Intact cell MALDI-TOF mass spectrometry on single bovine oocyte and follicular cells combined with top-down proteomics: A novel approach to characterise markers of oocyte maturation

Valérie Labas, Ana-Paula Teixeira-Mechin, Laura Bouguereau, Audrey Gargaros-Ratajczak, Lucie Spina, Aurélie Marestaing, Svetlana Uzbekova

HAL Id: hal-01605876
https://hal.science/hal-01605876

Submitted on 26 May 2020

Accepted Manuscript

Intact cell MALDI-TOF mass spectrometry on single bovine oocyte and follicular cells combined with top-down proteomics: A novel approach to characterise markers of oocyte maturation

Valérie Labas, Ana-Paula Teixeira-Gomes, Laura Bouguereau, Audrey Gargaros, Lucie Spina, Aurélie Marestaing, Svetlana Uzbekova

Please cite this article as: Valérie Labas, Ana-Paula Teixeira-Gomes, Laura Bouguereau, Audrey Gargaros, Lucie Spina, Aurélie Marestaing, Svetlana Uzbekova , Intact cell MALDI-TOF mass spectrometry on single bovine oocyte and follicular cells combined with top-down proteomics: A novel approach to characterise markers of oocyte maturation. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Jprot(2017), doi: 10.1016/j.jprot.2017.03.027

Intact cell MALDI-TOF mass spectrometry on single bovine oocyte and follicular cells combined with top-down proteomics: a novel approach to characterise markers of oocyte maturation

Short Title: Proteomic analysis of single bovine oocytes and follicular cells

Authors:

Valérie LABAS[1,3], Ana-Paula TEIXEIRA-GOMES[2,3], Laura BOUGUEREAU[2-3], Audrey GARGAROS[1,3], Lucie SPINA [3,4], Aurélie MARESTAING[1,3], Svetlana UZBEKOVA[1, 3*]

1) *UMR PRC, INRA 85, CNRS, Université de Tours, IFCE, 37380 Nouzilly, France*

2) *UMR ISP, INRA, Université de Tours, 37380 Nouzilly, France*

3) *INRA, Plateforme d'Analyse Intégrative des Biomolécules, Laboratoire de Spectrométrie de Masse, 37380 Nouzilly, France*

4) *INSA/CNRS 5504 - UMR INSA/INRA 792, Toulouse*

* Corresponding author :

Svetlana UZBEKOVA

Institut National de la Recherche Agronomique

Centre de Recherche Val de Loire

UMR INRA85, CNRS7247, UFR, IFCE

Physiologie de la Reproduction et des Comportements

37380 Nouzilly, France.

E-mail: svetlana.uzbekova@inra.fr

1

Abstract

Intact cell MALDI-TOF mass spectrometry (ICM-MS) was adapted to bovine follicular cells from individual ovarian follicles to obtain the protein/peptide signatures (<17 kDa) of single oocytes, cumulus cells (CC) and granulosa cells (GC), which shared a total of 439 peaks. By comparing the ICM-MS profiles of single oocytes with CC before and after *in vitro* maturation (IVM), 71 different peaks were characterised, and their relative abundance was found to vary depending on the stage of oocyte meiotic maturation. To identify these endogenous biomolecules, top-down workflow using high resolution MS/MS (TD HR-MS) was performed on the protein extracts from oocytes, CC and GC. The TD HR-MS proteomic approach allowed for: (1) identification of 386 peptide/proteoforms encoded by 194 genes; and (2) characterisation of proteolysis products likely resulting from the action of kallikreins and caspases. In total, 136 peaks observed by ICM-MS were annotated by TD HR-MS (ProteomeXchange PXD004892). Among these, 16 markers of maturation were identified, including IGF2 binding protein 3 and hemoglobin B in the oocyte, thymosins beta-4 and beta-10, histone H2B and ubiquitin in CC. The combination of ICM-MS and TD HR-MS proved to be a suitable strategy to identify non-invasive markers of oocyte quality using limited biological samples.

Significance:

Intact cell MALDI-TOF mass spectrometry on single oocytes and their surrounding cumulus cells, coupled to an optimised top-down HR-MS proteomic approach on ovarian follicular cells, was used to identify specific markers of oocyte meiotic maturation represented by whole low molecular weight proteins or products of degradation by specific proteases.

Highlights

- ICM-MS fingerprints of single oocytes and surrounding ovarian follicular cells.
- Top-down analysis of follicular cells by μLC-HR-MS/MS with or without pre-fractionation.
- Differential ICM-MS and TD HR-MS revealed oocyte meiotic maturation markers.
- Markers are small proteins or products of proteolysis by kallikreins and caspases.

Introduction

In mammals, the competence of female germ cells to be fertilised and develop into viable embryos, representing oocyte quality, is acquired throughout folliculogenesis. Oocyte maturation is the final step of female gamete development before ovulation, and is crucial for oocyte quality. In farm animals, including bovine, *in vitro* maturation (IVM), followed by *in vitro* fertilisation and *in vitro* embryo culture, is routinely used for reproductive biotechnologies to ensure livestock production. However, following IVM, the development of bovine oocytes to the blastocyst stage prior to transfer to the recipient cow is limited to about 30% [1]. Therefore, a better understanding of the molecular mechanisms involved in the regulation of oocyte maturation, and identification of potential markers related to oocyte developmental competence, may help to improve the efficiency of current IVM methods. Transcription does not occur in the oocyte during maturation, however, a number of proteins are produced to ensure the progression of oocyte meiosis from prophase-I to metaphase-II, and subsequently, early development of the embryo post-fertilisation prior to genome activation. Thus, proteomics appear to be a suitable approach for monitoring the dynamic changes functionally associated with oocyte phenotypes and to identify the proteins involved in these processes, not only in the oocytes, but also in the surrounding follicular cells. Indeed, within the ovarian follicle the oocyte is surrounded by cumulus cells (CC) and granulosa cells (GC) which regulate oocyte growth and metabolism. CC are physically and metabolically coupled to an oocyte, and play also an important role in maturation and fertilisation [2]. During maturation, both the oocyte and CC undergo significant morphological and molecular changes both at mRNA and protein levels. The expression of molecular factors by surrounding cells may reflect oocyte maturation and developmental potential, and therefore, these have been intensively studied using transcriptomic and proteomic approaches [3]. However, the scarcity of biological material (single oocytes and CC) limits the use of proteomics.

In bovine antral follicles, the oocyte, which is a large cell of about 100 μm in diameter, contains only 80–100 ng of protein [4], whereas approximately 1 μg of protein is available from the ≈ 20,000 cumulus cells attached to oocyte [5]. Furthermore, approximatively 50-times more protein can be obtained from the GC in small follicles, and up to 60-times more from preovulatory follicles, as reported in primates [6, 7]. In bovine, two

3

large-scale qualitative proteomic experiments have been performed on oocytes and CC, and both showed significant differences between the proteomes of these two cell types [8, 9]. Analysis of 500 oocyte-cumulus complexes using 2D-LC-MS/MS showed that over 1092 proteins are expressed in oocytes and 4395 in the corresponding CC [9]. These analyses constitute reference proteomes for bovine oocytes and CC, however, the proteins that could reliably reflect oocyte maturation and quality remain to be identified. Several studies have reported functional proteomic approaches performed on pools of oocytes or CC in different conditions in bovine [10-13], porcine [14-16], murine [17-19] and humans [20-22]. However, analyses of the single oocytes and their corresponding CC have so far remained inaccessible for proteomic approaches, for which analysis at the individual follicle level is required to study cellular heterogeneity and to gain a better insight into oocyte biology. Nevertheless, after efficient sample preparation using paramagnetic bead technology [23], more than 450 proteins were identified in single human oocytes [24]. However, this efficient sample preparation procedure is complex, and is not convenient for high-throughput quantitative analysis.

Therefore, in this study, we aimed to carry out a global approach based on profiling using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) on follicular cells from individual follicles. MALDI-TOF MS can detect the most abundant intact molecular ions present in a biological sample over a relatively wide mass range (typically 1000–30,000 Da), with a high sensitivity (from the picomolar to femtomolar range) and a high tolerance for salts. Consequently, MALDI MS is a powerful tool for the direct top-down analysis of complex peptide/protein mixtures within biological specimens, such as intact cells. Nowadays, the intact cell MALDI-TOF mass spectrometry (ICM-MS) approach is a well-established phenotyping method, allowing the direct detection, at low resolution, of endogenous peptides and proteins in whole cells without the need for extraction, pre-fractionation or separation of biomolecules prior to analysis. Early studies applied ICM-MS for the phyloproteomic classification of bacteria [25-27]. More recently, ICM-MS approaches have been mainly applied to isolated homogenous eukaryotic cellular populations for phenotyping mammalian or insect cell lines [28, 29], for differentiating between similar cell lines (e.g., glial or immune cells) [30-32], evaluating the cell response to different physiological [33] and pathological conditions [34, 35], or response to treatment with toxic chemicals [36, 37] or drugs [38], as well as for screening productive cell lines for antibody production [39-41]. In our laboratory, we have already shown that ICM-MS profiles from epididymal spermatozoa reflect their proteomic changes during maturation, predominantly

4

those related to differences in the activity of intracellular proteases and peptidases of sperm [42]. Furthermore, using ICM-MS, we also demonstrated that profiling of sperm from ejaculate could be used to identify proteome patterns linked with fertility in chickens [43], and to develop a molecular diagnostic test enabling the selection of high quality semen [44].

In the current study, we propose to adapt this fast, reproducible and sensitive method to single bovine oocytes at different stages of meiotic maturation, in addition to the surrounding follicular cells (CC and GC), for the characterisation of biomarkers that may be involved in oocyte maturation, and related to oocyte quality.

In order to perform a detailed characterisation of endogenous molecular species observed in the ICM-MS profiles, and to identify markers of interest, we chose a top-down (TD) proteomic approach. In this method, intact proteins are directly fragmented by high-resolution mass spectrometry (HR-MS), which allows for the direct identification of peptidoforms and proteoforms and their structural characterisation [45-47]. Advances in TD HR-MS proteomics have made it possible to apply this technique even when only a small amount of biological material is available, as previously demonstrated for 100–200 µg of HeLa cell lysate for LC-MS analysis [48], or detection of protein standards as low as 0.3 pmol [48].

Using bovine oocyte, CC and GC samples, we proposed to employ different analytical strategies, such as direct infusion and/or reverse-phase µliquid chromatography coupled to tandem high resolution mass spectrometry (µLC-HR-MS/MS) combined with several pre-fractionation methods, in order to identify proteins. The objectives of the present study were: (1) to adapt ICM-MS for use with bovine follicular cells, and acquire specific molecular profiles from GC, CC and oocytes derived from a single follicle; (2) to analyse the endogenous peptides/proteins in follicular cells by TD HR-MS in order to identify representative biomolecules in the ICM-MS spectra; and (3) to combine two TD approaches based on low (ICM-MS) and high resolution MS (µLC-HR-MS/MS) to propose new markers involved in oocyte maturation.

5

Materials and methods

Ethics

In the present study, no experiments with living animals were performed.

Chemicals

All chemicals were provided by Sigma-Aldrich (Saint-Quentin Fallavier, France) unless others stated.

Experimental workflow is presented in Fig.1 and described in details below.

Biological materials

For all experiments, bovine ovaries were collected in commercial slaughterhouse. Follicles of 3-8 mm in size were aspirated, oocyte-cumulus complexes (OCC) were retrieved under binocular microscope and 24h IVM culture in serum free TCM199-enriched medium at 38.8°C was performed as described earlier [49].

1. For ICM-MS analysis, follicular cells were prepared as follow:

    a) Oocyte either before IVM (immature oocytes), or after 6h (GVBD), 10h (Metaphase-I), 18h (Telophase-Anaphase I) and 24h of IVM (Metaphase-II mature oocytes) were stripped from their cumulus and rinsed with Phosphate-buffered saline (PBS), pH 7.4. The absence of CC was checked under binocular microscope, oocytes were then rinsed twice in 100µL drops of Tris-Sucrose buffer (TSB : 20 mM Tris-HCl, pH 6.8, and 260 mM sucrose) and then used for ICM-MS analysis.

    b) Cumulus was recovered from each OOC by mechanical stripping of individual oocyte in 100µL drop of PBS. Oocyte was removed and CC were centrifuged 5 min at 3000g, then pellet was rinsed twice with TSB, and used for ICM-MS analysis.

    c) Granulosa layer was collected from the selected follicles of 5±0.2 mm in size (n=25) which were opened individually by cutting. Interior of each follicle was scrapped and GC were recovered and put into Eppendorf tube. GC were washed in 500µL PBS and centrifuged 5 min at 3000g, then GC pellets were rinsed two times with TSB with subsequent centrifugation. GC pellets with remaining 2-3µL TBS were used for ICM-MS analysis.

2. For Top-Down HR-MS proteomics, different pools of immature oocytes (n=200), mature oocytes (n=140), two groups of whole OCC (n=40 and n=300) and pools of GC recovered from 3-6 mm ovarian follicles (40 ovaries) were used. Cells were washed once with PBS, then

6

twice with TSB and after centrifugation and discarding of TSB, pellets were kept at -80°C until analysis.

Intact Cell MALDI-TOF Mass Spectrometry

Each pellet of GC and CC were mixed with 2 µL of a sinapinic acid (SA) matrix solution freshly prepared by dissolving 20 mg of the solid crystals in 1 mL of 50% acetonitrile / 50% H20 / 0.2% trifluoroacetic acid reagent (TFA) and sonicated for 10 min prior to application. 0.5 µL of follicular cells/matrix mix was spotted onto the MALDI sample probe and were immediately mixed with 1.5 µL of SA matrix. Single oocytes in 0.5 µL of TSB were individually deposited onto the MALDI sample plate; the most of TSB was removed and oocyte was covered with 0.5µL of SA matrix. The matrix/sample was allowed to evaporate slowly at room temperature for 30 min, before MALDI analysis.

All mass spectra were generated on a MALDI/LR (Waters Corporation, Micromass Ltd, Manchester, UK) MALDI-TOF mass spectrometer, operating in positive ion mode and linear MS mode. Operation of the mass spectrometer was performed using MassLynx[TM] v4.0 software, (Waters Corporation, Manchester, UK). The instrument was fitted with a nitrogen laser (wavelength 337 nm; time lag focusing (TLF) 500 ns). The voltage of the micro-channel plate detector was set at 1,950 V, matrix suppression at 500 mass units and the source voltage at 15,000 V. External calibration of the time of flight analyzer was performed by one acquisition using 1 µL of a mixture of peptides and proteins standard (1:1; v/v) with a SA matrix solution (20 mg/mL) containing 50% acetonitrile /50% H20 in presence of 0.1% TFA. The mixture contained Glu1-fibrinopeptide B, ACTH (fragment18-39), insulin and ubiquitin, all at 1 pmol/µL, 2 pmol/µL cytochrome C, 4 pmol/µL myoglobin and 8 pmol/µL trypsinogen. Spectral profiles were collected in the 1,500-25,000 m/z mass range, acquiring 10 shots per spectrum at a laser firing rate of 10 Hz. Data processing was performed using MassLynx[TM] v4.1 software. Two hundred spectra were combined. The background was subtracted from spectra obtained for each sample well using a polynomial order of 10% below the curve and smoothed with the minimum peak width at half height set for 15 channels. Two smoothings were performed using the Savitzky Golay algorithm. All spectra were processed using the same parameters. To increase mass accuracy (mass error < 0.05%), internal calibration was subsequently applied to all spectra. The latter was achieved by performing a lock mass correction with the mass corresponding to the ubiquitin protein (8,565 and 8,452 m/z) or to the thymosin β4 peptide (4,964.5 m/z). Each spectrum was converted and saved in a text file using MassLynx[TM] v4.1 software (Waters Corporation, Manchester, UK).

7

ICM-MS data processing and statistical analysis

1) For a global comparative analysis between follicular cells, MALDI-TOF spectra from immature and mature oocytes (n=25), CC before and after IVM (n=24) and GC from different follicles (n=25) were compared. A total of 74 spectra were analyzed in this study.

2) For the differential analyses on CC or oocytes between immature (imm-) an mature (mat-) states, 10 samples per condition were analysed (imm-CC vs mat-CC, imm-Oo vs mat-Oo). A total of 40 spectra were generated in this study.

3) For differential analysis on oocytes during IVM kinetics, 10 oocytes per time-point (0h, 6h, 10h, 18h, 24h) were analysed. A total 60 spectra were generated.

The spectra were analyzed by Progenesis MALDI™ version 1.2 software (NonLinear Dynamics, Newcastle upon Tyne, UK). All peak lists from each study were generated with Progenesis MALDI software and submitted to fitting of peaks for all the samples. For each cellular type, all spectra profiles were pre-processed with baseline subtraction (Top Hat filter 300), denoising (Noise filter 2) and aligned. Automatic peak detection was applied to the reference spectrum (a weighted average of all experimental spectra). Peaks were detected with a signal/background noise greater than 2 (150 counts). Normalization on peak height was performed using the Total Ionic Count (TIC) in order to display and compare all spectra on the same scale. Quantitative profiling analysis was based on the comparison of single and individual spectra generated from the 25 GC, 24 CC, 25 oocytes or from the 10 CC and 10 oocytes at different maturation stages.

The reproducibility for whole cell analysis linked directly to the biological and individual variability in addition to the spectrometer process (technical variability) was evaluated by a coefficient of variation (CV) on the 10 CC and oocyte samples for each condition (immature or mature state). CV was calculated using normalized peak height of the 10 most intense peaks. For CC, mean CV values did not exceed 44 and 39 % for immature and mature state, respectively. For oocytes, mean CV value was 35 and 36 % for immature and mature state, respectively.

Differential analysis between the groups was performed using one way analysis of variance (ANOVA) on normalized peak heights values (normalized intensity) of all the samples using Progenesis MALDI™ 1.2 software. Fold Change (FC) was calculated as a ratio between the mean normalized intensity values; maximum FC was calculated between the maximum and minimum mean values. Principal component analyses (PCA) and Hierarchical clustering were

8

performed using XLSTAT 3.01 (Addinsoft, Paris, France) software, using differential peaks (P<0.01).

For the comparative analysis of individual peaks between GC, CC and oocyte, Kruskal-Wallis statistics with post-hoc Dunn's Multiple comparison test was used (GraphPad Prism 5.0, CA). At p-value < 0.05, difference was considered significant.

For comparative analysis between immature and mature oocytes or CC, significantly different peaks were determined using t-test (Progenesis MALDI™ 1.2) and Mann-Whitney test (GraphPad Prism 5.0),

To compare ICM-MS profiles through five time points of oocyte maturation kinetics, one way ANOVA was performed to normalized peak height spectra values of the oocytes at 0h, 6h, 10h, 18h and 24h IVM (n=10 per condition). PCA was done using significantly different peaks (p < 0.01, FC> 2).

To ensure the statistical relevance of generated results, we verified that 100% of the data had a power > 0.9.

Sample preparations and pre-fractionations for TD HR-MS proteomics

Depending on the amount of biological material available, the TD HR-MS proteomics experiments, operated in discovery mode, was carried out by either i) TD direct infusion (immature OCC only), ii) direct injection on μLC-HR-MS/MS system of total protein extracts from immature OCC, isolated oocytes or cumulus cells at immature and mature state and GC, iii) by combining multiple off-line pre-fractionations of GC protein extract using RP or GF liquid chromatography methods to μLC-HR-MS/MS.

All biological samples (immature OCCs (n=40 and n=300), isolated oocytes or cumulus cells at immature (n=200) or mature state (n=140) and granulosa cells) were collected and rinsed with PBS and TSB. After frozen-thaw process, peptides and proteins were extracted by sonication using Tris-Urea (TU) buffer (6 M Urea, 50 mM Tris-HCl pH 8.8 buffer containing protease inhibitor cocktail (Sigma-Aldrich, Saint-Quentin Fallavier, France) and centrifuged during 45 minutes at 10,000 g and 4ºC. Protein concentration was determined using the DC Protein Assay (Bio-Rad, Marnes-La-Coquette, France). Protein extractions were performed using TU buffer, at the exception of immature bovine OCCs (n=40) for which we carried out an acid extraction at 5% formic acid (FA) under sonication to perform a TD HR-MS/MS analysis by direct infusion. In this case, the acid extract was desalted and concentrated using ZipTip C_4 solid-phase extraction (ZT-C_4-SPE) (Millipore Corporation, Billerica, MA) and

9

enriched, in small molecular species, by a differential elution with acetonitrile (ACN).
Fractions eluted with a solution composed of 20% ACN in water in presence of 1% FA were
dried using speedVac 1010 (Thermofisher). Micropurified species dissolved with 5µL of a
solution of 50% MetOH in water in presence of FA 1% were introduced in a metalized
nanoelectrospray needle (PicoTip emitters, New Objective) for a TD HR-MS/MS analysis by
direct infusion.

All biological models were analysed by µLC-HR-MS/MS after direct injection of around 10-
15 µg of protein total extract previously desalted and concentrated using ZT-C$_4$-SPE.

All biological models were analysed by µLC-HR-MS/MS. Around 10-15 µg of protein total
extract were desalted and concentrated using ZT-C$_4$-SPE.

Fractions eluted with a solution composed of 20% ACN in water in presence of 1% FA were
dried using speedVac 1010 (Thermofisher). Molecular species (< 5µg) were dissolved with 10
µL of a solution of FA 1%, sonicated, and directly injected (8µL).In order to reduce sample
complexity of GC protein extract, the peptides/proteins from about 0.8 g of GC (pool of
aspirated cells from different follicles) were subjected to multiple off-line pre-fractionations
through reversed phase (RP) and gel filtration (GF) chromatographic separations on an
UltiMate 3000 RSLC system controlled by Chromeleon version 6.80 SR13 software (Thermo
Electron, Courtaboeuf, France) as described in Data in Brief article [50].

Briefly, RP1, RP2 and RP3 separation methods were based on the biomolecules separation by
hydrophobicity with RP HPLC using an XBridge BEH C18 column (250 × 4.6 mm i.d.,
particule size 5 µm; Waters, Guyancourt, France). Additionally, RP3 condition was combined
to an ultracentrifugation process to enriched small proteoforms. For each condition, one mg of
peptides/proteins were injected and separated. Fourth separation (GF) consisted to separate
peptidoforms and proteoforms by molecular weight using a Superdex 75 10/300 GL gel
filtration (GF) column (GE Healthcare Europe GmbH, Velizy-Villacoublay, France).

In total, 43, 54, 46 and 40 fractions were manually collected from the RP1, RP2, RP3 and GF
liquid chromatography processes, respectively, following UV detection peaks at 214 nm.
After vacuum-drying, all fractions (at the exception of RP2 and RP3) were desalted and
enriched a second time using ZipTip C4 SPE (Millipore, Saint-Quentin-en-Yvelines, France)
and eluted with a solution of 50% ACN in presence of FA 1%. Samples were immediately
vacuum-dried and kept at -20ºC until µLC-MS/MS.

10

Top-Down HR-MS

All experiments of µLC-HR-MS/MS was performed using an Ultimate® 3000 RSLC Ultra High Pressure Liquid Chromatographer (Dionex, The Netherlands) controlled by Chromeleon Software (version 6.8 SR11; Dionex, The Netherlands). All µLC conditions are provided in Data in Brief article [50]. Briefly, each sample was loaded on a Dionex trap column (Monolithic PS-DVB PepSwift, 200 µm inner diameter x 5 mm long). Solution A was composed of 2% acetonitrile in water in presence of 0.1% formic acid, whereas solution B consisted of 16% water in acetonitrile and 0.1% formic acid. Biomolecules were pre-concentrated for 10 min at 10 µL/min with 4% solvent B. For all direct injections of protein extracts (OCC, oocytes, CC, GC), the separation was conducted using a Dionex column (Monolithic PS-DVB PepSwift, 200 µm inner diameter x 5 cm long) whereas for pre-fractionated GC samples analyses, we used an analytic column more longer (Monolithic PS-DVB PepSwift, 200 µm inner diameter x 25 cm long). The flow rate was set to 1 µL/min for both and the gradient consisted of 10-95% B for 60 min. A SilicaTip emitter with 30 µm inner diameter and 360 µm outer diameter (New Objective, Woburn, MA, USA) into a Thermo Finnigan Nanospray Ion Source 1 was used for µLC-MS.

All TD HR-MS experiments were performed using a LTQ Orbitrap Velos instrument (Thermo Fisher Scientific, Germany) operating in positive mode. Data were acquired using Xcalibur software v2.1 (Thermo Fisher Scientific, San Jose, CA). Standard mass spectrometric conditions for all analyses are provided in Data in Brief article [50].

The molecular species were analysed by FT-MS in the 400-2000 m/z mass range with a target resolution in the Orbitrap set to 100,000. After direct infusion, all major m/z were manually selected to the fragmentation using HCD (High energy Collisional Dissociation) and/or CID (Collision Induced Dissociation) modes. The selected precursor width for fragmentation was 2-3 m/z. Spectra corresponded to the accumulation of scans over approximately 1 min, yet good signal to noise ratios could be obtained within less time. For automatic data-dependent MS/MS analyses during µLC-MS, the 10 most intense ions observed after full-MS in profile mode were selected and fragmented by HCD (Higher-Energy Collisional Dissociation). Mass spectrometric parameters for MS/MS analyses are provided in Data in Brief article [50].

a) MALDI-TOF mass fingerprint of HPLC fractions

One microliter of each fraction was spotted onto a MTP Ground Steel 384 MALDI plate (Bruker Daltonics, Germany) and overlayed with 1 µL of a saturated α-cyano-4-hydroxycinnamic (CHCA) acid matrix solution containing 50% acetonitrile /50% H20 in

11

presence of 0.2% TFA. Spectra were acquired three consecutive times per spot using a Bruker UltrafleXtreme MALDI-TOF instrument (Bruker Daltonics, Germany) equipped with a Smartbeam laser at 2 kHz laser repetition rate following an automated method controlled by FlexControl 3.0 software (Bruker Daltonics, Germany). Spectra were obtained in positive linear ion mode in the m/z 1,000 – 20,000 range, and collected from each spot as a sum of 1,000 laser shots in 3 shot steps (total of 3,000 spectra per spot). The parameters used for spectra acquisition were: ion source 1, 25 kV; ion source 2, 23.55 kV; lens, 7 kV; pulsed ion extraction; and laser parameter set, large. External calibration was followed using a mixture of peptides and proteins containing Glu1-fibrinopeptide B, ACTH (fragments 18–39), insulin and ubiquitin, both at 1 pmol/µL, 2 pmol/µL cytochrome C, 4 pmol/µL myoglobin and 8 pmol/µL trypsinogen. Peak lists of single spectra were obtained using FlexAnalysis v3.4 (Bruker Daltonics, Germany) and default processing parameters. Spectral processing and gel view analysis were performed with ClinProTools v3.0 software (Bruker Daltonics, Germany). Data analysis began with an automated raw data pre-treatment workflow, comprising baseline subtraction (Top Hat, 10% minimum baseline width), normalization (total ion count), smoothing (Savitzky-Golay algorithm 2 cycle, m/z range = 15) without spectra alignment. Automatic peak detection was applied to the total average spectrum (a weighted average of all experimental spectra at 800 ppm resolution) with a signal/background noise greater than 2.

b) Top-Down HR-MS data processing and identification

Identifications and structural characterisations were performed using ProSight PC software 2.0 and 3.0 SP1 (Thermo Scientific, San Jose). Parameters for data processes and database searches are provided in Data in Brief article [50].

Briefly, for manually acquired TD HR-MS data, Raw data files were individually processed by THRASH algorithm. From PUF files (ProSight Upload Format), searches were performed using the "Absolute Mass and Biomarker" search options against the UniprotKB Swiss-Prot Bos taurus release 2015_06 database (bos_taurus_2015_06_top_down_simple.pwf). Proposed sequences with E-value < 1 x 10^{-6} were considered positively identified with a minimum of 10 matching fragment ions. Interpretation was performed with a manual, iterative process with variable parameters independently tested to maximize the E-value.

For automated data acquired by µLC-MS/MS, raw files were automatically processed inside ProSightHT of the ProSight PC software v 3.0 SP1 (Thermo Fisher, San Jose). All data files (*.raw) were processed using the cRAWler application and the THRASH algorithm. Automated searches were performed on PUF files using the "Biomarker and Absolute mass"

12

search options against the selected database. For the biomarker searches, an iterative search tree was designed to begin with high mass accuracy for monoisotopic precursors. If a top result was matched with a p-score of $\leq 1 \times 10^{-6}$ the search engine accepted this result as valid. A second search was performed for invalid results using larger intact mass tolerances. For all searches, all post-translation modifications (PTM) were considered. Then, all the *.puf files were additionally searched in "Absolute mass" mode.

For identification of endogenous peptidoforms and proteoforms, the conservative threshold for intact protein identifications by TD using the Bonferroni-corrected Poisson model was significantly exceeded (minimum E-value at $9.9\ E^{-6}$). Furthermore, according to recent TD proteomic study showing that the E-value higher than $1\ E^{-4}$ corresponded to 1% FDR threshold [51], we validated all the peptidoforms/proteoforms with E value $< 1E^{-8}$ and $> 1E^{-5}$ only after being manually controlled using Sequence Gazer (sequence with at least 4 consecutive b or y fragment ions).

Data mining and bioinformatics analyses

Raw data derived from µLC-MS/MS top-down analysis have been deposited to the ProteomeXchange Consortium [52] via the PRIDE partner repository [53] with the dataset identifier PXD004892.

In TD results lists, the redundancy was eliminated on the base of the Prosight PC observed mass (theoretical mass plus delta mass). The entry names and gene names were recovered from the UniProtKB accession numbers from *Bos taurus* annotated proteins using the Retrieve/ID mapping of Uniprot (http://www.uniprot.org/uploadlists/) and listed. Uncharacterized proteins were mapped to the corresponding *Homo sapiens* orthologues by identifying the reciprocal-best-BLAST hits using blastp program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Gene Ontology and system biology analysis were performed using list of identified proteins. Gene symbols correspondent to these proteins were analyzed for their implication in biological process, molecular functions and pathways using GeneAnalytics (http://geneanalytics.genecards.org/) and DAVID (https://david.ncifcrf.gov/) software. Also, we compared our dataset with the CASBAH database (http://bioinf.gen.tcd.ie/casbah/), which contained a comprehensive list of caspases substrates (downloaded at September 2016, 777 entries).

13

Immunoblotting analysis

Western blots were performed on groups of immature and *in vitro* mature oocytes (n=50) or correspondent CC as previously described [54]. Primary polyclonal anti-human H2B antibodies (1:500 dilution, Santa Cruz Biotechnology, CA) and anti-ubiquitin antibodies (1:400 dilution) and secondary HRP-conjugated goat anti-rabbit antibody (Cell Signaling Technology (Ozyme, Saint Quentin Yvelines, France) were used. Membranes were visualized using the ECL Plus™ Western Blotting Detection Reagent (Amersham Biosciences Orsay, France) and densitometry of specific bands was analyzed using Scion Image for Windows (Scion Corporation, Maryland, USA).

Results

Intact cell MALDI-TOF MS phenotyping of follicular cells

The ICM-MS procedures were adapted for bovine follicular cells, and molecular profiles in the range 1500–25,000 m/z were obtained using samples of bovine GC (n = 25), CC (n = 24) and single oocytes (n = 25) (Fig. 2A). Using this approach, 271, 182 and 248 m/z were detected in the mass range 1500–18,000 m/z for GC, CC and oocytes, respectively. The distribution of the detected m/z over the mass range was similar for the three types of follicular cells (Fig. 2B). A total of 439 m/z molecular species were characterised by ICM-MS, representing the basal molecular phenotype of bovine follicular cells (supplementary data, Table S1-1). The masses detected in the ICM-MS spectra were compared, with 107, 33 and 105 m/z observed only in the GC, CC or single oocytes, respectively (mass tolerance 0.1%). Of the 439 peaks, 68 (15.4%) were detected in all three cell types, whereas 23.9, 24.3 and 7.5% of the molecular species were preferentially detected in oocytes, GC and CC, respectively (Fig. 2C).

Differential analysis determined 133 peaks (ANOVA, $p < 0.001$, maximum FC >3) (supplementary data, Table S1-2). The PCA performed with the different peaks was able to clearly discriminate between GC, CC and oocyte groups (Fig. 3A).

Hierarchical clustering of 133 different peaks, presented as a heatmap (Fig. 3B), showed clusters of the peptides and proteins that were more abundant either in the oocytes (cluster 1), GC (cluster 2) or CC (cluster 3).

Examples of the normalised peak height variations are shown by scatter plots (Fig. 3C) for three proteins (m/z 4,718, 6,024 and 8,296), which were found to be significantly more abundant in oocytes, CC and GC, respectively ($p < 0.05$).

14

Top-down HR-MS proteomic analysis

In order to identify the endogenous peptidoforms and proteoforms in follicular cells by ICM-MS, TD HR-MS proteomic analyses were performed without prior chemical treatment of samples (no reduction or alkylation).

From the direct infusion of proteins prepared from bovine OCCs, manual precursor selection and CID or HCD fragmentation, 15 biomolecules corresponding to the thymosin family proteins (thymosin β-4, thymosin β-10, thymosin α-1), thioredoxin, ubiquitin, hemoglobin, actin and vimentin were identified (supplementary Table DB1-A in Data in Brief article [50])). The masses of the intact or fragmented proteins were detected in a mass range from 2,200 to 15,000 Da, which were then confidently identified using ProSight PC software with E-values ranging from 6.99×10^{-6} to 9.92×10^{-91}. Among them, nine had been previously observed in ICM-MS spectra.

In order to identify a greater number of proteins, we developed an automated TD HR-MS method using μLC-HR-MS/MS. After direct injection of total protein extract samples, we were able to identify 52 peptidoforms and proteoforms, with 17, 15 and 20 molecular species identified for oocytes, OCC and GC/CC, respectively (supplementary Table DB1-B in Data in Brief article [50])). These peptido- and proteoforms were detected in the 2,000-9,000 Da mass range, and were confidently identified with E-values ranging from 6.98×10^{-6} to 1.45×10^{-59}. Among these, 11 masses of interest were identified as potential candidates for biomarkers of oocyte maturation. However, this approach was again too limiting for overall annotation of ICM-MS spectra in the 1500-20,000 Da mass range.

To increase the number of identified proteins, pre-fractionation of the most abundant sample (pool of follicular cells containing mainly GC) was performed prior to the TD HR-MS analyses. In order to reduce sample heterogeneity, we combined four separation methods based on liquid chromatography, including reverse phase (RP1 and RP2) and gel filtration (GF) prior to μLC-MS/MS, using the same protein extract. In addition, in order to enrich the peptides and small proteins, ultracentrifugation enrichment was combined with RP fractionation (RP3).

Before the TD analysis, the 43, 54, 46 and 40 fractions collected from RP1, RP2, RP3 and GF LC processes, respectively, were analysed by MALDI-TOF MS. The MS spectra of these 183 fractions were used to generate a peak list (using Flex Analysis software) containing more than 12218 peaks (with redundancy). In order to obtain an overview of the biomolecules detected, we used ClinProTools software to generate an expanded gel view of the four

15

fractionations which showed their molecular species distribution (Fig. 4A). For the RP1, RP2, RP3 and GF fractionations, the corresponding average spectra displayed 230, 169, 174 and 185 peaks, respectively, in the mass range of 1500–20,000 m/z (Fig. 4B). Comparison of these four average profiles (mass tolerance of 0.1%) showed that the four fractionations were complementary. Among the 464 peaks detected, 115, 44, 60 and 83 m/z were specific to RP1, RP2, RP3 and GF, respectively, and only 19 were common to all fractionations (Fig. 4C). When 464 m/z species obtained by the four liquid chromatographic separations and 439 peaks obtained by ICM-MS were compared, only 136 common masses were observed (Fig. 4D). The distribution of the peaks in 1,500-15,000 Da mass range detected using fractionations was globally similar in number of molecular species with those determined by ICM-MS, however the mass range was increased to 20 kDa (Fig. 4E).

Conventional µLC-MS/MS analysis of all fractions identified 1424 molecular species. A total of 170, 49, 79 and 74 non-redundant molecular species were identified for the RP1, RP2, RP3 and GF processes, respectively, corresponding to 98, 35, 44 and 55 different UniprotKB accession numbers and 97, 33, 42 and 54 gene names, respectively (Fig. 1A in Data in Brief [50]). From these subsets, after redundancy had been suppressed, a total of 372 unique peptido- and proteoforms in the mass range of 1,000–17,000 Da were identified (approximately 25-fold more biomolecules than those identified by the direct infusion strategy) with E-values ranging from 9.96×10^{-6} to 2.87×10^{-65}. These represented 190 unique accession numbers and 173 unique genes (supplementary Table DB1-C in Data in Brief article [50])). Comparison of the biomolecules identified, accession numbers and gene names between the four fractionations showed that they were complementary for the TD approach (Fig. 1B in Data in Brief article [50]). There were no common biomolecules found between the RP1, RP2, RP3 and GF approaches, however, three accession numbers and gene names were shared by all fractionations.

Proteins identified using Top-Down HR-MS approaches

The complete inventory of endogenous peptido- and proteoforms identified in bovine follicular cells by TD HR-MS was obtained by merging the protein lists obtained from different analytical approaches, which included direct infusion, µLC-HR-MS/MS with or without prefractionation. From the 439 masses identified, 11, 39 and 316 were specific to the direct infusion, µLC-HR-MS/MS and prefractionation approaches combined to µLC-HR-MS/MS, respectively, with only two masses common to the three approaches. After suppression of redundancy, a total of 386 different molecular species were identified in a mass

16

range between 1,055 and 16,790 Da ($[M+H]^+$ theoretical average mass), with E-values ranging from 9.96×10^{-6} to 2.87×10^{-65} (supplementary Table DB1-D in Data in Brief article [50])). Regarding the mass range distribution, molecular species identified by TD showed a bias when compared with biomolecules identified from MALDI (Fig. 2A in Data in Brief article [50]). Although the overall molecular weight range was as broad as that observed for the ICM-MS spectra, the number of proteoforms identified in the range of 10–17 kDa largely decreased. However, a similar proportion of isoelectric points for the identified biomolecules was observed for the most acidic and the most alkaline species (Fig. 2B in Data in Brief article [50]).

The peptido- and proteoforms derived from the 386 identified species corresponded to 209 unique UniprotKB accession numbers and 194 unambiguously identified genes. Of these 386 species, 211 were annotated in UniprotKB (bovine taxonomy) with a reviewed status (175 unreviewed), 167 had previously been described at the protein level, 82 at the transcript level, 74 were inferred from homology and 63 were predicted (supplementary data, Table S2-D).

Among the 386 molecular species identified, 15% corresponded to whole proteins, 20% to N-terminal fragments of original proteins, 23% to C-terminal fragments of original proteins and 42% to internal fragments (Table 1). Many of the identified proteoforms cleaved from the same protein contained common N-ter or C-ter amino acids. A series of cleaved products produced from sequential truncations of either N- or C-terminal directions were also observed. Of the 386 identified biomolecules, 113 presented PTMs, including: (i) proteolytic events involving N-terminal methionine excision (101 m/z); (ii) N-terminal proteins blocked by acetylation (90 m/z) or cyclisation of the N-terminal Gln into pyroglutamate (4 m/z); and (iii) covalent modifications such as internal acetylation (5 m/z), tri-methylation (3 m/z), methionine oxidation (3 m/z), phosphorylation (3 m/z), or disulfide bridges (4 m/z).

From the 328 protein fragments (N-ter, C-ter and internal fragments), 105 corresponded to tryptic or semi-tryptic sequences (cleavages at the C-ter R or K), and the other 223 were either aliphatic (A, G, L, P) or aromatic (F) amino acid residues. To obtain an overview of the residues engaged in these cleavage sites, we noted the occurrence of the amino acids at the end of the N-ter or internal fragments, and those above the internal or C-ter fragments. The frequencies of the top five amino acids are presented in Fig. 5, with the most frequent residues found at the end and above the fragment being D, R, A, L and S. These data suggest that the biomolecules identified by TD are essentially products of protease activities

17

from trypsin-like or kallikrein enzymes and other endopeptidases targeting specific sites and/or specific substrates.

Analysis of gene ontology of the proteins identified by Top-Down HR-MS

The identified proteins (194 gene IDs) were analysed in relation to their molecular function, biological process and the pathways in which they are involved (Supplementary table DB2 in Data in Brief article [50]). The most representative functions were protein binding (126 proteins), poly(A) RNA binding (37 proteins) including mRNA 3'-UTR binding (IGF2BP3, PUM1, SERB1, YBX3 and CPEB3), protein hetero-dimerization activity (17 proteins) and cytochrome-c oxidase activity (7 proteins). The biological processes of the identified proteins included nucleosome assembly (13 proteins), chromosome organisation (13 proteins), respiratory electron transport chain (12 proteins), cellular metabolic processes (22 proteins), glucose metabolism (10 proteins) and regulation of gene expression (26 proteins). Among the most representative pathways was mitochondrial oxidative phosphorylation (12 proteins, 5.8-fold enrichment). Also, 36 proteins are involved in transport to Goldgi and subsequent modifications and 40 - in gene expression regulation. The pathways with the most significant scores were related to packaging of telomere ends and DNA double strand break response (Supplementary table DB2 in Data in Brief article [50]).

Annotation of intact cell MALDI-TOF MS peaks.

The complete dataset of 386 proteins identified by TD HR-MS (Supplementary table DB1-D in Data in Brief article [50]) was compared to ICM-MS spectral profiles of follicular cells (supplementary data, Table S1-1). The observed MALDI peaks were aligned to theoretical or observed TD HRMS masses considering their PTMs, with a mass tolerance less than 0.05% (500 ppm). A total of 247 out of 439 peaks observed in the ICM-MS spectra were identified by 136 accession numbers corresponding to masses <17 kDa. From these, 57, 58, 66 and 60 peaks were characterised from profiles specific to GC, CC, oocytes and the global follicular cell phenotype, respectively, corresponding to 59, 59, 66 and 63 accession numbers, respectively (supplementary data, Table S2). The peptido- and proteoforms derived from 91 unique UniprotKB accession numbers corresponded to 87 of the confirmed genes identified (Table 2).

Thus, our qualitative ICM-MS/TD HR-MS strategy allowed for the direct identification of 136 low weight endogenous protein and peptide species with several different PTMs, which were detectable in intact bovine follicular cells, including potential biomarkers.

18

Differential intact cell MALDI-TOF MS analysis to identify potential markers of oocyte maturation

In order to identify potential oocyte maturation biomarkers, differential analyses using ICM-MS profiling were performed on either the oocyte or the surrounding CC before and during *in vitro* maturation. Comparison of the ICM-MS spectra from single immature oocytes (n = 10) or mature oocytes after IVM (n = 10) identified 21 m/z species ($p < 0.01$, FC > 2; supplementary data, Table S3). For CC, 20 peaks differed between these groups ($p < 0.01$, FC > 2). These different species allowed for clear discrimination between immature and mature cells, both in CC (Fig. 6A) and oocytes (Fig. 6B), as shown by the PCA score plots. Examples of m/z species up- and down-regulated after IVM in CC and oocytes are shown in Fig. 6A and B, respectively. Using TD HR-MS proteomics, six peaks were significantly more abundant in CC after IVM. Of these, the peak at 4,717.15 m/z was annotated as TYB10_BOVIN (thymosin beta-10) and that at 4,964.87 m/z was TYB4_BOVIN (thymosin beta-4). On the other hand, among the 14 most abundant m/z species in immature CC, five of these were identified, including Q2KII5_BOVIN (histone H2B) at 2,885.31 m/z, F6S1Q0_BOVIN (keratin, type I cytoskeletal 18; also known as cell proliferation-inducing gene 46 protein) at 3,681.11 m/z, TYB10_BOVIN (thymosin beta-10) at 4,516.33 m/z, E1BK75_BOVIN (histone H2B) at 4,572.15 m/z and RM52_BOVIN (39S ribosomal protein L52, mitochondrial) at 11,555.11 m/z.

In the oocytes, 9 of 21 peaks were found decreased after IVM (Fig. 6B). Two of these were identified, one as F1MPB2_BOVIN (insulin-like growth factor 2 mRNA-binding protein 3, IGF2BP3) at 4,718.77 m/z and the other as G3X8G9_BOVIN (probable E3 ubiquitin-protein ligase HECTD4) at 2,764.44 m/z. Twelve peaks were found to be significantly increased in the oocytes after IVM.

After the changes throughout oocyte maturation had been determined, ICM-MS analyses were performed on the oocytes at 0, 6, 10, 18 and 24 hours of IVM. These time-points were chosen to correspond to progressive stages of oocyte meiotic maturation: germinal vesicle stage (GV), GV breakdown (GVBD), metaphase-I (Meta-I), anaphase-I (Ana-I), and metaphase-II (Meta-II), respectively. Thirty different peaks ($p < 0.05$, supplementary data S3) were used for PCA, which allowed for the discrimination of immature oocytes (0 hours IVM) not only from mature oocytes (24 hours IVM), but also from those progressing through meiotic maturation at 6, 10 and 18 hours IVM (Fig. 7A). Hierarchical clustering of the abundant species at these five IVM time-points was presented as a heat map,

19

in which different expression clusters could be distinguished (Fig. 7B). The relative abundance of four peaks, representing the four main clusters, is shown in Fig. 7C. The abundance of the protein at m/z 4,519.76 (cluster 1) was found to drop drastically at 6 hours IVM, which was then maintained for the remainder of the maturation process, whereas the protein at m/z 11,119.37 (cluster 2) progressively decreased during IVM. The expression of proteins representing cluster 3, such as m/z 8,718.57 was increased at the end of maturation (18 and 24 hours IVM), whereas those from cluster 4 (m/z 11,235.31) progressively increased in abundance during IVM.

Among the 30 peaks characterised by the IVM kinetic study, six were identified: G3X8G9_BOVIN (probable E3 ubiquitin-protein ligase HECTD4) at 2,764.44 m/z, Q2KII5_BOVIN (histone H2B) at 2885.42, TYB10_BOVIN (thymosin beta-10) at 4,519.76 m/z, TYB4_BOVIN (thymosin beta-4) at 4,965.66 m/z, UBC_BOVIN (polyubiquitin-C) at 8,452.07 m/z and HBB_BOVIN (hemoglobin subunit beta) at 8,718.57 m/z.

In order to validate the differences in proteins abundance between the immature and mature stages, evidenced by ICM-MS, we performed immunoblots to follow the change in expression in the proteins identified by TD HR-MS proteomics in the CC and oocytes before and after IVM. One example of this was for histone H2B, characterised as a fragment at 2,885.42 m/z, for which the ICM-MS analysis showed that this peak was significantly less abundant in CC after IVM when compared to immature cells ($p < 0.01$, FC>2, Fig. 8A). When quantified by western blot, the level of the whole H2B protein (~14 kDa) was also decreased after IVM ($p < 0.01$, Fig. 8B). Another protein, ubiquitin, was identified by TD at 8,565.53 m/z. As shown by ICM-MS, this molecular species was significantly upregulated during IVM in CC ($p < 0.01$, FC=1.86) but not in the oocytes (Fig. 8C). Ubiquitin was detected as a protein lower than 10 kDa in the western blot, and was less abundant in immature CC than those during IVM, whereas the abundance of ubiquitin was similar in immature and mature oocytes (Fig. 8D)

Table 2 presents a summary of the 136 identified m/z corresponding to whole proteins or fragments of original proteins. Among them, 30 m/z (represented by 19 protein IDs) were differential between different follicular cells and 16 m/z (represented by 10 protein IDs) differed between the immature and mature stage in oocytes and/or CC, and therefore, these peaks may represent markers of oocyte maturation.

20

Discussion

In the present study, we coupled TD HR-MS and ICM-MS quantitative proteomic analyses in order to determine gamete quality biomarkers at the single-cell level, using bovine oocyte maturation as a model. Oocytes are one of the largest cells in an organism, and proteomic analysis in mammals at the single-cell level has been limited by its size. Recent technological innovations, both in protein preparation and MS equipment, have led to significant advances in the oocyte proteome, for which there are now more than 2900 identified proteins in mice [18], 1000 in bovine [55] and 2154 proteins in human oocytes [24]. These studies reported proteome variations during oocyte maturation by comparing the proteins extracted from immature and mature oocyte pools, but not from single cells. However, the possible detection of more than 400 proteins from a single human oocyte has been reported [24]. Here, ICM-MS technology was applied for the first time to single oocytes and biopsies of ovarian follicular cells, followed by quantitative analysis of low molecular weight (1–17 kDa) proteins and their derivatives, which are hardly accessible by classic bottom-up proteomic approaches. A TD proteomic approach using HR-MS was employed to annotate proteins identified as potential markers of oocyte maturation. This was achieved by quantitative differential analysis of ICM-MS profiles between immature and mature oocytes.

Qualitative and quantitative intact cell MALDI-TOF MS profiling of ovarian follicular cells

Performing ICM-MS on whole intact cells has proved to be a remarkable high-throughput proteomic strategy for providing data for the global characterisation of peptido- and proteoforms with molecular weights lower than 20 kDa in different mammalian cells and cell lines [39], human blood [33] and male germinal cells [42, 43]. Moreover, the combination of ICM-MS and powerful statistical tests enables the determination of cellular heterogeneity and identification of fertility biomarkers, also reported in chicken sperm [44].

Indeed, ICM-MS allowed for the direct detection of low molecular weight peptide and protein species in a small amount of fresh or cryopreserved cells without prior sample preparation. In fact, a single oocyte and only part of the granulosa or cumulus layers, including several thousand cells, was sufficient to obtain the ICM-MS spectra. Moreover, the acidic environment of the organic matrix inhibits the enzymatic activity of most proteases, and thus, permits analysis of biomolecules in their biological context without any risk of

21

degradation due to sample preparation. In our study, the ICM-MS showed repeatable spectra, with the CV% between the samples of the same group ranging from 35 to 44% for the oocytes and CC, reflecting more biological than technical variation. This is consistent with results from studies that used ICM-MS for sperm and epithelial cells [35, 42-44].

Our study demonstrated significant differences in ICM-MS profiles between GC, CC and oocytes. This was expected due to the specific function of these cells during oocyte development [56]. In fact, 133 out of 439 m/z detected in the mass range 1,500–18,000 m/z differed between GC, CC and oocytes, and only 15% of the peaks detected were common to all follicular cells. Furthermore, as they are differentiated from the mural GC layer, CC showed a much more similar profile to GC than with oocytes. This confirms the need for cell-specific analyses when investigating the oocyte proteome, which also provides an opportunity to annotate shared peaks determined from GC/CC peptidome/proteome characterisation.

Therefore, GC, CC and oocytes have specific molecular fingerprints, which can be further compared between different follicles to look for markers in GC and CC related to proliferation and apoptosis, in addition to other non-invasive biomarkers related to the quality of the enclosed oocyte.

Differential intact cell MALDI-TOF MS analysis of bovine single oocytes and surrounding cumulus cells

In our study, we aimed to find markers of oocyte maturation. During the maturation process, oocytes resume their first meiotic division, transforming from immature oocytes (prophase nucleus, GV stage) to mature metaphase-II oocytes, which are able to be fertilised. To produce competent oocytes, meiotic maturation must be accompanied by cytoplasmic and molecular maturation, which includes protein synthesis, PTMs, and consequent modification of organelles, all of which are important for oocyte quality in terms of embryo development support [57].

In order to identify markers of oocyte maturation, quantitative analyses were performed on the ICM-MS profiles obtained from individual oocytes and surrounding CC during maturation, which revealed significant variation in several molecular species during this process. Using PCA and hierarchical clustering, the immature and mature oocytes could be clearly discriminated from the oocyte and CC profiles. Moreover, we demonstrated that the kinetics of the maturation progress was reflected by oocyte molecular fingerprints, allowing the characterisation of potential markers of the main meiotic events, such as GVBD or the transition from Meta-I to Meta-II. In fact, in bovine oocytes, protein synthesis and

22

phosphorylation occurs during IVM [58], and differs during meiosis progression [59]. For example, the m/z 4,519.76 and 11,119.37 drastically decreased from immature oocytes at the GV stage to the Meta-II stage, whereas the m/z 11,235.31 increased from the GV to Meta-I stage. The transition between the Meta-I and Meta-II stages was marked by an increase in the m/z 8,718.57 peak at the Ana-I and Meta-II stages, suggesting that these peaks could be markers of oocyte meiotic progression. The abundance of 71 m/z species were found to vary at different oocyte meiotic maturation stages, detected either in the oocytes themselves (differences in immature vs. mature oocytes, or in the maturation kinetics) or in the surrounding CC (immature vs. IVM). Expression in CC reflects the oocyte maturation status due to the tight intracellular and molecular interactions inside the oocyte-cumulus complex [60], and therefore, CC could be used for the detection of non-invasive predictive markers of oocyte quality [61]. Interestingly, several differential peaks were common to both cell types, and showed the same increase or decrease during IVM (m/z 2,358, 2,885, 4,516, 4,964. and m/z 5,171). This indicates common regulatory pathways and/or possible exchange of these molecules between the oocytes and CC, suggesting direct involvement of CC-derived factors in oocyte maturation.

Thus, ICM-MS appears to be an efficient method for the phenotyping of follicular cells, allowing for the identification of potential markers of oocyte maturation. However, endogenous biomolecules remain to be identified using this method. In order to identify these biomolecules, we choose to employ a top-down HR-MS proteomic strategy, which also considers the variety of biomolecules.

Top-down HR-MS approach to characterise the proteome and peptidome of follicular cells
The TD HR-MS proteomic approach is based on the analysis of intact proteins, allowing for direct identification of peptido- and proteoforms and structural characterisation. Therefore, this approach is more appropriate for the identification of MALDI-TOF MS signatures compared to classic bottom-up approaches using digested peptides. In order to identify the proteo- and peptidoforms detected by ICM-MS in the oocytes, we performed TD HR-MS using not only the oocytes, but also their surrounding cells, which are physically and metabolically coupled. Indeed, the ovarian follicular cells that enclose an oocyte share a significant proportion of molecular species, as previously demonstrated by comparison of the transcriptomic datasets of bovine oocytes, CC and GC [62] and proteomic analyses of oocytes and CC [8, 22]. Consistent with this, among the 248 m/z species detected by ICM-MS in the

23

oocytes in our study, 45.5% were common with GC and 39.5% with CC, therefore, it seemed pertinent to develop a TD approach for pools of follicular cells.

Given that use of HR-MS or an µLC-HR-MS/MS strategy alone is not sufficient for characterising the enormous complexity of a whole-cell proteome, we applied four different pre-fractionation methods to GC protein extracts to decrease the sample complexity and increase the dynamic range of detection, as reported previously [48]. Thus, an off-line multi-dimensional liquid chromatography approach, combining four fractionation methods based on reverse-phase or gel filtration, was coupled to µLC-HR-MS/MS, and approximately 30% (136/439) of the m/z masses represented in the follicular cell ICM-MS spectra were annotated. This resulted from combining three analytical strategies (direct infusion, direct injection to µLC-HR-MS/MS and prefractionations combined to µLC-HR-MS/MS), which were adapted for the amount of biological material available (approximately 10–50 µg for OCC, CC and oocytes, and several milligrams for GC-enriched pools). These approaches were highly complementary, as 11, 39 and 316 masses out of 386 TD identifications were specific for direct infusion, direct injection to µLC-HR-MS/MS and prefractionations combined to µLC-HR-MS/MS, respectively, and only two masses were common between the approaches. Furthermore, the four pre-fractionation methods based on reverse-phase and gel filtration chromatography were highly complementary. Of the 464 m/z detected in the average MALDI spectra of the RP1, RP2, RP3 and GF fractionations, only 19 masses were common to all methods. In addition, out of the 372 peptido- and proteoforms identified by TD HR-MS analyses of the 183 single fractions, no common masses were found between the four methods used, and only three accession numbers or gene names were shared.

In total, the TD HR-MS strategy resulted in the identification of 386 low molecular weight biomolecules (<17 kDa). Among these, we identified 219 proteins for which no evidence of protein expression was previously noted, as only detection at the transcript level or their description by homology or prediction was available. By applying our TD HR-MS workflow, we produced a list of endogenous biomolecules corresponding to whole proteins (15%), N-terminal (20%) and C-terminal (23%) fragments of original proteins, and internal fragments (42%), related to 209 unique accession numbers corresponding to 194 identified genes.

The Gene Ontology terms attributed to the corresponding proteins included the most important processes and functions of the oocyte and surrounding follicular cells in the periovulatory period [63]. In fact, proteins involved in poly(A) binding were more abundant, particularly those that bind to 3'-UTR (IGF2BP3, PUM1, SERB1, YBX3 and CPEB3),

24

corroborating the critical importance of polyadenylation and degradation of maternal RNA in bovine oocytes, which occurs during meiotic maturation [59, 64]. Furthermore, meiotic rearrangement of the chromatin from prophase to metaphase-II in the oocyte involves many proteins, several of which were identified in the current study, including different histones (H1, H2B, H3 and H4) and other factors (NPM1 and SMARCA5). In order to resume meiosis and complete meiotic divisions, the oocyte requires energy. Oocyte/cumulus energy metabolism is closely related to glucose [65], transformed to pyruvate and fatty acids by CC [66], which are both energy substrates that require mitochondrial oxidation to make ATP. Thus, the identification of proteins involved in glucose and carbohydrate metabolism (HECTD4, HSPG2, MAN2BA, MGAT4C, PPP2R5D, MDH2, GAPDH and PGAM2), fatty acid metabolism (FABP5, APOA2, APOC3, UBC and PPARGC1B), and oxidative phosphorylation (NDUFA4, UQCR10, COX7A2, UQCRC1, COX8A, COX7B, NDUFAB1, COX6A1, ATP5I, ATP5H, COX5B and COX6C) reflect these energy-providing mechanisms, which are crucial for oocyte maturation [54, 67]. Among the proteins detected, several had been reported in a recent study on human oocytes [24].

Therefore, using the TD HR-MS approach, we were able to obtain a protein dataset representative of the key pathways involved in oocyte maturation.

Top-down proteomics demonstrated post-translational modifications and degradation of proteins in follicular cells

About 30% of 386 molecules identified by TD HR-MS carried PTMs such as N-terminal protein acetylation after methionine excision, N-terminal pyroglutamate, acetylation, methylation, methionine oxidation, phosphorylation and disulfide bridges, among others. Of the N-terminus PTMs identified in 78 proteins, 55 (70%) represented acetylation, which began after methionine excision. N-terminal acetylation is a major PTM, and its functional implications include regulation of protein interactions or targeting to membranes, however, for numerous proteins the function remains unknown [68]. Recent reports have suggested potential roles as a degradation signal and the prevention of protein targeting to the secretory pathway, thus highlighting the importance of N-terminal acetylation as a major determinant for the life and death of proteins [69, 70].

Importantly, 85% (328) of the masses identified by TD HR-MS were protein fragments, and we were interested in defining the origin of these cleavages. We observed a high frequency of protein cleavages involving nine amino acid residues: R, K, D, A, G, L, F, P and S. Many of these residues are hydrophobic or polar, and represent specific targets for

25

proteases exhibiting either trypsin-like or chymotrypsin-like specificity such as kallikreins, a family of (chymo)trypsin-like serine proteinases. Other endopeptidases targeting specific sites and/or specific substrates (as metalloproteinases, cathepsin, caspases, plasmin or elastase-like) could also be involved in the generation of these fragments.

Kallikreins are known to be expressed in a wide range of tissues, including the ovary, where they have diverse physiological functions [71], as well as being involved in inflammatory processes and different pathologies, particularly endocrine-related malignancies such as ovarian cancer [72]. Some kallikreins are known to present specificity, with a strong preference for R or K as the cleavage residue, as well as A, L, F, Y, M and Q [71]. In our study, 105 out of the 328 identified protein fragments (32%) presented an R or K terminal residue, whereas 92 (28%) presented A, L, F, Y or Q residues, which supports the hypothesis of kallikrein-mediated degradation.

We observed that the D residue was frequently involved in the cleavage sites of 13% of the fragments, 26 of which were prior to the terminal residue and 18 were the terminal residues. This observation suggests that these cleaved proteins could be putative caspase substrates. Caspases are associated with the specific degradation of cellular components, controlled by signalling events which they induce [73]. These cysteine-dependent aspartate-specific proteases transmit downstream signals from the specific cleavage of key cellular components that induce a certain pathway [74]. The enzymatic properties are determined by their dominant specificity for protein substrates containing Asp (D), associated with proteolytic events of endogenous proteins related to apoptosis, involving membrane phosphatidylserine externalisation, cellular retraction, chromatin condensation and apoptotic body production. In oocyte-cumulus complexes, active caspases have been reported in CC, and the apoptosis rate in these cells was found to be an important indicator of bovine oocyte quality [75]. Moreover, caspase-3 activity in bovine granulosa cells was found to be correlated with the quality of the enclosed oocytes and with the concentration of low molecular weight IGF-binding proteins [76].

A large number of proteins have already been reported to be *in vivo* caspase substrates [77, 78]. The list of annotated substrates continues to increase, and most candidates lack functional evidence linking cleavage to a role in apoptosis. Comparison of our TD HR-MS dataset with the CASBAH database, containing a comprehensive list of caspases substrates identified by numerous proteomic studies, revealed that our list of peptidoforms included 17 known substrates. Several of these are known to be associated with specific alterations of nuclear architecture (LMNB1, TMPO, NUP153 and VIM) and nuclear RNA binding

26

(heterogeneous nuclear ribonucleoroteins A1, C, F, M and U-like), or with histone binding, proliferation and apoptosis (PTMA) [79].

Therefore, according to our dataset of biomolecules identified, ovarian follicular cells contained a greater amount of protein degradation products, and the profile of these fragments is specific to the cell types and maturation stage. Indeed, degradomic-peptidomic profiling of blood plasma was reported to be highly sensitive to physiopathological changes, and provides unique signatures that could potentially be useful for cancer diagnostics [80]. This reinforces the hypothesis that the degradomic and peptidomic data collected from CC could be predictive of oocyte quality.

Finally, our TD HR-MS approach allowed for the formal annotation of 136 molecular species corresponding to the peaks detected by ICM-MS in all three types of follicular cells. The peptido- and proteoforms obtained by TD HR-MS were annotated considering an error tolerance less than 0.05% (500 ppm). Thus, two different annotations could be attributed to one peak, which may have actually been two different species detected as one. In fact, m/z 4,717 and 4,718 were initially considered as one peak, with two identifications of similar probability: FIMPB2_BOV (IGF2BP3) and TYB10_BOV (thymosin beta-10). However, using transcriptomic data, we discovered that IGF2BP3 was over-expressed more than 1000-fold in the oocyte compared to CC and GC (not published), and was also found to be 4.5-fold more abundant in the oocytes than in CC or GC by ICM-MS differential analysis. Thus, IGF2BP3 may be considered a better marker in oocytes than in CC. This is one limitation of our ICM-MS/TD HR-MS workflow, which may be overcome by using *in situ* MALDI TD and higher resolution MS/MS in future analyses.

Although only half of these peaks were detectable in the oocyte, and more than 70% of the oocyte m/z species (182) have not yet been annotated, characterisation of ICM-MS spectra using TD HR-MS allowed us to identify several valuable markers involved in oocyte maturation.

Markers of oocyte maturation

Sixteen potential m/z markers identified by ICM-MS were annotated using TD HR-MS, corresponding to either whole low molecular weight proteins or to peptide fragments derived from larger proteins. From comparative analysis of the oocytes at different stages of IVM, we generated a list of eight proteins and fragments which showed a difference in abundance in the oocytes at metaphase-II (mature oocyte) or other specific stages of oocyte meiotic progression compared to immature oocytes (Table 2). Among them, IGF2BP3, ubiquitin-protein ligase

27

HECTD4, hemoglobin beta and cysteine hydrolase CMBL were all specific to oocytes, whereas histone H2B and H2B-like, thymosin beta-4 and thymosin beta-10 differed in abundance between oocytes and CC. This list was completed by CC-specific markers, including fragments of mitochondrial ribosomal protein MRPL52, ribosomal RPL39 and cytoskeletal keratin 18, the abundance of which differed in CC surrounding immature or mature oocytes.

The thymosins TMBX4 and TMSB10 were more abundant in CC after IVM, consistent with the significant over-expression of their corresponding genes during IVM, reported in bovine CC [81]. Described as anti-apoptotic and anti-inflammatory factors, these actin-binding proteins participate in various intracellular activities and signalling pathways related to cytoskeleton remodelling [82], therefore, they could potentially be involved in cumulus expansion and secretory activities, which are important for oocyte fertilisation.

Polyubiquitin was identified, and the differential ICM-MS analysis was validated by western blot, which showed the same patterns as by ICM-MS, in both CC and oocytes. The importance of the ubiquitin-proteasome pathway for cumulus expansion and oocyte meiosis has been demonstrated in porcine species [83], and therefore, the greater abundance of ubiquitin in expanded mature cumulus compared to compact immature cumulus appears to be relevant.

The most abundant proteins identified by TD HR-MS were histone H2B and H2B-like (5 m/z and 7 m/z, respectively). Furthermore, from western blot quantification of the 14 kDa histone H2B, we demonstrated the difference in relative abundance of the fragment peptide (m/z 2,825, annotated as H2B), as evidenced by ICM-MS, may be associated with similar variations in the corresponding whole protein.

In this study, IGF2BP3 (also known as IMP3) was identified as the acetylated N-terminal fragment (1–43) of a predicted protein (F1MPB2_BOVIN), which was less abundant in mature oocytes. In zebrafish eggs, IMP3 was shown to represses the translation of maternal cyclin B1 mRNA in immature oocytes, and its release from 3'-UTR of *CCBN1* mRNA triggered the translation of cyclin B1, and thus, meiosis. Moreover, IMP3 was shown to interact with the Pumilio and cytoplasmic polyadenylation-element (CPE) binding (CPEB) proteins [84]. Similar to fish, *de novo* synthesis of cyclin B1 is essential for the initiation of bovine oocyte meiotic maturation [85], which requires activation of maturation-promoting factor (MPF), a protein kinase consisting of the catalytic subunit Cdc2 and regulatory subunit cyclin B1. CPEB is one of the main regulators of the translation of cyclin B1 and other maternal mRNAs that contain CPEs in their 3'-UTR during meiosis. Pumilio 1 is also a well-

28

known regulator of mRNA polyadenylation, which interacts with CPEB proteins and binds in their proximity on many mRNAs [86]. Previously, we described a spatio-temporal expression pattern of the CPEB protein in bovine oocytes and showed its degradation in mature oocytes, as well as an increase in the cyclin B1 protein [87]. Here, internal fragments of CPEB3 and Pumilio 1 proteins were also identified by our TD HR-MS workflow, suggesting that combined with IGF2BP3, these form part of associated proteins or complexes in the oocytes. In the present study, the IGFBP2 fragment was found to be more abundant in the immature stage, suggesting either a decrease of its synthesis or more active degradation of this protein during IVM, similar to the decrease in CPEB. IGF2BP3 could be a caspase substrate due to the presence of the D amino acid residue at the C-terminal position of the TD-identified fragment, similar to cyclins A1, A2, E and T2, which are cleaved by caspases [88, 89].

Interestingly, using ICM-MS, we detected that the abundance of hemoglobin HBB (m/z 3,556) differed between follicular cells, as the abundance was 4.5-fold higher in the oocytes than in CC or GC. Moreover, a significant increase in HBB was found when oocytes transitioned from the GVBD to Meta-II stages. Indeed, expression of the *HBA* and *HBB* genes was recently shown in mice and human CC, and the corresponding proteins were detected in CC and oocytes, suggesting that hemoglobin may act as a molecule that sequesters O_2 or NO to provide a short-lived hypoxic environment to the oocyte before ovulation [90]. Moreover, the intra-oocyte level of hemoglobin was shown to be higher in more competent mature oocytes *in vivo* than in less competent oocytes after IVM, supporting the hypothesis that hemoglobin is crucial for oocyte quality.

Therefore, the possibility of detecting and quantifying markers using ICM-MS in cumulus cells suggests that this may be a valuable non-invasive method for predicting oocyte quality. Detection of such markers at the single-oocyte level can help to improve IVM conditions, and thus, increase the efficiency of *in vitro* embryo production.

Conclusion

Intact cell MALDI MS profiling was adapted for bovine ovarian follicular cells at the single-follicle level, demonstrating cell-specific protein and peptide signatures. Differential analyses on CC or enclosed single oocytes allowed for the determination of endogenous molecules that varied in abundance during *in vitro* maturation. The use of complementary TD analytical strategies, including pre-fractionations followed by HR-MS/MS, allowed for the identification of 386 peptidoforms and proteoforms. These included both low molecular weight proteins and

29

cleaved products of larger proteins, including those that had not previously been described in ovarian follicular cells at the protein level. Moreover, PTMs and degradation of these proteins was demonstrated. Furthermore, the identification of peaks in the ICM-MS spectra allowed for the identification of particular proteins whose relative abundance may be associated with oocyte meiotic maturation and quality.

Index

OCC : Oocyte-Cumulus cells complexe; CC : Cumulus Cells; GC: Granulosa Cells; ICM-MS: Intact cell MALDI-TOF Mass Spectrometry; TD: Top-Down; HR-MS/MS: High Resolution tandem Mass spectrometry; µLC: micro-Liquid Chromatography; IVM: In Vitro maturation.

Acknowledgements

30

Figure legends

Fig. 1. Workflow resuming strategies from ICM-MS qualitative and quantitative profiling to Top-Down HR-MS/MS

Fig. 2. MALDI-TOF spectra of intact bovine granulosa cells (GC), cumulus cells (CC) and single oocytes, as well as the distribution of peaks for the mass ranges of the three different cell types. (A) Three spectra ranging from 3,000 to 20,000 m/z, illustrating the representative molecular phenotypes obtained from GC, CC and single oocytes. (B) Distribution of the m/z fragments detected over the mass range for GC, CC and oocytes. (C) Venn diagram for comparison of all masses detected from the intact cell MALDI-TOF MS spectra of the three types of follicular cells.

Fig. 3. Comparative analysis of abundance of the different peaks determined by ICM-MS in ovarian follicular cells. This analysis included 133 peaks that presented at least one significant difference ($p < 0.001$; fold-change >3) between the samples of three cell types, granulosa cells (GC, n=24), cumulus cells (CC, n=25) and single oocytes (n=25). (A) Scatter plot of PCA scores from ICM-MS analysis of single oocytes, GC and CC. (B) Heat map representation of differential peaks between oocytes, CC and GC. (C) Relative abundance of three differential peaks between the oocytes and CC and GC samples.

Fig. 4. Masses detected for the RP1, RP2, RP3 and GF fractionations, and comparison of the peaks observed from MALDI-TOF MS on LC fractions and intact cells. (A) The analytical gels of the MALDI-TOF spectra obtained from each fraction, obtained using the four fractionation methods (RP1, RP2, RP3 and GF), showing typical separation of peptides and proteins and the distribution of biomolecules in the 2,000–20,000 mass range. (B) The average spectra revealed 230, 169, 174 and 185 peaks for RP1, RP2, RP3 and GF fractionations, respectively. (C) Venn diagram showing the complementary peaks obtained from the four approaches for the 464 peaks detected from the global average spectra (combining RP1, RP2, RP3 and GF). (D) Of the 464 and 439 peaks detected in the fractions and intact cell MALDI-TOF MS (ICM-MS) spectra, respectively, the Venn diagram shows that the 136 common peaks can potentially be identified by top-down high-resolution MS and used to annotate ICM-MS spectra. (E) The representation of the number of peaks detected within the 1,500–20,000 Da mass range showing good distribution within this range.

31

Fig. 5. Graphical representation of the occurrence of amino acid residues involved in protein cleavages generating N-terminal, C-terminal or internal fragments. The frequency of the top five residues found at the end and above the 328 peptide fragments generated from original proteins is presented. Most of the cleavages involved arginine (R) and lysine (K) residues, aliphatic amino acids (alanine (A), glycine (G), leucine (L) and proline (P)) or aromatic amino acids (phenylalanine (F)).

Fig. 6. Principal component analysis (PCA) and relative intensity of the most abundant peaks in (A) immature and mature cumulus cells (CC) and (B) the corresponding oocytes .

Fig. 7. Intact cell MALDI-TOF MS (ICM-MS) analysis of *in vitro* oocyte maturation kinetics. (A) Schematic of nuclear stages during the oocyte meiotic maturation process and the corresponding *in vitro* maturation (IVM) time-points (0, 6, 10, 18 and 24 hours). Principal component analysis (PCA) of the significant peaks detected in the oocytes at different IVM time-points. (B) Heat map representation of hierarchical clustering, using the relative peak intensity values for the different peaks during IVM. (C) Graphical representation of abundance kinetics during IVM for several peaks. Annotation of m/z 8,718 (HBB, haemoglobin B) and m/z 4,519 (TMSB10, thymosin) was carried out according to that observed in the top-down high-resolution MS analysis.

Fig. 8. Validation of differential intact cell MALDI-TOF MS (ICM-MS) analysis by western blot. (A) Relative abundance of histone H2B (m/z 2,885) in CC samples before (immature, imm) or after *in vitro* maturation (mature, IVM), as detected by ICM-MS. (B) Western blot quantification of the H2B protein in cumulus cells (CC) before and after 24 hours IVM. Four independent CC samples were analysed (10 µg of protein loaded per sample). (C) Relative abundance of ubiquitin (m/z 8,565) in CC samples and in individual oocytes before (imm) or after IVM, as detected by ICM-MS. (D) Western blot detection of ubiquitin in CC and in oocytes before (imm) and after IVM (10 µg of CC protein and 50 oocytes were loaded per lane). Vinculin was analysed as the internal loading control. The asterisk (*) represents a significant difference ($p < 0.01$). For Western blot, red ponceau coloration of the membranes and correspondent images of chimioluminescence protein detection are shown.

32

Table legends

Table 1. Number of endogenous peptidoforms and proteoforms with or without post-translational modifications identified by top-down high-resolution MS in bovine follicular cells.

The number of whole proteins and peptides as either N-terminal, C-terminal or internal fragments was identified using ProSight PC 3.0 software. The number of identifications (IDs), total number of characterised post-translational modifications (PTM) and their distributions are presented for proteoforms and peptidoforms.

Table 2. List of the peptides/proteins detected by intact cell MALDI-TOF MS and identified by top-down HR-MS/MS in bovine ovarian follicular cells. Markers were found from ICM-MS profiling analysis of follicular compartments (F), oocytes (Oo) and cumulus cells (CC). Potential markers of oocyte maturation are shaded grey.

33

References

[1] M.A. Sirard, Follicle environment and quality of in vitro matured oocytes, J Assist Reprod Genet (2011).

[2] R. Buccione, A.C. Schroeder, J.J. Eppig, Interactions between somatic cells and germ cells throughout mammalian oogenesis, Biology of reproduction 43(4) (1990) 543-7.

[3] E. Seli, C. Robert, M.A. Sirard, OMICS in assisted reproduction: possibilities and pitfalls, Mol Hum Reprod 16(8) (2010) 513-30.

[4] D. Cohen, J.A. Dickerson, C.D. Whitmore, E.H. Turner, M.M. Palcic, O. Hindsgaul, N.J. Dovichi, Chemical cytometry: fluorescence-based single-cell analysis, Annu Rev Anal Chem (Palo Alto Calif) 1 (2008) 165-90.

[5] S. Hashimoto, K. Saeki, Y. Nagao, N. Minami, M. Yamada, K. Utsumi, Effects of cumulus cell density during in vitro maturation of the developmental competence of bovine oocytes, Theriogenology 49(8) (1998) 1451-63.

[6] D.K. Lobb, S.R. Soliman, S. Daya, E.V. Younglai, Steroidogenesis in luteinized granulosa cell cultures varies with follicular priming regimen, Hum Reprod 13(8) (1998) 2064-7.

[7] A. Gougeon, Regulation of ovarian follicular development in primates: facts and hypotheses, Endocrine reviews 17(2) (1996) 121-55.

[8] D. Peddinti, E. Memili, S.C. Burgess, Proteomics-based systems biology modeling of bovine germinal vesicle stage oocyte and cumulus cell interaction, PloS one 5(6) (2010) e11240.

[9] E. Memili, D. Peddinti, L.A. Shack, B. Nanduri, F. McCarthy, H. Sagirkaya, S.C. Burgess, Bovine germinal vesicle oocyte and cumulus cell proteomics, Reproduction 133(6) (2007) 1107-20.

[10] M. Bhojwani, E. Rudolph, W. Kanitz, H. Zuehlke, F. Schneider, W. Tomek, Molecular analysis of maturation processes by protein and phosphoprotein profiling during in vitro maturation of bovine oocytes: a proteomic approach, Cloning and stem cells 8(4) (2006) 259-74.

[11] A. Bettegowda, O.V. Patel, K.B. Lee, K.E. Park, M. Salem, J. Yao, J.J. Ireland, G.W. Smith, Identification of novel bovine cumulus cell molecular markers predictive of oocyte competence: functional and diagnostic implications, Biology of reproduction 79(2) (2008) 301-9.

[12] F.J. Berendt, T. Frohlich, P. Bolbrinker, M. Boelhauve, T. Gungor, F.A. Habermann, E. Wolf, G.J. Arnold, Highly sensitive saturation labeling reveals changes in abundance of cell cycle-associated proteins and redox enzyme variants during oocyte maturation in vitro, Proteomics 9(3) (2009) 550-64.

[13] L. Chen, L. Zhai, C. Qu, C. Zhang, S. Li, F. Wu, Y. Qi, F. Lu, P. Xu, X. Li, D. Shi, Comparative Proteomic Analysis of Buffalo Oocytes Matured in vitro Using iTRAQ Technique, Scientific reports 6 (2016) 31795.

[14] Z. Ellederova, P. Halada, P. Man, M. Kubelka, J. Motlik, H. Kovarova, Protein patterns of pig oocytes during in vitro maturation, Biology of reproduction 71(5) (2004) 1533-9.

[15] M.D. Powell, G. Manandhar, L. Spate, M. Sutovsky, S. Zimmerman, S.C. Sachdev, M. Hannink, R.S. Prather, P. Sutovsky, Discovery of putative oocyte quality markers by comparative ExacTag proteomics, Proteomics. Clinical applications 4(3) (2010) 337-51.

[16] J. Kim, J.S. Kim, Y.J. Jeon, D.W. Kim, T.H. Yang, Y. Soh, H.K. Lee, N.J. Choi, S.B. Park, K.S. Seo, H.M. Chung, D.S. Lee, J.I. Chae, Identification of maturation and protein synthesis related proteins from porcine oocytes during in vitro maturation, Proteome science 9 (2011) 28.

[17] A.M. Vitale, M.E. Calvert, M. Mallavarapu, P. Yurttas, J. Perlin, J. Herr, S. Coonrod, Proteomic profiling of murine oocyte maturation, Molecular reproduction and development 74(5) (2007) 608-16.

34

[18] S. Wang, Z. Kou, Z. Jing, Y. Zhang, X. Guo, M. Dong, I. Wilmut, S. Gao, Proteome of mouse oocytes at different developmental stages, Proceedings of the National Academy of Sciences of the United States of America 107(41) (2010) 17639-44.

[19] S. Cao, X. Guo, Z. Zhou, J. Sha, Comparative proteomic analysis of proteins involved in oocyte meiotic maturation in mice, Molecular reproduction and development 79(6) (2012) 413-22.

[20] S. Hamamah, V. Matha, C. Berthenet, T. Anahory, V. Loup, H. Dechaud, B. Hedon, A. Fernandez, N. Lamb, Comparative protein expression profiling in human cumulus cells in relation to oocyte fertilization and ovarian stimulation protocol, Reproductive biomedicine online 13(6) (2006) 807-14.

[21] I. Virant-Klun, J. Krijgsveld, Proteomes of animal oocytes: what can we learn for human oocytes in the in vitro fertilization programme?, BioMed research international 2014 (2014) 856907.

[22] S. McReynolds, M. Dzieciatkowska, B.R. McCallie, S.D. Mitchell, J. Stevens, K. Hansen, W.B. Schoolcraft, M.G. Katz-Jaffe, Impact of maternal aging on the molecular signature of human cumulus cells, Fertility and sterility 98(6) (2012) 1574-80 e5.

[23] C.S. Hughes, S. Foehr, D.A. Garfield, E.E. Furlong, L.M. Steinmetz, J. Krijgsveld, Ultrasensitive proteome analysis using paramagnetic bead technology, Molecular systems biology 10 (2014) 757.

[24] I. Virant-Klun, S. Leicht, C. Hughes, J. Krijgsveld, Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes, Molecular & cellular proteomics : MCP 15(8) (2016) 2616-27.

[25] M.A. Claydon, S.N. Davey, V. Edwards-Jones, D.B. Gordon, The rapid identification of intact microorganisms using mass spectrometry, Nature biotechnology 14(11) (1996) 1584-6.

[26] R.D. Holland, J.G. Wilkes, F. Rafii, J.B. Sutherland, C.C. Persons, K.J. Voorhees, J.O. Lay, Jr., Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry, Rapid communications in mass spectrometry : RCM 10(10) (1996) 1227-32.

[27] T. Krishnamurthy, P.L. Ross, Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells, Rapid communications in mass spectrometry : RCM 10(15) (1996) 1992-6.

[28] G. Vogel, A. Strauss, B. Jenni, D. Ziegler, E. Dumermuth, S. Antz, C. Bardouille, B. Wipf, C. Miscenic, G. Schmid, V. Pfluger, Development and validation of a protocol for cell line identification by MALDI-TOF MS, BMC proceedings 5 Suppl 8 (2011) P45.

[29] A. Karger, B. Bettin, M. Lenk, T.C. Mettenleiter, Rapid characterisation of cell cultures by matrix-assisted laser desorption/ionisation mass spectrometric typing, Journal of virological methods 164(1-2) (2010) 116-21.

[30] X. Zhang, M. Scalf, T.W. Berggren, M.S. Westphall, L.M. Smith, Identification of mammalian cell lines using MALDI-TOF and LC-ESI-MS/MS mass spectrometry, Journal of the American Society for Mass Spectrometry 17(4) (2006) 490-9.

[31] R. Ouedraogo, C. Flaudrops, A. Ben Amara, C. Capo, D. Raoult, J.L. Mege, Global analysis of circulating immune cells by matrix-assisted laser desorption ionization time-of-flight mass spectrometry, PloS one 5(10) (2010) e13691.

[32] J. Hanrieder, G. Wicher, J. Bergquist, M. Andersson, A. Fex-Svenningsen, MALDI mass spectrometry based molecular phenotyping of CNS glial cells for prediction in mammalian brain tissue, Analytical and bioanalytical chemistry 401(1) (2011) 135-47.

[33] B. Munteanu, C. von Reitzenstein, G.M. Hansch, B. Meyer, C. Hopf, Sensitive, robust and automated protein analysis of cell differentiation and of primary human blood cells by intact cell MALDI mass spectrometry biotyping, Analytical and bioanalytical chemistry 404(8) (2012) 2277-86.

35

[34] R. Ouedraogo, A. Daumas, E. Ghigo, C. Capo, J.L. Mege, J. Textoris, Whole-cell MALDI-TOF MS: a new tool to assess the multifaceted activation of macrophages, Journal of proteomics 75(18) (2012) 5523-32.

[35] L.F. Marvin-Guy, P. Duncan, S. Wagniere, N. Antille, N. Porta, M. Affolter, M. Kussmann, Rapid identification of differentiation markers from whole epithelial cells by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry and statistical analysis, Rapid communications in mass spectrometry : RCM 22(8) (2008) 1099-108.

[36] S.L. Kober, H. Meyer-Alert, D. Grienitz, H. Hollert, M. Frohme, Intact cell mass spectrometry as a rapid and specific tool for the differentiation of toxic effects in cell-based ecotoxicological test systems, Analytical and bioanalytical chemistry 407(25) (2015) 7721-31.

[37] N.H. Chiu, Z. Jia, R. Diaz, P. Wright, Rapid differentiation of in vitro cellular responses to toxic chemicals by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Environmental toxicology and chemistry / SETAC 34(1) (2015) 161-6.

[38] B. Munteanu, B. Meyer, C. von Reitzenstein, E. Burgermeister, S. Bog, A. Pahl, M.P. Ebert, C. Hopf, Label-free in situ monitoring of histone deacetylase drug target engagement by matrix-assisted laser desorption ionization-mass spectrometry biotyping and imaging, Analytical chemistry 86(10) (2014) 4642-7.

[39] J.F. Povey, C.J. O'Malley, T. Root, E.B. Martin, G.A. Montague, M. Feary, C. Trim, D.A. Lang, R. Alldread, A.J. Racher, C.M. Smales, Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling, Journal of biotechnology 184 (2014) 84-93.

[40] H.T. Feng, L.C. Sim, C. Wan, N.S. Wong, Y. Yang, Rapid characterization of protein productivity and production stability of CHO cells by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Rapid communications in mass spectrometry : RCM 25(10) (2011) 1407-12.

[41] H.T. Feng, N.S. Wong, L.C. Sim, L. Wati, Y. Ho, M.M. Lee, Rapid characterization of high/low producer CHO cells using matrix-assisted laser desorption/ionization time-of-flight, Rapid communications in mass spectrometry : RCM 24(9) (2010) 1226-30.

[42] V. Labas, L. Spina, C. Belleannee, A.P. Teixeira-Gomes, A. Gargaros, F. Dacheux, J.L. Dacheux, Analysis of epididymal sperm maturation by MALDI profiling and top-down mass spectrometry, Journal of proteomics 113 (2015) 226-43.

[43] V. Labas, I. Grasseau, K. Cahier, A. Gargaros, G. Harichaux, A.P. Teixeira-Gomes, S. Alves, M. Bourin, N. Gerard, E. Blesbois, Qualitative and quantitative peptidomic and proteomic approaches to phenotyping chicken semen, Journal of proteomics 112 (2015) 313-35.

[44] L. Soler, V. Labas, A. Thelie, I. Grasseau, A.P. Teixeira-Gomes, E. Blesbois, Intact Cell MALDI-TOF MS on Sperm: A Molecular Test For Male Fertility Diagnosis, Molecular & cellular proteomics : MCP 15(6) (2016) 1998-2010.

[45] N.L. Kelleher, From primary structure to function: biological insights from large-molecule mass spectra, Chemistry & biology 7(2) (2000) R37-45.

[46] F.W. McLafferty, E.K. Fridriksson, D.M. Horn, M.A. Lewis, R.A. Zubarev, Techview: biochemistry. Biomolecule mass spectrometry, Science 284(5418) (1999) 1289-90.

[47] N.L. Kelleher, Top-down proteomics, Analytical chemistry 76(11) (2004) 197A-203A.

[48] J.F. Kellie, J.C. Tran, J.E. Lee, D.R. Ahlf, H.M. Thomas, I. Ntai, A.D. Catherman, K.R. Durbin, L. Zamdborg, A. Vellaichamy, P.M. Thomas, N.L. Kelleher, The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput, Molecular bioSystems 6(9) (2010) 1532-9.

[49] S. Auclair, R. Uzbekov, S. Elis, L. Sanchez, I. Kireev, L. Lardic, R. Dalbies-Tran, S. Uzbekova, Absence of cumulus cells during in vitro maturation affects lipid metabolism in

36

bovine oocytes, American journal of physiology. Endocrinology and metabolism 304(6) (2013) E599-613.

[50] V. Labas, A. Teixeira-Gomes, L. Bouguereau, A. Gargaros, L. Spina, A. Marestaing, S. Uzbekova, Data on endogenous bovine follicular cells peptides and small proteins obtained through Top-down High Resolution Mass Spectrometry, Data in Brief (2017).

[51] A.D. Catherman, K.R. Durbin, D.R. Ahlf, B.P. Early, R.T. Fellers, J.C. Tran, P.M. Thomas, N.L. Kelleher, Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence, Molecular & cellular proteomics : MCP 12(12) (2013) 3465-73.

[52] J.A. Vizcaino, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.J. Kraus, J.P. Albar, S. Martinez-Bartolome, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nature biotechnology 32(3) (2014) 223-6.

[53] J.A. Vizcaino, R.G. Cote, A. Csordas, J.A. Dianes, A. Fabregat, J.M. Foster, J. Griss, E. Alpi, M. Birim, J. Contell, G. O'Kelly, A. Schoenegger, D. Ovelleiro, Y. Perez-Riverol, F. Reisinger, D. Rios, R. Wang, H. Hermjakob, The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013, Nucleic acids research 41(Database issue) (2013) D1063-9.

[54] L. Sanchez-Lazo, D. Brisard, S. Elis, V. Maillard, R. Uzbekov, V. Labas, A. Desmarchais, P. Papillier, P. Monget, S. Uzbekova, Fatty acid synthesis and oxidation in cumulus cells support oocyte maturation in bovine, Molecular Endocrinology 28(9) (2014) 1502-1521.

[55] D.R. Deutsch, T. Frohlich, K.A. Otte, A. Beck, F.A. Habermann, E. Wolf, G.J. Arnold, Stage-specific proteome signatures in early bovine embryo development, Journal of proteome research 13(10) (2014) 4363-76.

[56] M.L. Hennet, C.M. Combelles, The antral follicle: a microenvironment for oocyte differentiation, The International journal of developmental biology 56(10-12) (2012) 819-31.

[57] M.A. Sirard, F. Richard, P. Blondin, C. Robert, Contribution of the oocyte to embryo quality, Theriogenology 65(1) (2006) 126-36.

[58] W. Tomek, H. Torner, W. Kanitz, Comparative analysis of protein synthesis, transcription and cytoplasmic polyadenylation of mRNA during maturation of bovine oocytes in vitro, Reprod Domest Anim 37(2) (2002) 86-91.

[59] J.M. Traverso, I. Donnay, A.S. Lequarre, Effects of polyadenylation inhibition on meiosis progression in relation to the polyadenylation status of cyclins A2 and B1 during in vitro maturation of bovine oocytes, Molecular reproduction and development 71(1) (2005) 107-14.

[60] C. Charlier, J. Montfort, O. Chabrol, D. Brisard, T.T.V. Nguyen, A. Le Cam, L. Richard-Parpaillon, F. Moreews, P. Pontarotti, S. Uzbekova, F. Chesnel, J. Bobe, Oocyte-somatic cells interactions, lessons from evolution, BMC Genomics 13:560 (2012) 1-18.

[61] V. Puard, T. Tranchant, V. Cadoret, C. Gauthier, E. Reiter, F. Guerif, D. Royere, Semi-quantitative measurement of specific proteins in human cumulus cells using reverse phase protein array, Reproductive biology and endocrinology : RB&E 11 (2013) 100.

[62] D.R. Khan, D.A. Landry, E. Fournier, C. Vigneault, P. Blondin, M.A. Sirard, Transcriptome meta-analysis of three follicular compartments and its correlation with ovarian follicle maturity and oocyte developmental competence in cows, Physiological genomics 48(8) (2016) 633-43.

[63] A.R. Krauchunas, M.F. Wolfner, Molecular changes during egg activation, Curr Top Dev Biol 102 267-92.

[64] C. Krischek, B. Meinecke, In vitro maturation of bovine oocytes requires polyadenylation of mRNAs coding proteins for chromatin condensation, spindle assembly, MPF and MAP kinase activation, Anim Reprod Sci 73(3-4) (2002) 129.

[65] M.L. Sutton-McDowall, R.B. Gilchrist, J.G. Thompson, The pivotal role of glucose metabolism in determining oocyte developmental competence, Reproduction 139(4) (2010) 685-95.

[66] K. Dunning, D.L. Russell, R. Robker, Lipids and oocyte developmental competence: the role of fatty acids and B-oxidation, Reproduction (2014).

[67] N. Songsasen, Energy metabolism regulating mammalian oocyte maturation, in: Meiosis - Molecular Mechanisms and Cytogenetic Diversity, edited by Andrew Swan. ISBN: 978-953-51-0118-5, InTech, DOI: 10.5772/33752. Available from: http://www.intechopen.com/books/meiosis-molecular-mechanisms-and-cytogenetic-diversity/energy-metabolism-regulating-mammalian-oocyte-maturation (2012) 173-186.

[68] T. Arnesen, Towards a functional understanding of protein N-terminal acetylation, PLoS biology 9(5) (2011) e1001074.

[69] C.S. Hwang, A. Shemorry, A. Varshavsky, N-terminal acetylation of cellular proteins creates specific degradation signals, Science 327(5968) (2010) 973-7.

[70] A. Varshavsky, Discovery of cellular regulation by protein degradation, The Journal of biological chemistry 283(50) (2008) 34469-89.

[71] M. Debela, V. Magdolen, N. Schechter, M. Valachova, F. Lottspeich, C.S. Craik, Y. Choe, W. Bode, P. Goettig, Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences, The Journal of biological chemistry 281(35) (2006) 25678-88.

[72] C.A. Borgono, I.P. Michael, E.P. Diamandis, Human tissue kallikreins: physiologic roles and applications in cancer, Molecular cancer research : MCR 2(5) (2004) 257-80.

[73] D.R. McIlwain, T. Berger, T.W. Mak, Caspase functions in cell death and disease, Cold Spring Harbor perspectives in biology 5(4) (2013) a008656.

[74] C. Pop, G.S. Salvesen, Human caspases: activation, specificity, and regulation, The Journal of biological chemistry 284(33) (2009) 21777-81.

[75] Y.Q. Yuan, A. Van Soom, J.L. Leroy, J. Dewulf, A. Van Zeveren, A. de Kruif, L.J. Peelman, Apoptosis in cumulus cells, but not in oocytes, may influence bovine embryonic developmental competence, Theriogenology 63(8) (2005) 2147-63.

[76] B. Nicholas, R. Alberio, A.A. Fouladi-Nashta, R. Webb, Relationship between low-molecular-weight insulin-like growth factor-binding proteins, caspase-3 activity, and oocyte quality, Biology of reproduction 72(4) (2005) 796-804.

[77] R.C. Taylor, G. Brumatti, S. Ito, M.O. Hengartner, W.B. Derry, S.J. Martin, Establishing a blueprint for CED-3-dependent killing through identification of multiple substrates for this protease, The Journal of biological chemistry 282(20) (2007) 15011-21.

[78] A.U. Luthi, S.J. Martin, The CASBAH: a searchable database of caspase substrates, Cell death and differentiation 14(4) (2007) 641-50.

[79] K.P. Letsas, M. Frangou-Lazaridis, Surfing on prothymosin alpha proliferation and anti-apoptotic properties, Neoplasma 53(2) (2006) 92-6.

[80] Y. Shen, N. Tolic, T. Liu, R. Zhao, B.O. Petritis, M.A. Gritsenko, D.G. Camp, R.J. Moore, S.O. Purvine, F.J. Esteva, R.D. Smith, Blood peptidome-degradome profile of breast cancer, PloS one 5(10) (2010) e13133.

[81] M. Salhab, P. Papillier, C. Perreau, C. Guyader-Joly, J. Dupont, P. Mermillod, S. Uzbekova, Thymosins beta-4 and beta-10 are expressed in bovine ovarian follicles and upregulated in cumulus cells during meiotic maturation, Reproduction Fertility and Development 22(8) (2010) 1206-1221.

[82] L. Renault, Intrinsic, Functional, and Structural Properties of beta-Thymosins and beta-Thymosin/WH2 Domains in the Regulation and Coordination of Actin Self-Assembly Dynamics and Cytoskeleton Remodeling, Vitamins and hormones 102 (2016) 25-54.

38

[83] Y.J. Yi, E. Nagyova, G. Manandhar, R. Prochazka, M. Sutovsky, C.S. Park, P. Sutovsky, Proteolytic activity of the 26S proteasome is required for the meiotic resumption, germinal vesicle breakdown, and cumulus expansion of porcine cumulus-oocyte complexes matured in vitro, Biology of reproduction 78(1) (2008) 115-26.

[84] K. Takahashi, T. Kotani, Y. Katsu, M. Yamashita, Possible involvement of insulin-like growth factor 2 mRNA-binding protein 3 in zebrafish oocyte maturation as a novel cyclin B1 mRNA-binding protein that represses the translation in immature oocytes, Biochemical and biophysical research communications 448(1) (2014) 22-7.

[85] J.T. Levesque, M.A. Sirard, Resumption of meiosis is initiated by the accumulation of cyclin B in bovine oocytes., Biology of reproduction 55. 55(6. 6) (1996) 1427-36-1427-36.

[86] W. Mak, C. Fang, T. Holden, M.B. Dratver, H. Lin, An Important Role of Pumilio 1 in Regulating the Development of the Mammalian Female Germline, Biology of reproduction 94(6) (2016) 134.

[87] S. Uzbekova, Y. Arlot-Bonnemains, J. Dupont, R. Dalbies-Tran, P. Papillier, S. Pennetier, A. Thelie, C. Perreau, P. Mermillod, C. Prigent, R. Uzbekov, Spatio-temporal expression patterns of aurora kinases a, B, and C and cytoplasmic polyadenylation-element-binding protein in bovine oocytes during meiotic maturation, Biology of reproduction 78(2) (2008) 218-33.

[88] J.H. Stack, J.W. Newport, Developmentally regulated activation of apoptosis early in Xenopus gastrulation results in cyclin A degradation during interphase of the cell cycle, Development 124(16) (1997) 3185-95.

[89] S. Mahrus, J.C. Trinidad, D.T. Barkan, A. Sali, A.L. Burlingame, J.A. Wells, Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini, Cell 134(5) (2008) 866-76.

[90] H.M. Brown, M.R. Anastasi, L.A. Frank, K.L. Kind, D. Richani, R.L. Robker, D.L. Russell, R.B. Gilchrist, J.G. Thompson, Hemoglobin: a gas transport molecule that is hormonally regulated in the ovarian follicle in mice and humans, Biology of reproduction 92(1) (2015) 26.

39

40

Authors declared no conflicts of interest.

Table 1. Number of endogenous peptidoforms and proteoforms with or without post-translational modifications identified by top-down high-resolution MS in bovine follicular cells.

|	Ds	IDs with PTM	Methionine excision	N-acetylation	Q => PyroGlu	Internal acetylation	Tri-methylation	Methionine Oxidation	Phosphorylation	Di sulfide bridge
Whole protein	8	45	45	35	0	3	0	3	2	2
N-ter fragment	8	60	56	55	2	1	0	0	1	0
C-ter fragment	9	5	0	0	0	0	3	0	0	2
Internal fragment	61	3	0	0	2	1	0	0	0	0

41

Table 2. List of peptides/proteins detected by ICM-MS and identified by TD HR-MS/MS proteomics in bovine ovarian follicular cells.

Entry name	Protein description	Theoretical av [M+H]$^+$	Gene	Marker
F1MPB2_BOVIN	Insulin-like growth factor 2 mRNA-binding protein 3	4,718.37	IGF2BP3	Oo
G3X8G9_BOVIN	Probable E3 ubiquitin-protein ligase HECTD4	2,766.04	HECTD4	Oo
E1BK75_BOVIN	Histone H2B	4,572.26	HIST1H2BA	CC
G3N053_BOVIN		6,026.87	HIST1H2BA	-
Q2KII5_BOVIN	Histone H2B	2,886.36	HIST1H2BI	Oo, CC
		3,39		-
		8.96		
TYB4_BOVIN	Thymosin beta-4	4,964.46	TMSB4X	Oo, CC
		4,161.67		-
		4,495.00		-
		4,922.46		-
		5,044.46		-
F6S1Q0_BOVIN	Keratin, type I cytoskeletal 18	3,680.93	KRT18	CC
UBC_BOVIN	Polyubiquitin-C	8,451.74	UBC	Oo
		8,565.84		CC, F
		8,182.39		F
		8,295.55		F
		2,098.45		-
		6,418.29		-
TYB10_BOVIN	Thymosin beta-10	4,717.24	TMSB10	Oo, CC, F
		4,517.99		Oo, CC
		4,004.44		-
HBB_BOVIN	Hemoglobin subunit beta	8,718.97	HBB	Oo, F
		3,557.05		F
		5,193.06		-
		3,485.97		-
RM52_BOVIN	39S ribosomal protein L52, mitochondrial	11,560.21	MRPL52	CC
F1N2I5_BOVIN	Carboxymethylenebutenolidase homolog	1,604.89	CMBL	F
RL39_BOVIN	60S ribosomal protein L39	6,276.49	RPL39	F
F1MP41_BOVIN	Metalloendopeptidase	3,626.05	ASTL	F
		5,043.5		F
		5,286.76		F
		5,330.9		F
		5,547.1		F
		6,620.22		F
		7,036.74		F
		5,199.71		-
		5,460.2		-
FIBB_BOVIN	Fibrinogen beta chain	1,951.04	FGB	F
		2,192.33		-
		2,209.33		-
E1B7N2_BOVIN	Histone H4	11,309.24	HIST1H4A	F
		3,790.37		-
MRP_BOVIN	MARCKS-related protein	6,107.23	MARCKSL1	F
		5,183.29		-
F1N3H1_BOVIN	Calumenin	5,088.54	CALU	F
CO3_BOVIN	Complement C3	5,134.96	C3	F
F1MV04_BOVIN	E3 ubiquitin-protein ligase	4,281.55	ZFP91	F

42

|				
HP1B3_BOVIN	Heterochromatin protein 1-binding protein 3	4,677.33	*HP1BP3*	F
G5E6I9_BOVIN	Histone H2B type 1 like	6,025.92	*LOC107133263*	F
		13,773.97		F
E1BAU3_BOVIN	Immunoglobulin superfamily member 10	3,401.79	*IGSF10*	F
F1N6B7_BOVIN	Serine/threonine-protein phosphatase 2B	1,722.95	*PPP3CC*	F
F1MMM4_BOVIN	Synaptotagmin-like protein 5	1,854.02	*SYTL5*	F
TMA7_BOVIN	Translation machinery-associated protein 7	6,978.11	*TMA7*	F
CLPT1_BOVIN	Cleft lip and palate transmembrane protein 1 homolog	4,914.21	*CLPTM1*	F
E1BEL7_BOVIN	Heat shock protein beta-1	2,185.43	*HSPB1*	-
		3,441.90		-
E1BB06_BOVIN	Formin-like protein 2	1,767.12	*FMNL2*	-
		2,080.26		-
G3N135_BOVIN		2,422.91	*FMNL2*	-
H13_BOVIN	Histone H1.3	1,785.90	*HIST1H1D*	-
		1,947.15		-
		1,975.20		-
A6QQ28_BOVIN	Histone H2B	4,173.79	*H2B*	-
		5,741.56		-
		5,897.75		-
E1BGW2_BOVIN	Histone H2B	2,374.69	*HIST1H2BJ*	-
		2,714.17		-
G3N0F3_BOVIN	Histone H2B	1,985.24	*N/A (H2B)*	-
G3N1C9_BOVIN	Histone H2B	6,025.92	*N/A (H2B)*	-
PTMA_BOVIN	Prothymosin alpha	3,805.01	*PTMA*	-
		3,109.28		-
		11,983.68		-
VIME_BOVIN	Vimentin	4,066.37	*VIM*	-
		4,323.66		-
		5,041.67		-
HSPB1_BOVIN	Heat shock protein beta-1 (HspB1)	2,244.97	*HSPB1*	-
		2,997.33		-
		3,978.44		-
ALDR_BOVIN	Aldose reductase (20-alpha-HSD)	2,994.50	*AKR1B2*	-
		3,807.29		-
CALM_BOVIN	Calmodulin	6,612.23	*CALM*	-
		16749.39		-
IF4H_BOVIN	Eukaryotic translation initiation factor 4H	3,460.82	*EIF4H*	-
		3,955.46		-
HBA_BOVIN	Hemoglobin subunit alpha	3,474.93	*HBA*	-
		15,054.18		-
F1N312_BOVIN	Zinc finger protein 608	1,829.13	*ZNF608*	-
DSS1_BOVIN	26S proteasome complex subunit DSS1	8,189.46	*SHFM1*	-
68MP_BOVIN	6.8 kDa mitochondrial proteolipid	6,835.08	*MP68*	-
F1MUZ9_BOVIN	60 kDa heat shock protein, mitochondrial	3,487.11	*HSPD1*	-
APOA2_BOVIN	Apolipoprotein A2	3,888.26	*APOA2*	-
ATP5H_BOVIN	ATP synthase subunit d, mitochondrial	4,527.18	*ATP5H*	-
E1BIW2_BOVIN	BEN domain-containing protein 4	2,255.44	*BEND4*	-
A5D7Q4_BOVIN	CSDA protein	5,928.39	*YBX3*	-
QCR9_BOVIN	Cytochrome b-c1 complex subunit 9	7,327.43	*UQCR10*	-
CX6A1_BOVIN	Cytochrome c oxidase subunit 6A1, mitochondrial	2,285.53	*COX6A1*	-
COX6C_BOVIN	Cytochrome c oxidase subunit 6C	8,521.95	*COX6C*	-
COX8A_BOVIN	Cytochrome c oxidase subunit 8A, mitochondrial	5,046.94	*COX8A*	-
NDUA4_BOVIN	Cytochrome c oxidase subunit NDUFA4	9,325.72	*NDUFA4*	-

43

F1ME65_BOVIN	Cytoskeleton-associated protein 4	6,193.75	*CKAP4*	-
G5E5G8_BOVIN	Diacylglycerol kinase (DAG kinase)	2,944.45	*DGKZ*	-
F1MWR3_BOVIN	Electron transfer flavoprotein subunit alpha, mitochondrial	3,491.16	*ETFA*	-
EPCR_BOVIN	Endothelial protein C receptor (CD antigen CD201)	2,114.40	*PROCR*	-
FABP5_BOVIN	Fatty acid-binding protein 5, epidermal	4,595.35	*FABP5*	-
FIBA_BOVIN	Fibrinogen alpha chain	1,934.92	*FGA*	-
F1MZX2_BOVIN	Glia-derived nexin (Peptidase inhibitor 7)	3,831.51	*SERPINE2*	-
HMGN3_BOVIN	High mobility group nucleosome-binding domain-containing protein 3	10,578.77	*HMGN3*	-
HINT2_BOVIN	Histidine triad nucleotide-binding protein 2, mitochondrial	3,155.49	*HINT2*	-
E1BMT1_BOVIN	Homeobox protein	3,297.75	*SIX3*	-
KRT81_BOVIN	Keratin, type II cuticular Hb1	2,829.24	*KRT81*	-
F1MCA8_BOVIN	Lamina-associated polypeptide 2	6,405.29	*TMPO*	-
F1N052_BOVIN	matrix-associated actin-dependent regulator of chromatin	2,959.19	*SMARCA5*	-
ATOX1_BOVIN	Metal transport protein ATX1	7,235.30	*ATOX1*	-
E1BIX5_BOVIN	Nuclear pore complex protein Nup153 (nucleoporin)	2,966.43	*NUP153*	-
E1BPA2_BOVIN	Olfactory receptor	1,561.06	*OR2A7*	-
F1N1Y3_BOVIN	Olfactory receptor	6,000.13	*OR5W2*	-
PTMS_BOVIN	Parathymosin	3,488.87	*PTMS*	-
E1BG53_BOVIN	PPAR gamma coactivator 1-beta	4,195.85	*PPARGC1B*	-
F1N5H8_BOVIN	Probable ATP-dependent RNA helicase	2,007.30	*DDX31*	-
THRB_BOVIN	Prothrombin (Coagulation factor II)	5,729.26	*F2*	-
RT63_BOVIN	Ribosomal protein 63, mitochondrial	11,867.73	*MRPL57*	-
F6RBQ9_BOVIN	RNA-binding protein 3	3,176.33	*RBM3*	-
F1N0Q3_BOVIN	RNA-binding Raly-like protein	2,226.40	*RALYL*	-
E1BE29_BOVIN	Ryanodine receptor 2	1,886.24	*RYR2*	-
SPA31_BOVIN	Serpin A3-1 (Endopin-1A)	4,049.91	*SERPINA3*	-
A1L5B7_BOVIN	SERPINE1 mRNA binding protein 1	2,431.61	*SERBP1*	-
RUXF_BOVIN	Small nuclear ribonucleoprotein F	2,406.93	*SNRPF*	-
Q05B82_BOVIN	TEAD2 protein	1,915.23	*TEAD2*	-
THIO_BOVIN	Thioredoxin	11,680.42	*TXN*	-
G3X7U0_BOVIN	Transmembrane protein 28	2,344.81	*FAM155B*	-
TTHY_BOVIN	Transthyretin (Prealbumin)	2,782.16	*TTR*	-
TR112_BOVIN	tRNA methyltransferase 112 homolog	2,605.01	*TRMT112*	-
UBXN1_BOVIN	UBX domain-containing protein 1	6,458.21	*UBXN1*	-
F1MFG7_BOVIN	Valacyclovir hydrolase (VACVase)	4,320.93	*BPHL*	-
F1MC90_BOVIN	Zinc finger protein 106 (Zfp-106)	2,119.43	*ZNF106*	-

44

45

Figure 1

45

Figure 2

Figure 3

A. Cumulus cells / Oocytes / Granulosa cells (PCA plot: PC 1 (41.82 %) vs PC 2 (11.80 %))

B. 133 differential m/z — Cluster 3, Cluster 2, Cluster 1 — Oocytes, CC, GC

C. m/z 4718.34, m/z 6024.89, m/z 8296.25

47

Figure 4

49
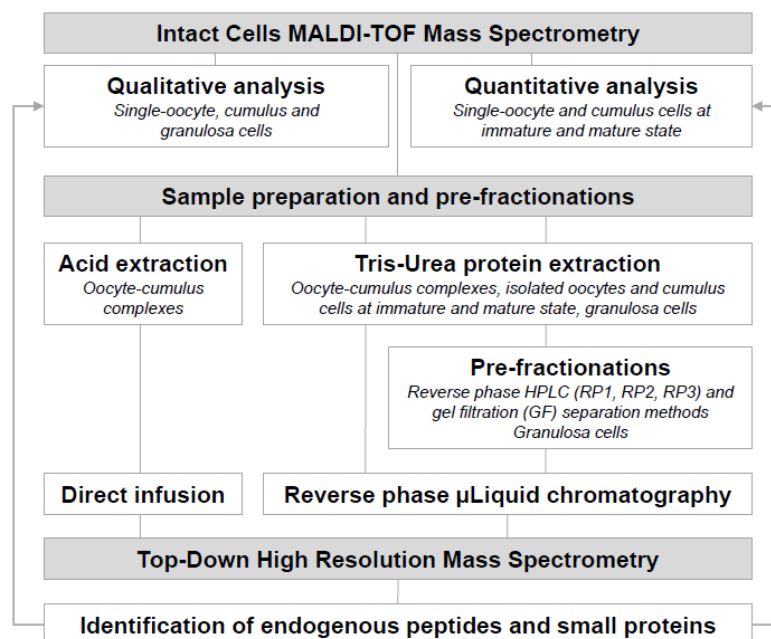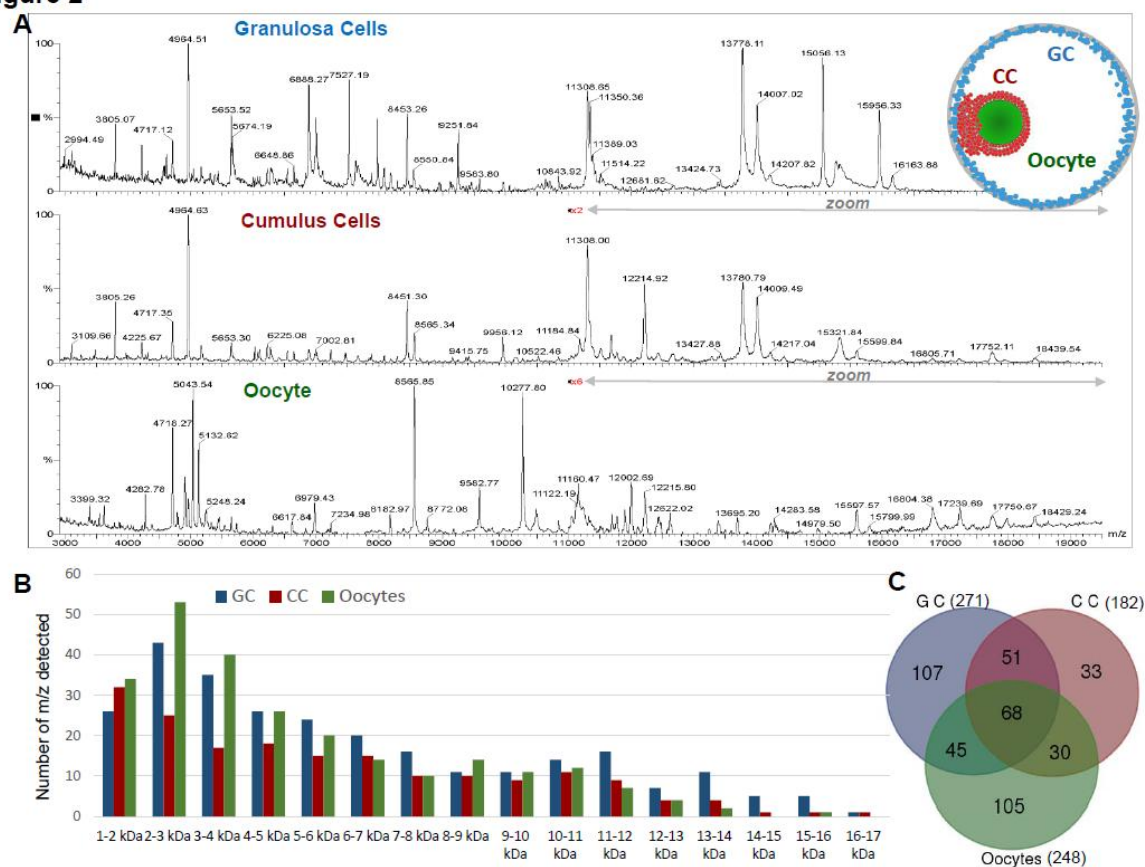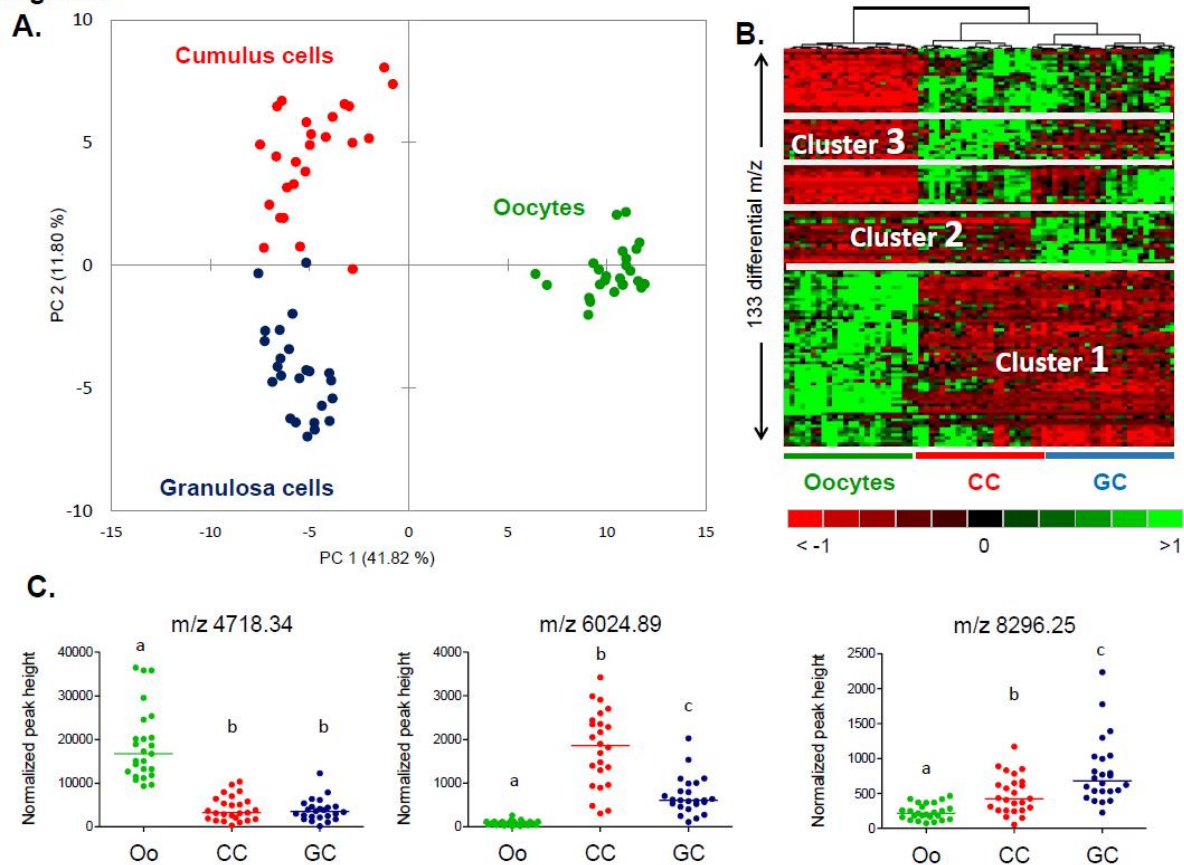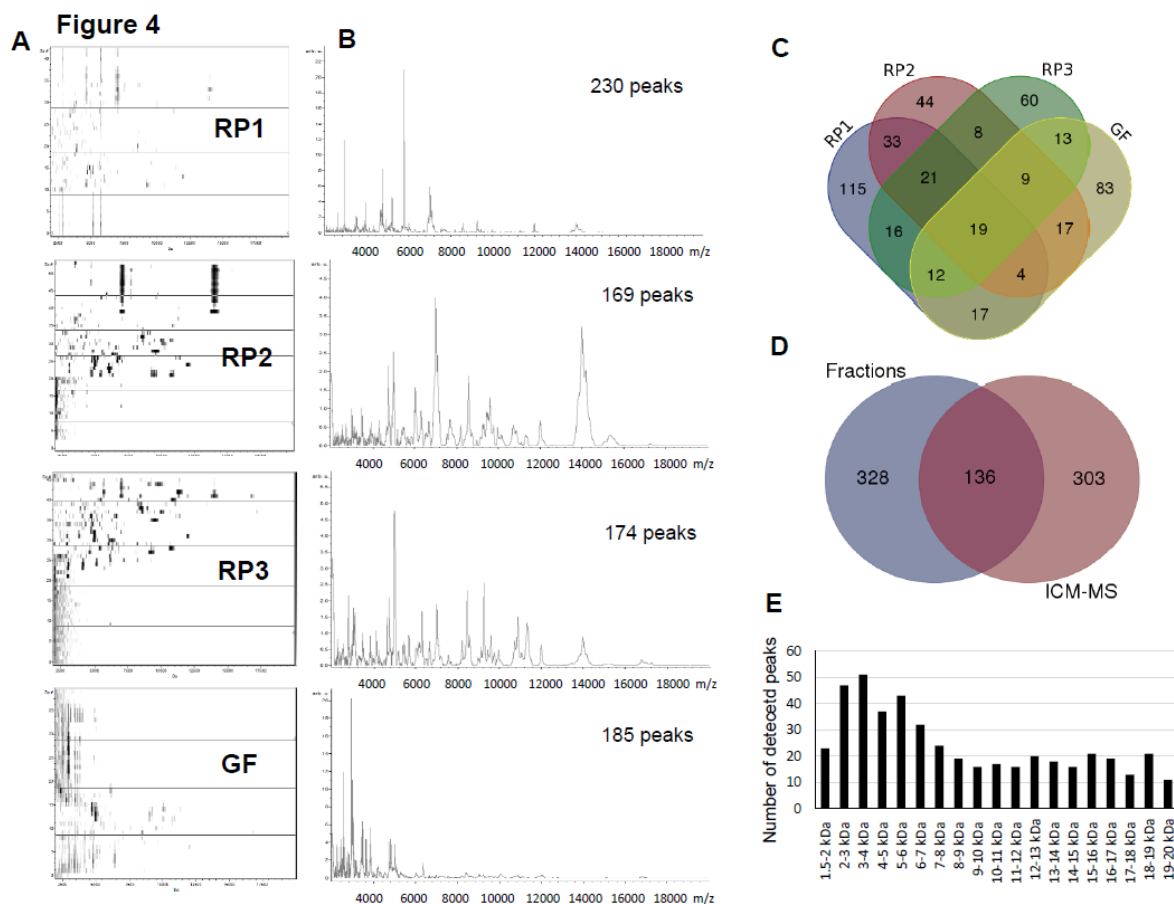
Figure 5

	Previous residues **X-SEQUENCE**	Terminal residues **SEQUENCE-**
N-terminal fragment	-	ₐDGK**R**
Internal fragment	DGLRS	ₐFLₚ**R**
C-terminal fragment	ADₖLR	-

Figure 6

50

Figure 7

Figure 8

52

Graphical Abstract

Granulosa Cells

Cumulus Cells

Single oocyte

Intact Cell MALDI-TOF MS

DIFFERENTIAL ANALYSIS

TOP-DOWN HR-MS/MS

386 IDs
58 proteoforms + 328 peptidoforms
16 markers of oocyte maturation

Immature — GV — GVBD — Meta-I — Ana-I — *Mature* — Meta-II

0h — 6h — 10h — 18h — 24h

IN VITRO MATURATION

Version postprint

53