid of women with polycystic ovary syndrome and its correlation with IVF outcome. Reprod BioMed Online. 2018;36(4):483–9.
69. Lédée-Bataille N, Laprée-Delage G, Taupin JL, Dubanchet S, Taieb J, Moreau JF, et al. Follicular fluid concentration of leukaemia inhibitory factor is decreased among women with polycystic ovarian syndrome during assisted reproduction cycles. Hum Reprod. 2001;16(10):2073–8.
70. Cheung LP, Leung HY, Bongso A. Effect of supplementation of leukemia inhibitory factor and epidermal growth factor on murine embryonic development in vitro, implantation, and outcome of offspring. Fertil Steril. 2003;80(SUPPL. 2):727–35.
71. De Matos DG, Miller K, Scott R, Tran CA, Kagan D, Nataraja SG, et al. Leukemia inhibitory factor induces cumulus expansion in immature human and mouse oocytes and improves mouse two-cell rate and delivery rates when it is present during mouse in vitro oocyte maturation. Fertil Steril. 2008;90(6):2367–75.
72. Liu J, Park E-S, Curry TE, Jo M, Jo M. Periovulatory expression of hyaluronan and proteoglycan link protein 1 (Hapln1) in the rat ovary: hormonal regulation and potential function. Mol Endocrinol The Endocrine Society. 2010;24(6):1203–17.
73. Shah D, Patel S. Polycystic ovarian syndrome as a cause of recurrent pregnancy loss. J Obs Gynecol India. 2007;57(5):391–7.
74. Di Pietro M, Pascuali N, Parborell F, Abramovich D. Ovarian angiogenesis in polycystic ovary syndrome. Reproduction. 2018;155(5):R199–209.
75. Karnezis T, Shayan R, Fox S, Achen MG, Stacker SA. The connection between lymphangiogenic signalling and prostaglandin biology: a missing link in the metastatic pathway. Oncotarget. 2012;3(8):893–906.
76. Mantel A, Carpenter-Mendini AB, Vanbuskirk JB, De Benedetto A, Beck LA, Pentland AP. Aldo-keto reductase 1C3 is expressed in differentiated human epidermis, affects keratinocyte differentiation, and is upregulated in atopic dermatitis. J Invest Dermatol. 2012;132(4):1103–10.
77. Rižner TL, Penning TM. Role of aldo-keto reductase family 1 (AKR1) enzymes in human steroid metabolism. Steroids. 2014;79:49–63.
78. Nio-Kobayashi J, Kudo M, Sakuragi N, Iwanaga T, Duncan WC. Loss of luteotropic prostaglandin E plays an important role in the regulation of luteolysis in women. MHR Basic Sci Reprod Med. 2017;23(5):271–81.
79. Trau HA, Davis JS, Duffy DM. Angiogenesis in the primate ovulatory follicle is stimulated by luteinizing hormone via prostaglandin E2. Biol Reprod. 2015;92(1):15.
80. Trau HA, Brännström M, Curry TE, Duffy DM, Duffy DM. Prostaglandin E2 and vascular endothelial growth factor A mediate angiogenesis of human ovarian follicular endothelial cells. Hum Reprod. 2016;31(2):436–44.
81. Chandras C, Harris TE, López Bernal A, Abayasekara DRE, Michael AE. PTGER1 and PTGER2 receptors mediate regulation of progesterone synthesis and type 1 11β-hydroxysteroid dehydrogenase activity by prostaglandin E2in human granulosa-lutein cells. J Endocrinol. 2007;194(3):595–602.
82. Aleshire SL, Osteen KG, Maxson WS, Entman SS, Bradley CA, Parl FF. Localization of transferrin and its receptor in ovarian follicular cells: morphologic studies in relation to follicular development. Vol. 51. Fertil Steril. 1989;51(3):444–9.
83. Gomme PT, McCann KB. Transferrin: structure, function and potential therapeutic actions. Elsevier Current Trends. 2005;10(4):267–73.
84. Wu Y-T, Wu Y, Zhang J-Y, Hou N-N, Liu A-X, Pan J-X, et al. Preliminary proteomic analysis on the alterations in follicular fluid proteins from women undergoing natural cycles or controlled ovarian hyperstimulation. J Assist Reprod Genet. 2015;32(3):417–27.
85. Chattopadhayay R, Ganesh A, Samanta J, Jana SK, Chakravarty BN, Chaudhury K. Effect of follicular fluid oxidative stress on meiotic spindle

formation in infertile women with polycystic ovarian syndrome. Gynecol Obstet Investig. 2010;69(3):197–202.
86. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril. 2004;81(1):19–25.
87. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res. 2012;40(W1):W478–83.
88. Leo CP, Vitt UA, Hsueh AJW. The ovarian kaleidoscope database: an online resource for the ovarian research community. Endocrinology. 2000;141(9): 3052–4.

OPEN ACCESS Freely available online

PLoS one

Gene Bionetwork Analysis of Ovarian Primordial Follicle Development

Eric E. Nilsson[1], Marina I. Savenkova[1], Ryan Schindler[1], Bin Zhang[2], Eric E. Schadt[2], Michael K. Skinner[1]*

1 Center for Reproductive Biology, School of Molecular Biosciences, Washington State University, Pullman, Washington, United States of America, **2** Sage Bionetworks, Seattle, Washington, United States of America

Abstract

Ovarian primordial follicles are critical for female reproduction and comprise a finite pool of gametes arrested in development. A systems biology approach was used to identify regulatory gene networks essential for primordial follicle development. Transcriptional responses to eight different growth factors known to influence primordial follicles were used to construct a bionetwork of regulatory genes involved in rat primordial follicle development. Over 1,500 genes were found to be regulated by the various growth factors and a network analysis identified critical gene modules involved in a number of signaling pathways and cellular processes. A set of 55 genes was identified as potential critical regulators of these gene modules, and a sub-network associated with development was determined. Within the network two previously identified regulatory genes were confirmed (i.e., Pdgfa and Fgfr2) and a new factor was identified, connective tissue growth factor (CTGF). CTGF was tested in ovarian organ cultures and found to stimulate primordial follicle development. Therefore, the relevant gene network associated with primordial follicle development was validated and the critical genes and pathways involved in this process were identified. This is one of the first applications of network analysis to a normal developmental process. These observations provide insights into potential therapeutic targets for preventing ovarian disease and promoting female reproduction.

Citation: Nilsson EE, Savenkova MI, Schindler R, Zhang B, Schadt EE, et al. (2010) Gene Bionetwork Analysis of Ovarian Primordial Follicle Development. PLoS ONE 5(7): e11637. doi:10.1371/journal.pone.0011637

Editor: Hugh Clarke, McGill University, Canada

Received March 18, 2010; **Accepted** June 11, 2010; **Published** July 16, 2010

Funding: National Institutes of Health (NIH) R01 HD043093-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: skinner@wsu.edu.edu

Introduction

An emerging concern in the field of biomedical research is that the common reductionist approach to studying biological processes may not be adequate to fully understand the complex interplay of cellular signaling, gene expression, and other complex molecular processes that occur within a tissue or organ. Examples of reductionist studies that have driven much of our understanding of biological processes associated with complex phenotypes like disease include the knockout mouse experiments and *in vitro* cytokine treatment to assess the effects of gene-specific perturbations on cell or tissue biology. Results from these types of studies provide information on candidate regulatory factors, but typically do not elucidate the network of factors or processes required for a normal developmental biology or pathobiology. A holistic, systems biology, approach to studying normal developmental processes can be a powerful tool that is complementary to the more reductionist experiments. In the spirit of a systems-based approach to development, the current study was designed to identify gene networks involved in ovarian primordial follicle development and to characterize critical regulatory factors involved in this development process.

In mammals, all the oocytes (eggs) that will be used over a female's lifetime are present in the ovary at birth in a finite pool. These oocytes are arrested in prophase of the first meiotic division and are each surrounded by flattened pre-granulosa cells to form a structure called a primordial follicle [1]. During the reproductive lifespan of a female, follicles gradually leave the arrested pool to undergo a primordial to primary follicle transition. A follicle undergoing follicle transition has an increase in oocyte diameter and the associated granulosa cells proliferate and change from a flattened to cuboidal in shape. Once primordial to primary follicle transition has occurred the follicle either continues to develop to the point of ovulation or undergoes atresia [1,2,3,4]. Previously cell-to-cell communication with extra-cellular growth factors has been shown to regulate the initiation of primordial follicle development. These studies have primarily used a reductionist approach to test candidate growth factors one at a time for their ability to affect follicle transition. A number of paracrine growth factors have been identified as having a role in early follicle development (reviews [4,5]).

To move beyond examining single gene effects on this development process, gene network analysis can be employed to identify groups (e.g. modules) of genes whose expression is regulated in a coordinated manner (gene network) [6,7,8]. In this type of analysis, a biological system is surveyed in the context of disease (or other interesting phenotypes) with microarrays multiple times with and without perturbations that cause the system to change. A novel bioinformatics analysis is used to identify modules of genes associated with biological systems (bionetwork). The great majority of network analyses have focused on disease states and been used to better understand the systems biology of disease processes and identify potential therapeutic targets [9,10,11, 12,13,14,15]. The current study was designed to determine if

network analysis can be applied to study a normal development process.

The current study used whole rat ovaries cultured *in vitro* in a manner that allowed primordial to primary follicle transition. The ovaries were treated with one of eight different growth factors previously shown to regulate primordial follicle transition in comparison to untreated control cultures. The mRNA was isolated from the ovaries and used for microarray transcriptome analysis to globally survey gene expression under these different treatment conditions. The effects of each growth factor on gene expression were analyzed to determine similarities and differences in gene expression between the different growth factor treatments. Those genes whose mRNA expression changed with any treatment were subjected to network analysis to identify pathways and genes with a high degree of connectivity between other genes and pathways. From the networks constructed from these data we identified a list of critical modules of regulated genes forming gene sub-networks that were used to identify regulatory genes involved in primordial follicle development. Not only were previously identified regulatory factors/genes associated with this process identified from this network analysis, but a number of putative regulators of follicle transition not previously associated with this process were also determined. One of the new candidate genes, connective tissue growth factor (Ctgf) [16], was tested experimentally and found to promote primordial to primary follicle transition. Observations demonstrate the utility of this network analysis to be used as a systems biology approach to study normal developmental processes in complex systems.

Number of Genes and Pathways Overlapped

	#PW	AMH	FGF2	BMP4	GDNF	FGF7	KITLG	LIF	PDGFab	CTGF
# genes		56	41	22	20	36	54	56	41	12
AMH	268		28	18	14	29	37	40	32	7
FGF2	248	10		13	11	19	27	29	17	7
BMP4	79	4	5		9	11	14	16	13	6
GDNF	148	14	7	3		9	9	11	11	3
FGF7	123	36	5	1	5		28	25	17	5
KITLG	271	8	5	3	1	2		39	24	8
LIF	349	7	18	13	4	3	18		30	9
PDGFab	275	18	22	3	14	5	7	10		6
CTGF	155	5	7	2	1	2	6	2	6	

Figure 1. Number of genes and pathways overlapped between signature (growth factor treatment group) lists. Total number of differentially expressed genes for each growth factor is shown in dark yellow column, number of genes overlapped between each pair of signature lists – in light yellow columns. Total number of KEGG pathways affected by each growth factor is shown in dark green row; number of KEGG pathways overlapped for each pair of growth factors is shown in light green row. CTGF analysis separate from the network analysis.
doi:10.1371/journal.pone.0011637.g001

Results

Primordial Follicle Transcriptome Analysis

A number of regulatory factors have been shown to affect primordial to primary follicle transition, including Amh [17,18], Fgf2 [19,20,21], Bmp4 [22,23], Gdnf [24], Fgf7/KGF [25], Kitlg [19,26,27], Lif [28] and Pdgfa [29]. In order to determine the underlying gene networks and processes involved in primordial follicle development, microarray analysis was performed on RNA from whole rat ovaries treated for two days in vitro with each of the above listed growth factors independently. There were three independent RNA samples of pooled ovaries for each growth factor treatment (except for GDNF, which had only two sample replicates), and corresponding control samples for a total of 38 RNA samples. These were evaluated using 38 Affymetrix Rat Gene 1.0 ST microarrays. The array data were analyzed together using normalization and pre-processing described in the Methods. Each growth factor treatment resulted in 79 to 349 genes with altered expression compared to controls (Figure 1). The lists of the genes affected by each treatment are presented in Table S1. There were relatively few genes with altered expression in common between the different treatments (Figure 1). Less than 10% of the genes changed by any one growth factor treatment were found to be changed in any other treatment. The exception was Fgf7/KGF, which had a more than 30% overlap of altered genes with Amh. There were no individual genes that changed expression levels in response to more than three of the eight original treatments (Table S1).

The complete list of genes whose expression levels changed with any of the treatments was compared to curated lists of genes from the KEGG database to identify processes that may be important for primordial follicle development. Automated unbiased matching of lists of affected genes to KEGG pathways was performed with Pathway Express (Intelligent Systems and Bioinformatics Laboratory; http://vortex.cs.wayne.edu/ontoexpress/). Pathways heavily impacted by genes whose expression altered in response to the growth factor treatments (Table 1) included pathways involved in cell surface and extracellular matrix regulation (cell adhesion molecules, adherens junction, focal adhesion, tight junction, gap junction, regulation of actin cytoskeleton), known signaling pathways (MAPK, notch, B-cell receptor, adipocytokine, toll-like receptor, ErbB, GnRH, Wnt, hedgehog, VEGF, Jak-STAT, TGF-beta, p53, insulin, PPAR), the complement cascade, axon guidance, glycan structure biosynthesis and pathways listing cell communication ligand-receptor interactions (cytokine-cytokine receptor, neuroactive ligand-receptor, ECM-receptor). There was a high degree of overlap of affected pathways between different growth factor treatments (Figure 1). For the list of pathways containing altered genes, from 70% to 82% of those pathways are shared with at least one other treatment. Application of the hypergeometrical Fisher Exact Test to assess whether the number of overlapped pathways was significantly greater than expected by chance, revealed that the majority were statistically significant. The pathways containing altered genes from several growth factor treatments are presented in Table 1. Although few altered genes were found to overlap between different treatments (Figure 1), each growth factor treatment influenced similar pathways, Table 1 and Table S1. Therefore, each growth factor affects similar pathways via different genes.

Bionetwork Analysis

The complete list of genes whose expression levels changed with any growth factor treatment was subjected to a network analysis as described in Methods. Potential batch effects for culture date, RNA processing data and microarray performance date were corrected during the analysis, with no major effect on the analysis. The data were fit using a robust linear regression model (rim function from R statistical package), and then the residuals with respect to the model fit were carried forward in all subsequent analysis. The network analysis scores each gene according to how

Table 1. Number of Regulated Genes in Pathways.

Pathway Name	Total List (1540)		Module Lists																Signature Lists								
	Total Overlap Number	Impact Factor	Short/Connected List	turquoise	blue	brown	yellow	green	red	black	pink	magenta	purple	green-yellow	tan	salmon	cyan	grey	AMH	bFGF	BMP4	GDNF	KGF	KL	LIF	PDGF	CTGF
Number of genes in the list			**55**	**194**	**182**	**158**	**150**	**139**	**112**	**99**	**85**	**68**	**45**	**32**	**29**	**28**	**22**	**157**	**268**	**248**	**79**	**148**	**123**	**271**	**349**	**275**	**155**
Neuroactive ligand-receptor interaction	**22**	**1.23**	1	2		5	2	1	3			1	1	2	1		1	3	4	5		1	4	5	4	6	3
Focal adhesion	**19**	**3.62**	3	7	1	7	1		1			1	1						3	3			4	8	3	1	
MAPK signaling pathway	**19**	**3.00**	6	7	1	3	4	1						1			1	1	6		4		5	4	6		
Cell adhesion molecules (CAMs)	**15**	**107**	1	1	3	2	1		1			3		1	1			2	6	2	3	2		3	2	2	1
Axon guidance	**15**	**6.34**	1	4	1	2	4	1						1	1				6		1	3	2	3	3	3	
Cytokine-cytokine receptor interaction	**15**	**2.24**	1	3	3	4	3			1								1	2	1	4		1	6	5		1
Calcium signaling pathway	**13**	**1.72**	1	3		5	1		2					1			1		3	2		1	2	3	4	1	
Regulation of actin cytoskeleton	**12**	**2.18**	2	5	1	4	2												4	2			5	4			
Cell Communication	**11**	**4.70**		2		4	3		1			1							3	1	2		1	4	1	1	
Jak-STAT signaling pathway	**11**	**1.88**		3	1	3	1			2				1					1	1			1	4	4		1
Adipocytokine signaling pathway	**10**	**2.94**		2			2	2			1		1	1		1			1	1	2				7	1	1
ECM-receptor interaction	**9**	**4.23**	2	2		4	2		1										1	2	1			5	2		
Complement and coagulation cascades	**9**	**3.63**		1	1	2	4							1						2	4			1	4		1
Apoptosis	**9**	**2.94**		1		2	2	1		2	1								2	1	2			2	5		
Toll-like receptor signaling pathway	**9**	**2.92**		3	1	1		1	1					1				1	2	1	1		1	2	1	1	
Leukocyte transendothelial migration	**9**	**2.74**		1	2	3	1					1						1	3	1				5	1		
Gap junction	**9**	**2.73**	1	2	1	2	2	1						1					2				1	3	1	3	
Insulin signaling pathway	**9**	**1.35**			1	1		1	2			1	2			1				2	1		1	4		2	1
Adherens junction	**8**	**33.1**		3		3						1						1	2					3	4	1	
Glycan structures - biosynthesis 2	**8**	**3.83**	1	2	2			1	1							1		1	1	3		1	3		1		
T cell receptor signaling pathway	**8**	**2.71**		2	1	1	1						1					2	3			1	2	2	1		
Tight junction	**7**	**3.28**		1		1	1	1	1			1						1		1				2	4		
GnRH signaling pathway	**6**	**2.84**	1	2		2											1	1	1				2	2	1		
TGF-beta signaling pathway	**6**	**1.40**		1	2						1	2							1		2	1	2		1	1	1

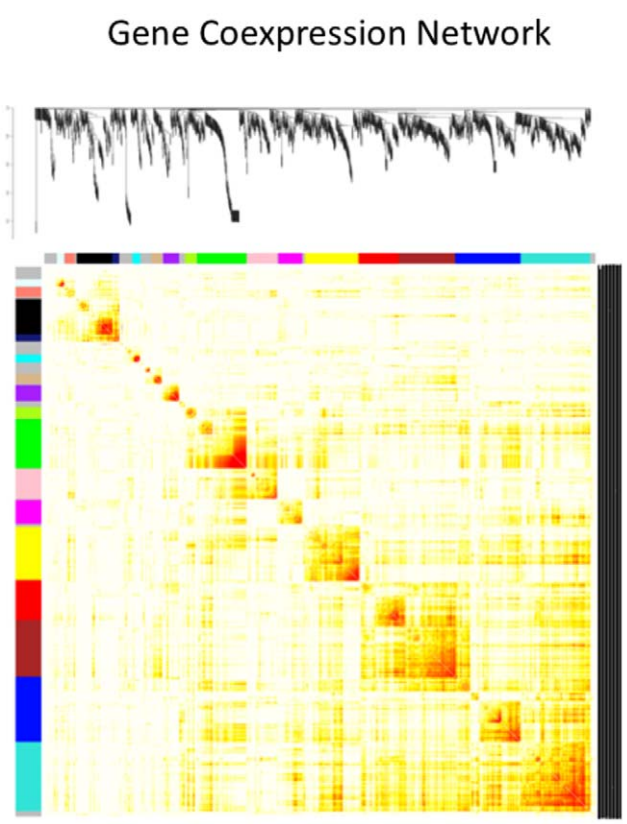
Table 1. Cont.

Pathway Name	Total List (1540)		Module Lists																Signature Lists								
	Total Overlap Number	Impact Factor	Short/Connected List	turquoise	blue	brown	yellow	green	red	black	pink	magenta	purple	green-yellow	tan	salmon	cyan	grey	AMH	bFGF	BMP4	GDNF	KGF	KL	LIF	PDGF	CTGF
Number of genes in the list			**55**	**194**	**182**	**158**	**150**	**139**	**112**	**99**	**85**	**68**	**45**	**32**	**29**	**28**	**22**	**157**	**268**	**248**	**79**	**148**	**123**	**271**	**349**	**275**	**155**
PPAR signaling pathway	**6**	**1.24**	1	1		1	1	1								1		1		1		1		1	2		
Antigen processing and presentation	**5**	**36.8**			1			1				2						1	1	1	1	1			1		
Melanogenesis	**5**	**3.34**	1	2		1	1				1								1	1				1	2	1	
B cell receptor signaling pathway	**5**	**2.95**		2	1	2													2	1		1		2			
Wnt signaling pathway	**5**	**2.84**		3			1								1				2	1				1	3		
Hematopoietic cell lineage	**5**	**1.28**		1	1		2											1	1	2	2		1		1		
ErbB signaling pathway	**4**	**2.89**				2	1											1	1				1	3			
Hedgehog signaling pathway	**4**	**2.38**		2		2													1	2					1	1	
VEGF signaling pathway	**4**	**2.17**		1		2							1						1					1	1	1	
p53 signaling pathway	**4**	**1.37**			1		1			1				1							1			1	2	1	
Glycan structures - biosynthesis 1	**4**	**0.85**		1	1				1										1	1	2	1	1				
Renin-angiotensin system	**3**	**3.21**															1	2		1				2			
Notch signaling pathway	**3**	**2.95**				1	1	1											1					1	1		1
mTOR signaling pathway	**3**	**2.82**				1						1	1							1				1	1	1	
Basal transcription factors	**3**	**2.03**							1		1					1				1		1	1				
Fc epsilon RI signaling pathway	**3**	**1.87**		1		1							1						1					2			
Nucleotide excision repair	**3**	**1.49**														1		2	2	1			1			1	
Cell cycle	**3**	**1.13**	1	1					1					1							1		1	1	1		
Ubiquitin mediated proteolysis	**3**	**0.98**	1	1		1					1									1			1	1			
Phosphatidylinositol signaling system	**2**	**28.0**						1										1							2		

Number of regulated genes in affected KEGG pathways for each of modules and signature (Growth Factor Treatment) lists.

Pathway enrichment analysis was performed using KEGG database, current for July 2009, for which gene symbols were used as input data. From 1,540 probe sets, 1,137 were identified as annotated genes using Affymetrix RaGene-1_0-st-v1 Transcript Cluster Annotations, release 28 (March, 2009). Top 44 pathways shown except for disease pathways (not shown). No KEGG pathways were affected by genes from light-cyan or midnight-blue modules.

doi:10.1371/journal.pone.0011637.t001



module	size
turquoise	194
blue	182
brown	158
yellow	150
green	139
red	112
black	99
pink	85
magenta	68
purple	45
greenyellow	32
tan	29
salmon	28
cyan	22
midnightblue	20
lightcyan	20

Figure 2. The ovary gene co-expression network and corresponding gene modules. A topological overlap matrix of the gene co-expression network consisting of the 1540 genes regulated by the various growth factors. Genes in the rows and columns are sorted by an agglomerative hierarchical clustering algorithm (see Methods). The different shades of color signify the strength of the connections between the nodes (from white signifying not significantly correlated to red signifying highly significantly correlated). The hierarchical clustering (top) and the topological overlap matrix strongly indicate highly interconnected subsets of genes (modules). Modules identified are colored along both column and row and are boxed. The number of genes in each module is listed as size of module.
doi:10.1371/journal.pone.0011637.g002

well, under different treatments, its changes in gene expression are correlated with the changes in expression of every other gene. This gives a connectivity score for each gene. High connectivity scores indicate that expression of this gene changes in concert with that of many other genes. In addition, the network analysis identifies gene modules in which the member genes have similar changes in expression in response to the various growth factor treatments. Gene modules are functional components of the network that are often associated with specific biological processes. To identify modules comprised of highly interconnected expression traits within the co-expression network, we examined the topological overlap matrix [30] associated with this network. The topological overlap between two genes not only reflects their more proximal interactions (e.g., two genes physically interacting or having correlated expression values), but also reflects the higher order interactions that these two genes may have with other genes in the network. Figure 2 depicts a hierarchically clustered topological overlap map in which the most highly interconnected modules in the network are readily identified. The specific details of the gene co-expression network analysis (Figure 2) are presented in the Methods section. To identify gene modules (sub-networks) formally from the topological overlap map, we employed a previously described dynamic cut-tree algorithm with near optimal performance on complicated dendrograms [31] (see Methods for details). Figure 2 shows the topological overlap map of the co-expression network with gene modules color-coded for the 16 modules identified. The membership of each module can be found in Table S1. The sixteen modules contained 1,383 genes with the remaining 157 genes (colored as gray) not failing into any module.

The pathways containing genes whose expression changed with growth factor treatment were compared to the genes from each module that were associated with specific pathways (Table 1 and Table S1 for full list). For most pathways genes from several network modules were present. However, several pathways were associated with selected modules. For example, out of 19 altered genes present in the focal adhesion pathway, seven were from the turquoise module and seven from the brown. Similarly, of the five altered genes in the Wnt signaling pathway three were from the turquoise module. For the fifteen changed genes in the cell adhesion molecule pathways three were from blue and three from magenta modules. Those genes whose expression changed with specific growth factor treatments were cross-matched with the genes assigned to each network module to determine if specific modules were heavily influenced by particular growth factors. Interestingly, each module was biased toward having many genes in common with selected growth factors (Table S1). In contrast, some growth factor treatments induced changes in genes that were distributed among several different modules.

In order to identify genes that could be key regulators of primordial follicle development, a shorter list of candidate genes was generated from the results of the network analysis. Six modules were chosen for having the highest numbers of known

regulatory genes and pathways (yellow, turquoise, blue, brown, red, purple). The top 10% of most connected transcripts in each of the six modules were identified as potential important regulators [13,15], except for the blue module for which the top 20% were chosen since so many of that module's most highly connected transcripts were not annotated as genes. The compiled list included 55 transcripts annotated as genes (Table 2), and these genes were subjected to more intensive investigation.

An automated unbiased analysis of published scientific literature was applied to the lists of differentially expressed genes described above using Genomatix/BiblioSphere software, as described in the Methods. Figure 3 shows a small integrated gene network among the short list of the 55 candidate regulators (Table 2). These relationships in literature raise the possibility that physiological interactions exist between these genes. The gene Ctgf (Connective tissue growth factor) was seen to relate to several other genes with high connectivity, Figure 3. Interestingly, two of the identified genes were previously shown to influence primordial follicle development, Pdgfa [29] and Fgfr2 [19,20,21].

The entire set of 1540 transcripts differentially expressed with growth factor treatment was also subjected to analysis using BiblioSphere. Only 632 were recognized by BiblioSphere and 613 were connected. A diagram of literature relationships between these genes is presented in Figure S1. Five major gene clusters were identified as associated with Nfkb1, Vegfa, Gadd45a, Esr1 and Egfr1. This analysis is useful to compare with the expression network analysis, but is biased toward the literature and finding relationships among more heavily studied factors.

Analysis of CTGF Actions

Critical regulatory candidates for primordial follicle development were selected due to their being differentially expressed in response to treatment with growth factors, having a high connectivity score, and being related in literature to other highly connected genes. For the purpose of the current study, candidate regulatory genes that were also extracellular growth factors were considered. Therefore, CTGF was selected based on all these criteria for further analysis. Experiments were performed to see if CTGF could regulate primordial to primary follicle transition. Ovaries from four-day old rats were treated with 50ng/ml CTGF protein for ten days in an organ culture system as described in the Methods (Figure 4A and 4B). Transforming growth factor beta 1 (TGFB1), which is known to interact with CTGF [32,33], was also tested. Untreated cultured ovaries were used as a negative control, and ovaries treated with 50ng/ml each of Kit ligand (KITL) and Fibroblast growth factor 2 (FGF2) were used as a positive control. CTGF treatment resulted in a significant ($p<0.05$) increase in developing follicles compared to untreated controls, as did treatment with the combination of KITL and FGF2 (Figure 4C). TGFB1 had no effect, either alone or in combination with CTGF.

RNA was collected from CTGF and control cultured ovaries as described in the Methods from three replicate experiments. The RNA was used for microarray analysis using the same criteria as for the other growth factors used in the network analysis. One hundred fifty-five transcripts were differentially expressed in CTGF-treated ovaries, Table S1. As was seen for the other growth factors used in the network analysis, there was little overlap of these changed genes with the genes showing changed expression in response to any other growth factor treatment, Figure 1. However, as seen among the other growth factors, there was a high degree of overlap between the pathways impacted by CTGF treatment and treatment with other growth factors (Figure 1). Therefore, a critical regulatory gene predicted from the network analysis was confirmed to regulate primordial follicle development.

Discussion

A systems biology approach was used to elucidate the changes in gene expression that are important for ovarian primordial to primary follicle transition. A gene network analysis was performed on the ovarian transcriptomes following treatment with 8 different growth factors. The rat ovary was used as a model system to test the utility of this approach in investigating a normal developmental process. This is one of the first applications of network analysis to a normal developmental process. The objective was to identify critical regulatory factors and pathways in primordial follicle development following a bionetwork analysis.

Microarray analysis determined the alterations in the ovarian transcriptome that occurred in response to treatment of ovaries with AMH, FGF2, BMP4, GDNF, FGF7, KITL, LIF, and PDGFB. All of these have previously been shown to effect follicle transition [17,18,19,20,21,22,23,24,25,26,27,28,29,34]. All these factors stimulate primordial follicle development except AMH that inhibits follicle development. The presence of both positive and negative factors provides a wider diversity of gene regulation to facilitate the network analysis. As expected the AMH regulated gene set is more distinct from the others. Surprisingly, there were few altered genes in common between all these growth factors and there were no genes that significantly changed in expression level in response to more than three of the eight growth factors. In contrast, the physiological processes impacted by these altered genes were found to have a higher level of overlap. Since a pathway includes groups of genes, it is expected that the overlap of pathways between growth factor treatments will be higher. The overlap of pathways was markedly high (70% to 82%) and statistically different, suggesting pathway associations provide a predicted capacity to identify regulatory factors. Certain pathways were significantly over-represented in the pool of genes with changed expression. This suggests that there are selected physiological pathways that are influenced by all the different growth factors (Figure 5), but that each growth factor affects different genes at different points in these pathways (Table 1). Multiple input points into these physiological pathways could allow for more precise regulation and more effective compensation between the growth factors. Since many growth factors are acting in parallel to regulate these pathways, any one pathway system is robust and maintains function if one growth factor becomes inoperative. Since primordial follicle development is essential for female reproduction, a complex network of regulatory factors influencing different aspects of critical signaling pathways has evolved.

For the eight growth factors evaluated the cellular processes affected in common (Figure 5) included changes in cell contact, morphogenesis, and cell proliferation and differentiation. These are processes that are necessary for the morphological changes that occur with primordial to primary follicle transition. During follicle transition granulosa cells change from squamous to cuboidal and the oocyte starts to grow in diameter (Figure 4B). Unexpectedly, what was also seen as an important affected cellular process was regulation of several key components of the complement and coagulation cascades (Figure S2). These genes are not known for having roles in ovary or follicle development, and merit further investigation.

Gene networks provide a convenient framework for exploring the context within which single genes operate. For gene networks associated with biological systems, the nodes in the network typically represent genes, and edges (links) between any two nodes indicate a relationship between the two corresponding genes. An important end product from the gene co-expression network analysis is a set of gene modules which member genes are more highly correlated with each other than with genes outside a module. It has been

Table 2. List of 55 Candidate Regulatory Genes.

Function	GeneSymbol	GeneBank	Probeset	k.in*	Module	Regulation by Growth Factor	GeneTitle
Apoptosis	Bcl2l10	NM_053733	10911690	10.0	red	KL-dwn	BCL2-like 10
Cell Cycle	Cdkn2c	NM_131902	10878705	7.6	red	KL-dwn	CDK4 inhibitor 2C
ECM	Cdh3	NC_005118	10807525	28.0	turq	AMH-dwn	cadherin 3, type 1, P-cadherin
	Col11a1	AJ005396	10818502	20.6	brown	KL-dwn	collagen, type XI, alpha 1
	Krt19	NM_199498	10747262	25.2	turq	AMH-dwn	keratin 19
	Lama5	NC_005102	10852270	30.9	turq	AMH-dwn	laminin, alpha 5
Development	Bnc2	NM_001106666	10877880	26.5	turq	AMH-dwn	basonuclin 2
	Emx2	NM_001109169	10716454	27.5	turq	KGF-dwn	empty spiracles homeobox 2
	Usmg5	NM_133544	10730633	13.5	blue	AMH, GDNF-dwn	muscle growth 5 homolog (mouse)
Epigenetics	Dnmt1	NM_053354	10915437	22.9	brown	KL-dwn	DNA (cytosine-5-)-methyltransferase 1
Golgi	B4galt6	NM_031740	10803394	11.7	blue	AMH-dwn, KGF-dwn	galactosyltransferase, polypeptide 6
Growth Factors	Ctgf	NM_022266	10717233	8.2	blue	KGF-up, LIF-dwn	connective tissue growth factor
	Il16	NM_001105749	10723351	31.9	turq	AMH-dwn, KGF-dwn	interleukin 16
	Pdgfa	NM_012801	10757129	27.0	turq	AMH-dwn, KGF-dwn	platelet-derived growth factor
Immune	LOC287167	NM_001013853	10741765	21.6	brown	KL-dwn	globin, alpha
Metabolism	Cacna2d3	NM_175595	10789819	19.9	brown	KL-dwn	calcium channel
	Hbq1	XM_001061675	10741761	24.0	brown	KL-dwn	hemoglobin, theta 1
	Hhatl	NM_001106868	10914424	15.7	yellow	PDGF-up	hedgehog acyltransferase-like
	Hmgcs2	NM_173094	10817759	19.5	brown	KL-dwn	Coenzyme A synthase 2
	Hsd11b1	NM_017080	10770795	17.0	yellow	BMP4-dwn	hydroxysteroid 11-beta dehydro
	Kirrel	NM_207606	10824123	14.1	yellow	AMH-dwn	kin of IRRE like (Drosophila)
	Plod2	NM_175869	10912255	18.3	brown	KL-dwn	procollagen lysine, 2-oxoglutarate
	Podxl	NM_138848	10861662	26.6	turq	AMH-dwn, KGF-dwn	podocalyxin-like
	Scn3a	NM_013119	10845809	17.4	brown	KL-dwn	sodium channel, type III, alpha
	Slc4a4	NM_053424	10775997	29.0	turq	AMH-dwn, KGF-dwn	solute carrier family 4
	Slc7a5	NM_017353	10811531	15.6	yellow	PDGF-up	solute carrier family 7
	Slc29a1	NM_031684	10921833	15.4	yellow	PDGF-up	solute carrier family 29
	Eno1	NM_012554	10874152	9.0	purple	PDGF-up	enolase 1, (alpha)
Receptors	Axl	NM_031794	10719900	15.4	yellow	BMP4-dwn	Axl receptor tyrosine kinase
	Ednrb	NM_017333	10785724	20.7	brown	KL-dwn	endothelin receptor type B
	Fgfr2	NM_012712	10726172	30.8	turq	AMH-dwn, KGF-dwn	fibroblueast growth factor receptor 2
	Itgb3bp	NM_001013213	10878272	11.0	blue	GDNF-dwn	integrin beta 3 binding protein
	Plxna4a	NM_001107852	10861678	30.9	turq	AMH-dwn, KGF-dwn	plexin A4, A
	Tmem151a	NM_001107570	10727725	26.2	turq	AMH-dwn, KGF-dwn	transmembrane protein 151A
Signaling	Nrgn	NM_024140	10916228	16.1	yellow	AMH, KGF-dwn, PDGF-up	neurogranin
	Dusp4	NM_022199	10792035	17.3	yellow	AMH-dwn	dual specificity phosphatase 4
	Dusp6	NM_053883	10895144	16.6	yellow	BMP4-dwn	dual specificity phosphatase 6
	Efna5	NM_053903	10930204	15.0	yellow	AMH, GDNF-dwn	ephrin A5
	Map3k1	NM_053887	10821276	26.8	turq	KGF-dwn	mitogen activated protein kinase
	Pde7b	NM_080894	10717069	17.8	brown	KL-dwn	phosphodiesterase 7B
	Rem1	NM_001025753	10840861	14.5	yellow	PDGF-up	RAS (RAD and GEM)-like
	Shc4	NC_005102	10849423	17.0	yellow	AMH-dwn	SHC family, member 4
	Ubash3b	AC_000076	10916476	16.9	yellow	PDGF-up	ubiquitin associated
Transcription	Btg4	NM_001013176	10909937	7.4	red	KL-dwn	B-cell translocation gene 4
	Etv5	NM_001107082	10752034	14.4	yellow	AMH-dwn	ets variant 5
	Fbxo15	NM_001108436	10803025	11.9	red	KL-dwn	F-box protein 15
Misc. & Unknown	Depdc2	NM_001107899	10875023	29.5	brown	KL-dwn	DEP domain containing 2
	Fam154a	AC_000073	10877890	11.4	red	KL-dwn	similarity 154, member A
	LOC686725	AC_000076	10915208	25.9	turq	AMH-dwn	hypothetical protein LOC686725

Table 2. Cont.

Function	GeneSymbol	GeneBank	Probeset	k.in*	Module	Regulation by Growth Factor	GeneTitle
	RGD1306186	BC090317	10881318	9.2	red	KL-dwn	similar to RIKEN cDNA 4930569K13
	RGD1306622	XM_001074493	10728647	32.6	turq	AMH-dwn, KGF-dwn	similar to KIAA0954 protein
	RGD1308023	XR_006437	10850490	17.6	brown	KL-dwn, LIF-dwn	similar to CG5521-PA
	RGD1566021	AC_000086	10800122	19.6	brown	KL-dwn	similar to KIAA1772

Short list of 55 genes that are the most connected genes with known functions in the modules of interest.
Selected from the top 10% most connected genes in each module (except blue module for which considered top 20% as many hubs are not annotated. Abbreviations used: dwn - down-regulated; up - up-regulated; (*)- k in. is connectivity coefficient obtained/calculated in network analysis.
doi:10.1371/journal.pone.0011637.t002

demonstrate that these types of modules are enriched for known biological pathways for genes that associate with disease traits and for genes that are linked to common genetic loci [6,35].

The current study employed a weighted gene co-expression network approach that has been extensively used for uncovering biologically meaningful gene modules [7,13,15] to explore novel pathways involved in primordial follicle development. An unsupervised and unbiased approach was used to nominate potential regulatory candidates for these modules based on gene network connectivity. The connectivity score shows how well under different treatments the changes in gene expression for a gene are correlated with the changes in expression for every other gene. In the current study, the gene co-expression network analysis helped select 55 highly connected genes for further functional analysis. An automated literature search of these 55 genes revealed a sub-network relationship among them as presented in Figure 3. This sub-network suggested regulatory roles for Pdgfa and Fgfr2 (the receptor) for Fgf2 and Fgf7 (KGF). PDGF, KGF/FGF7 and FGF2 proteins have previously been shown to regulate primordial to primary follicle transition [29,34]. Therefore, the bionetwork predicted to be involved in the regulation of primordial follicle development identified two previously known regulatory factors which validated the utility of the network analysis for identifying candidate regulatory genes, consistent with previous network studies [13,15]. This sub-network also identified connective tissue growth factor (Ctgf) [16,36] as a putative regulator of primordial follicle development. An ovarian organ culture experiment confirmed that CTGF promotes primordial to primary follicle transition. Therefore, a regulatory factor predicted to be important for primordial follicle development was confirmed to be involved which further validated the bionetwork approach. A microarray analysis of CTGF-treated ovaries showed an altered gene set similar to those of the other growth factors known to regulate follicle transition. These observations validate the network-based systems biology approach to elucidate the regulation of a complex developmental process.

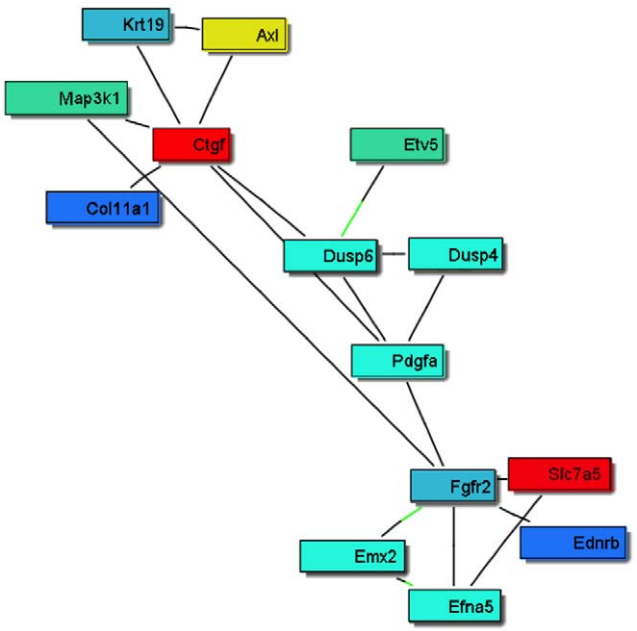


Figure 3. Sub-network connection scheme for the most highly connected 55 candidate genes obtained by global literature analysis. Only 15 connected genes from the list of 55 are shown, while the rest are not connected and not shown. Red and yellow colors represent up-regulated genes, blue and turquoise colors – down-regulated genes.
doi:10.1371/journal.pone.0011637.g003

Consideration of the 55 intra-module hub genes from critical regulatory modules revealed a number of signaling and cellular processes were influenced, Figure 5 and Figure S3. In the growth factor/chemokine family Pdgfa and Ctgf were confirmed to be involved. The IL16 identified is currently being investigated as a potential regulatory candidate. The specific genes identified in Table 2 and associated regulatory processes provide potential therapeutic targets to regulate primordial follicle development. The ability to inhibit or stimulate primordial follicle development with a therapeutic treatment has a number of clinical applications. A delay in primordial follicle development and maintenance of the primordial pool could delay the onset of menopause and extend the reproductive life span of a female. In addition, the ability to therapeutically inhibit primordial follicle development would provide a treatment for premature ovarian failure, a disease when the primordial pool is lost early in life causing female infertility. In contrast, the therapeutic stimulation of primordial follicle development could treat forms of female infertility [4]. The induction of primordial follicle development also could promote the loss of the primordial pool and induce female sterility. The bionetwork identified in the current study produced a number of potential therapeutic targets to manipulate primordial follicle development and female reproductive capacity.

The systems biology approach taken with this network analysis of primordial follicle development identified clusters and modules of genes involved in this critical development process. A number of the growth factors previously shown to be involved (e.g. PDGF and bFGF) were identified, but other factors known to be important for ovarian development were not identified. Often a reductionist approach such as a knockout mouse model can

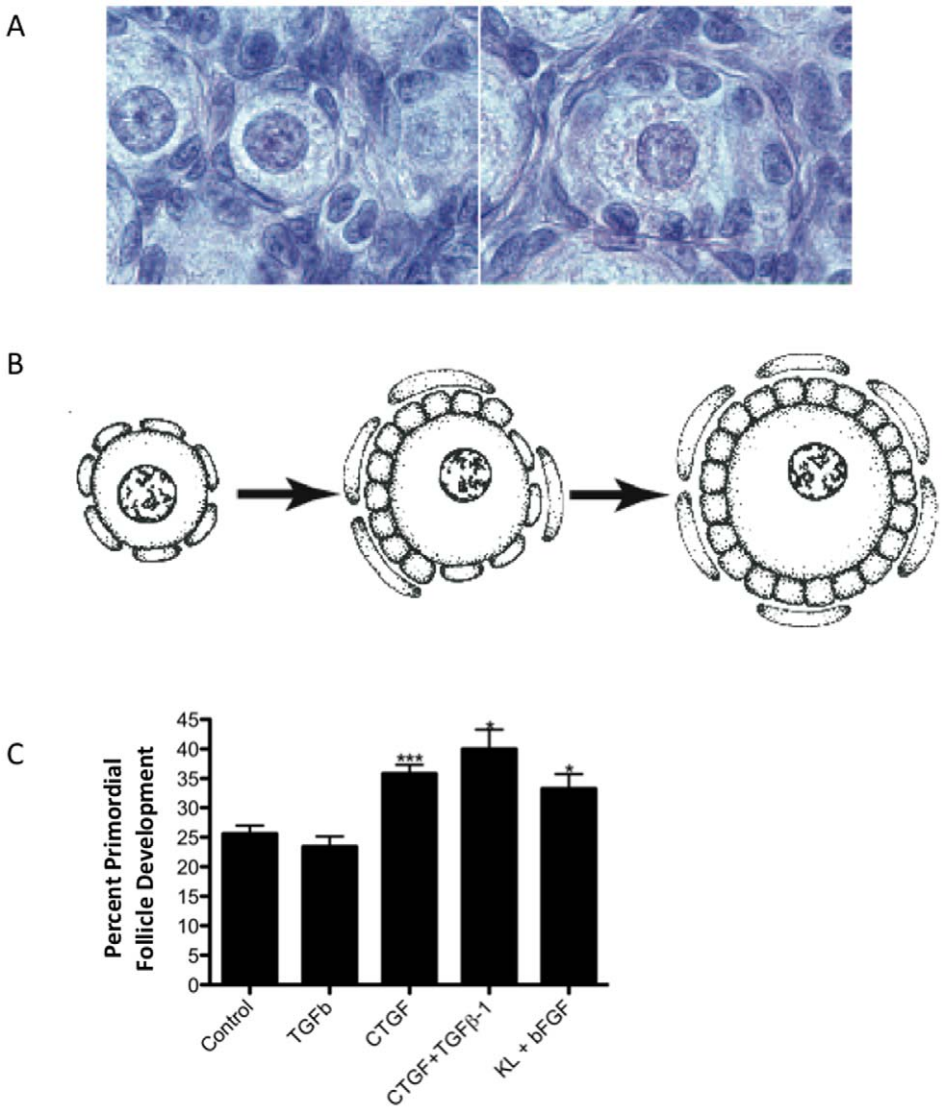


Figure 4. Analysis of the role of CTGF in primordial follicle transition. A) Hematoxylin-eosin stained ovary sections showing a representative arrested primordial follicle (left), and a developing primary follicle after having undergone primordial to primary follicle transition (right). **B**) Graphic representation of primordial and primary follicles. **C**) Effect of CTGF on cultured ovaries. Ovaries were cultured for 10 days with the treatments indicated. Ovarian histological analysis determined the percentage of primordial versus developing follicles for each ovary. Bars are the mean percent developing follicles ± SEM. N = 5–7 per treatment from four different replicate experiments. Asterisks indicate a significant (*$p<0.05$ or ***$p<0.01$) difference from control by Dunnet's post-hoc test after a significant result of ANOVA.
doi:10.1371/journal.pone.0011637.g004

identify a factor as being important for the maintenance of tissue development or function, but this does not mean the factor is regulated during the process. In addition, critical developmental processes such as primordial follicle development often have a set of compensatory factors that have evolved such that loss of any one will still allow the process to proceed. Therefore, knockout models often do not have phenotypes for these factors. This does not mean the factor is not important, but instead that the developmental process is essential and thus multiple factors compensate to assure the developmental process occurs. The current study takes a systems biology approach to identify networks of genes involved in the process without the bias of a reductionist model. Therefore, novel groups of factors and cellular processes were identified that now require further investigation.

The integrative analysis revealed a gene sub-network involved in primordial follicle development to elucidate the basic developmental biology of this process and provide potential therapeutic targets for ovarian disease and function. This sub-network was validated by the presence of two genes previously identified as being important. A new gene identified, Ctgf, was tested and found to regulate primordial follicle development. Therefore, the network based systems biology approach was partially validated for a normal developmental process. This type of approach will likely be invaluable to study development on a systems biology level in the future.

Materials and Methods

Ovarian organ culture

Four-day old female Sprague-Dawley rats (Harlan Laboratories, Inc., USA) were euthanized according to the laboratories Washington State University IACUC approved (#02568-014)

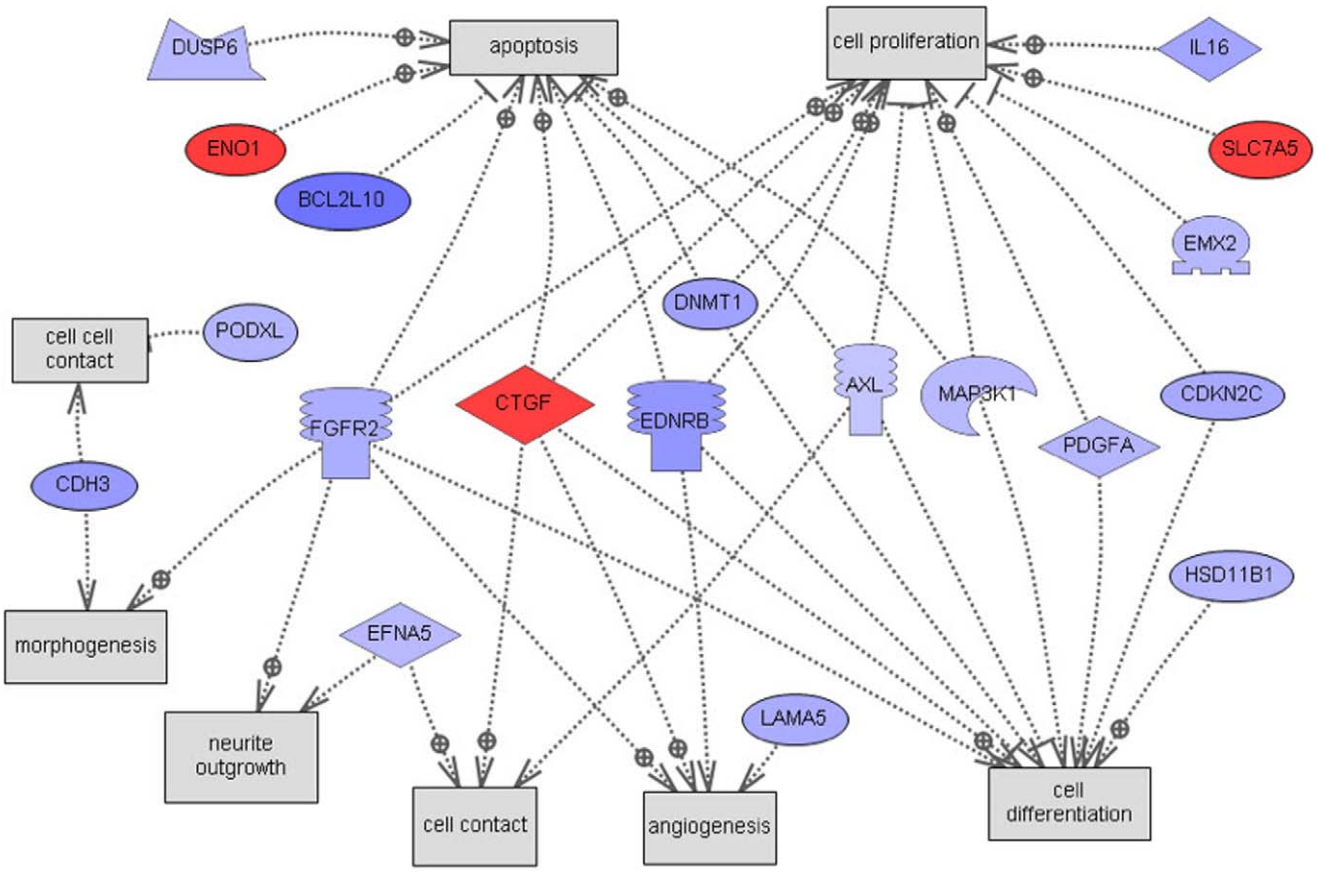


Figure 5. Scheme of direct connections to cellular processes for the 55 candidate regulatory genes obtained by global literature analysis. Only 19 connected genes from the list of 55 are shown, the rest are not connected and not shown. Node shapes code: oval and circle – protein; crescent – protein kinase and kinase; diamond – ligand; irregular polygon – phosphatase; circle/oval on tripod platform – transcription factor; ice cream cone – receptor. Red color represents up-regulated genes, blue color – down-regulated genes, grey rectangles represent cell processes; arrows with plus sign show positive regulation/activation, arrows with minus sign – negative regulation/inhibition.
doi:10.1371/journal.pone.0011637.g005

protocols and the ovaries removed and cultured whole as described previously [24]. Four-day old rat ovaries contain almost exclusively primordial follicles. For ovary culture experiments in which ovarian RNA was collected, 2–3 ovaries per well were cultured with media changes every 24 hours for two days in the absence (controls) or presence (treated) of either AMH (human Anti-Mülerian hormone)(50ng/mL, R&D Systems Inc., USA), FGF2 (rat Fibroblast growth factor 2)(50ng/mL, R&D Systems Inc., USA), BMP4 (human Bone morphogenetic protein 4)(50ng/mL, R&D Systems Inc., USA), GDNF (rat Glial derived neurotrophic factor)(50ng/mL, Calbiochem, USA), FGF7 (human fibroblast growth factor 7/keratinocyte growth factor)(50ng/mL, R&D Systems Inc., USA), KITLG (mouse Kit ligand)(50ng/mL, R&D Systems Inc., USA), LIF (rat leukemia inhibitory factor)(50ng/mL, Chemicon/Millipore, USA), PDGF-AB (rat platelet derived growth factor AB heterodimer)(50ng/mL, R&D Systems Inc., USA), TGFβ1 (human transforming growth factor beta 1)(50ng/mL, R&D Systems Inc., USA) or CTGF (human connective tissue growth factor)(50ng/mL, PeproTech Inc., NJ USA). After only two days of culture there are no morphological differences between control and growth factor-treated ovaries, so measurements of whole-ovary gene expression reflect differences in RNA transcription, rather than differing proportions of cell types due to differing cell proliferation between treatments.

In order to determine the effect of CTGF on primordial to primary follicle development, ovaries were cultured as above for ten days in the absence or presence of CTGF (50ng/mL), alone or in combination with TGF-beta1 (50ng/mL). After culture ovaries were fixed with Bouin's solution, paraffin embedded, sectioned onto microscope slides and stained with hematoxylin and eosin as described previously [24].

Morphometric Analysis

The number of follicles at each developmental stage was counted and averaged in two serial sections from the largest cross-section through the center of the ovary. Total follicle number has not been found to change between treatment groups. Rather, only the percentage of follicles at each developmental stage changes with treatment [28,34]. KL was used as a positive control for the organ culture experiments. Follicles in ovarian cross sections were classified as primordial (stage 0), or developing (stages 1–4: early primary, primary, transitional and preantral) as previously described [37]. Primordial follicles consist of an oocyte arrested in prophase I of meiosis that is partially or completely encapsulated by flattened squamous pregranulosa cells. Early transition primary follicles have initiated development (i.e., undergone primordial to primary follicle transition) and contain at least two cuboidal granulosa cells. Primary and preantral follicles exhibit one or more complete layers of cuboidal granulosa cells. Four-day old ovaries

contain predominately primordial follicles [26,38]. Hematoxylin/eosin stained ovarian sections were analyzed at 400× magnification using light microscopy. Follicles containing degenerating red eosin-stained oocytes were not counted.

RNA preparation

RNA was isolated from whole rat ovaries after homogenization in one ml Trizol™ reagent (Sigma-Aldritch, USA), according to manufacturer's instructions. Two or three ovaries from the same culture well (from different rat pups out of the same litter) and receiving the same treatment were pooled and homogenized together. On any given day a culture experiment was performed, the treatment groups included untreated control ovaries and one to three different growth factor treatments. Homogenized samples were stored at −70 C until the time of RNA isolation. After isolation from Trizol, RNA was further purified using RNeasy MinElute Cleanup Kits (Qiagen, USA) and stored in aqueous solution at −70 C.

Microarray Analysis

The microarray analysis was performed by the Genomics Core Laboratory, Center for Reproductive Biology, Washington State University, Pullman, WA using standard Affymetrix reagents and protocol. Briefly, mRNA was transcribed into cDNA with random primers, cRNA was transcribed, and single-stranded sense DNA was synthesized which was fragmented and labeled with biotin. Biotin-labeled ssDNA was then hybridized to the Rat Gene 1.0 ST microarrays containing 27,342 transcripts (Affymetrix, Santa Clara, CA, USA). Hybridized chips were scanned on Affymetrix Scanner 3000. CEL files containing raw data were then pre-processed and analyzed with Partek Genomic Suite 6.3 software (Partek Incorporated, St. Louis, MO) using an RMA GC-content adjusted algorithm. Raw data pre-processing was performed in 2 groups. The first group containing 38 samples CEL files were pre-processed in Partek all together as one experiment. Comparison of all array histogram graphs demonstrated the data for all 38 chips were similar and appropriate for further analysis. The second group of samples for microarray analysis consisting of 3 CTGF-treated and 3 corresponding control ovaries was run, pre-processed and analyzed *post factum*, separately from the rest of the samples as a result of a discovery from network analysis. Partek pre-processing algorithm for these 6 CEL files used the same criteria as used for the first group.

The microarray quantitative data involves over 10 different oligonucleotides arrayed for each gene and the hybridization must be consistent to allow a statistically significant quantitative measure of gene expression and regulation. In contrast, a quantitative PCR procedure only uses two oligonucleotides and primer bias is a major factor in this type of analysis. Therefore, we did not attempt to use PCR based approaches as we feel the microarray analysis is more accurate and reproducible without primer bias such as PCR based approaches.

All microarray CEL files (MIAME compliant raw data) from this study have been deposited with the NCBI gene expression and hybridization array data repository (GEO, http://www.ncbi.nlm.nih.gov/geo) (GEO Accession number: GSE20324), all arrays combined with one accession number, and can be also accessed through www.skinner.wsu.edu. For gene annotation, Affymetrix annotation file RaGene1_0stv1.na30.rn4.transcript.csv was used unless otherwise specified.

Network analysis

The network analysis was restricted to genes differentially expressed between the control and the treatment groups based on previously established criteria: (1) fold change of group means ≥ 1.2 or ≤ 0.83; (2) T test p-value ≤ 0.05; and (3) absolute difference of group means ≥ 10. The union of the differentially expressed genes from the different treatments resulted in 1,540 genes being identified and used for constructing a weighted gene co-expression network [7,8]. Unlike traditional un-weighted gene co-expression networks in which two genes (nodes) are either connected or disconnected, the weighted gene co-expression network analysis assigns a connection weight to each gene pair using soft-thresholding and thus is robust to parameter selection. The weighted network analysis begins with a matrix of the Pearson correlations between all gene pairs, then converts the correlation matrix into an adjacency matrix using a power function $f(x) = x^{\beta}$. The parameter β of the power function is determined in such a way that the resulting adjacency matrix (i.e., the weighted co-expression network), is approximately scale-free. To measure how well a network satisfies a scale-free topology, we use the fitting index proposed by Zhang & Horvath [7] (i.e., the model fitting index R^2 of the linear model that regresses $log(p(k))$ on $log(k)$ where k is connectivity and $p(k)$ is the frequency distribution of connectivity). The fitting index of a perfect scale-free network is 1. For this dataset, we select the smallest β $(=7)$ which leads to an approximately scale-free network with the truncated scale-free fitting index R^2 greater than 0.75. The distribution $p(k)$ of the resulting network approximates a power law: $p(k) \sim k^{-1.29}$.

To explore the modular structures of the co-expression network, the adjacency matrix is further transformed into a topological overlap matrix [30]. As the topological overlap between two genes reflects not only their direct interaction, but also their indirect interactions through all the other genes in the network. Previous studies [7,30] have shown that topological overlap leads to more cohesive and biologically meaningful modules. To identify modules of highly co-regulated genes, we used average linkage hierarchical clustering to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules [39]. A total of sixteen modules were identified and the module size was observed to range from 20 to 194 genes.

To distinguish between modules, each module was assigned a unique color identifier, with the remaining, poorly connected genes colored grey. The hierarchical clustering over the topological overlap matrix (TOM) and the identified modules is shown. In this type of map, the rows and the columns represent genes in a symmetric fashion, and the color intensity represents the interaction strength between genes. This connectivity map highlights that genes in the ovary transcriptional network fall into distinct network modules, where genes within a given module are more interconnected with each other (blocks along the diagonal of the matrix) than with genes in other modules. There are a couple of network connectivity measures, but one particularly important one is the within module connectivity (k.in). The k.in of a gene was determined by taking the sum of its connection strengths (co-expression similarity) with all other genes in the module which the gene belonged.

Gene Co-expression Network Analysis Clarification

Gene networks provide a convenient framework for exploring the context within which single genes operate. Networks are simply graphical models comprised of nodes and edges. For gene co-expression networks, an edge between two genes may indicate that the corresponding expression traits are correlated in a given population of interest. Depending on whether the interaction strength of two genes is considered, there are two different approaches for analyzing gene co-expression networks: 1) an

unweighted network analysis that involves setting hard thresholds on the significance of the interactions, and 2) a weighted approach that avoids hard thresholds. Weighted gene co-expression networks preserve the continuous nature of gene-gene interactions at the transcriptional level and are robust to parameter selection. An important end product from the gene co-expression network analysis is a set of gene modules in which member genes are more highly correlated with each other than with genes outside a module. Most gene co-expression modules are enriched for GO functional annotations and are informative for identifying the functional components of the network that are associated with disease [6].

This gene co-expression network analysis (GCENA) has been increasingly used to identify gene sub-networks for prioritizing gene targets associated with a variety of common human diseases such as cancer and obesity [11,12,13,14,15]. One important end product of GCENA is the construction of gene modules comprised of highly interconnected genes. A number of studies have demonstrated that co-expression network modules are generally enriched for known biological pathways, for genes that are linked to common genetic loci and for genes associated with disease [6,7,11,13,14,15,40,41,42]. In this way, one can identify key groups of genes that are perturbed by genetic loci that lead to disease, and that define at the molecular level disease states. Furthermore, these studies have also shown the importance of the hub genes in the modules associated with various phenotypes. For example, GCENA identified ASPM, a hub gene in the cell cycle module, as a molecular target of glioblastoma [15] and MGC4504, a hub gene in the unfolded protein response module, as a target potentially involved in susceptibility to atherosclerosis [13].

Pathway Analysis

Resulting lists of differentially expressed genes for each growth factor treatment as well as for each module generated in the network analysis were analyzed for KEGG (Kyoto Encyclopedia for Genes and Genome, Kyoto University, Japan) pathway enrichment using Pathway-Express, a web-based tool freely available as part of the Onto-Tools (http://vortex.cs.wayne.edu) [43]. Global literature analysis of various gene lists was performed using BiblioSphere PathwayEdition (Genomatix Software GmbH, Munchen, Federal Republic of Germany) software which performs pathway and interaction analysis and labels genes which belong to certain known metabolic and signal transduction pathways. A program based on literature analysis Pathway Studio (Ariadne, Genomics Inc. Rockville MD) was used to evaluate cellular processes connected to differentially expressed genes.

Supporting Information

Figure S1 Network scheme for 1540 differentially expressed genes obtained by global literature analysis using BiblioSphere Pathway Edition Software (Genomatix Software GmbH, Munchen, Federal Republic of Germany). Different node colors represent different modules. A - the whole scheme clearly indicates 5 distinguished groups of genes (each group is shown separately on pp. 2–6) connected to 5 central genes: Nfkb1 (B, page 2), Vegfa (C, page 3), Egfr (D, page 4), and Gadd45a (F, page 6). Only connected genes are shown.
Found at: doi:10.1371/journal.pone.0011637.s001 (2.07 MB PDF)

Figure S2 KEGG Pathway "Complement and Coagulation Cascades" enriched by regulated genes from 1,540 gene list. Red nodes represent up-regulated genes, blue - down-regulated, green - not affected genes.
Found at: doi:10.1371/journal.pone.0011637.s002 (0.07 MB PDF)

Figure S3 Scheme of shortest connections to cellular processes for 55 candidate regulatory genes, as obtained by global literature analysis using Pathway Studio 7.0 (Ariadne Genomics, Inc., Rockville, MD; trial version). Only 22 connected genes from the list out of 55 are shown, the rest from the list are not connected and not shown. Node shapes code: oval and circle - protein; crescent - protein kinase and kinase; diamond - ligand; irregular polygon - phosphatase; circle/oval on tripod platform - transcription factor; ice cream cone - receptor. Red color represents up-regulated genes, blue color - down regulated genes, grey nodes represent genes closely connected (next neighbor) to the list genes; grey rectangles represent cell processes; arrows color: grey solid or dotted - regulation, blue - expression, green - promoter binding; arrows with plus sign show positive regulation/activation, arrows with minus sign - negative regulation/inhibition.
Found at: doi:10.1371/journal.pone.0011637.s003 (0.35 MB PDF)

Table S1 Rat Genes Expressed Differentially After Growth Factor Treatment of Ovary. Legends: * - absolute value of difference between means of Control and GF Treatment expression values ** - abbreviations used for modules' color: trq -turquoise; brw - brown; blu- blue; ylw- yellow; prp - purple; gr - grey; grn - green; grlw - green-yellow; blc- black; mbl - midnight-blue; slm - salmon; lcn - light cyan; ***- k in. is connectivity coefficient determined in network analysis.
Found at: doi:10.1371/journal.pone.0011637.s004 (1.37 MB PDF)

Author Contributions

Conceived and designed the experiments: EES MKS. Performed the experiments: EEN MIS RS BZ. Analyzed the data: EEN MIS RS BZ EES MKS. Wrote the paper: EEN MIS RS BZ EES MKS.

References

1. Hirshfield AN (1991) Development of follicles in the mammalian ovary. Int Rev Cytol 124: 43–101.
2. Rajah R, Glaser EM, Hirshfield AN (1992) The changing architecture of the neonatal rat ovary during histogenesis. Dev Dyn 194: 177–192.
3. Peters H, Byskov AG, Himelstein-Braw R, Faber M (1975) Follicular growth: the basic event in the mouse and human ovary. J Reprod Fertil 45: 559–566.
4. Skinner MK (2005) Regulation of primordial follicle assembly and development. Hum Reprod Update 11: 461–471.
5. Pangas SA (2007) Growth factors in ovarian development. Semin Reprod Med 25: 225–234.
6. Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, et al. (2006) Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. J Neurochem 97 Suppl 1: 50–62.
7. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4: Article17.
8. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, et al. (2007) Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. PLoS Comput Biol 3: e69.
9. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, et al. (2005) An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 37: 710–717.
10. Winrow CJ, Williams DL, Kasarskis A, Millstein J, Laposky AD, et al. (2009) Uncovering the genetic landscape for multiple sleep-wake traits. PLoS ONE 4: e5161.

11. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, et al. (2008) Variations in DNA elucidate molecular networks that cause disease. Nature 452: 429–435.
12. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 452: 423–428.
13. Gargalovic PS, Imura M, Zhang B, Gharavi NM, Clark MJ, et al. (2006) Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proc Natl Acad Sci U S A 103: 12741–12746.
14. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, et al. (2006) Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2: e130.
15. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, et al. (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proc Natl Acad Sci U S A 103: 17402–17407.
16. Leask A, Abraham DJ (2003) The role of connective tissue growth factor, a multifunctional matricellular protein, in fibroblast biology. Biochem Cell Biol 81: 355–363.
17. Durlinger AL, Kramer P, Karels B, de Jong FH, Uilenbroek JT, et al. (1999) Control of primordial follicle recruitment by anti-Mullerian hormone in the mouse ovary. Endocrinology 140: 5789–5796.
18. Nilsson E, Rogers N, Skinner MK (2007) Actions of anti-Mullerian hormone on the ovarian transcriptome to inhibit primordial to primary follicle transition. Reproduction 134: 209–221.
19. Nilsson EE, Skinner MK (2004) Kit ligand and basic fibroblast growth factor interactions in the induction of ovarian primordial to primary follicle transition. Mol Cell Endocrinol 214: 19–25.
20. Garor R, Abir R, Erman A, Felz C, Nitke S, et al. (2009) Effects of basic fibroblast growth factor on in vitro development of human ovarian primordial follicles. Fertil Steril 91: 1967–1975.
21. Matos MH, Lima-Verde IB, Bruno JB, Lopes CA, Martins FS, et al. (2007) Follicle stimulating hormone and fibroblast growth factor-2 interact and promote goat primordial follicle development in vitro. Reprod Fertil Dev 19: 677–684.
22. Nilsson EE, Skinner MK (2003) Bone morphogenetic protein-4 acts as an ovarian follicle survival factor and promotes primordial follicle development. Biol Reprod 69: 1265–1272.
23. Tanwar PS, O'Shea T, McFarlane JR (2008) In vivo evidence of role of bone morphogenetic protein-4 in the mouse ovary. Anim Reprod Sci 106: 232–240.
24. Dole G, Nilsson EE, Skinner MK (2008) Glial-derived neurotrophic factor promotes ovarian primordial follicle development and cell-cell interactions during folliculogenesis. Reproduction 135: 671–682.
25. Kezele P, Nilsson EE, Skinner MK (2005) Keratinocyte growth factor acts as a mesenchymal factor that promotes ovarian primordial to primary follicle transition. Biol Reprod 73: 967–973.
26. Parrott JA, Skinner MK (1999) Kit-ligand/stem cell factor induces primordial follicle development and initiates folliculogenesis. Endocrinology 140: 4262–4271.
27. Hutt KJ, McLaughlin EA, Holland MK (2006) KIT/KIT ligand in mammalian oogenesis and folliculogenesis: roles in rabbit and murine ovarian follicle activation and oocyte growth. Biol Reprod 75: 421–433.
28. Nilsson EE, Kezele P, Skinner MK (2002) Leukemia inhibitory factor (LIF) promotes the primordial to primary follicle transition in rat ovaries. Mol Cell Endocrinol 188: 65–73.
29. Nilsson EE, Detzel C, Skinner MK (2006) Platelet-derived growth factor modulates the primordial to primary follicle transition. Reproduction 131: 1007–1015.
30. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297: 1551–1555.
31. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24: 719–720.
32. Cicha I, Goppelt-Struebe M (2009) Connective tissue growth factor: context-dependent functions and mechanisms of regulation. Biofactors 35: 200–208.
33. Gressner OA, Gressner AM (2008) Connective tissue growth factor: a fibrogenic master switch in fibrotic liver diseases. Liver Int 28: 1065–1079.
34. Nilsson E, Parrott JA, Skinner MK (2001) Basic fibroblast growth factor induces primordial follicle development and initiates folliculogenesis. Mol Cell Endocrinol 175: 123–130.
35. Ghazalpour A, Doss S, Sheth SS, Ingram-Drake LA, Schadt EE, et al. (2005) Genomic analysis of metabolic pathway gene expression in mice. Genome Biol 6: R59.
36. Harlow CR, Hillier SG (2002) Connective tissue growth factor in the ovarian paracrine system. Mol Cell Endocrinol 187: 23–27.
37. Oktay K, Schenken RS, Nelson JF (1995) Proliferating cell nuclear antigen marks the initiation of follicular growth in the rat. Biol Reprod 53: 295–301.
38. Kezele PR, Ague JM, Nilsson E, Skinner MK (2005) Alterations in the ovarian transcriptome during primordial follicle assembly and development. Biol Reprod 72: 241–255.
39. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 1: 54.
40. Schadt EE, Molony C, Chudin E, Hao K, Yang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107.
41. Zhu J, Zhang B, Schadt EE (2008) A systems biology approach to drug discovery. Adv Genet 60: 603–635.
42. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, et al. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet 40: 854–861.
43. Draghici S, Khatri P, Tarca AL, Amin K, Done A, et al. (2007) A systems biology approach for pathway level analysis. Genome Res 17: 1537–1545.