Intercollegiate Faculty of Biotechnology
of the University of Gdańsk and the Medical University of Gdańsk

DOCTORAL THESIS

Label-free mass spectrometry quantification of proteins linked to the oocyte quality in human follicular fluid: development of suitable methodology for clinical studies

Analiza ilościowa białek związanych z jakością oocytu w ludzkim płynie pęcherzykowym za pomocą metod spektrometrii mas bez znakowania: rozwinięcie metodologii odpowiedniej do badań klinicznych

ALEKSANDRA EWA BOGUCKA

Supervisor: Prof. Stanisław Ołdziej, Ph.D.
Laboratory of Biopolymers Structure
Intercollegiate Faculty of Biotechnology
of the University of Gdańsk and the Medical University of Gdańsk

GDAŃSK 2022

Acknowledgements

First and foremost, I want to thank my supervisor prof. Stanisław Ołdziej, Ph.D. for everything, as it is impossible to list all of his contributions to this project and my career. His guidance has shaped me as a scientist and as a person. I would also like to express my gratitude towards all members of the Laboratory of Biopolymers Structure, former and present, for the great atmosphere and frequent discussions on various topics. Here I want to give special thanks to: Wioletta Żmudzińska, Ph.D. – for the irreplaceable help with any problems and her great attitude which made me feel at ease in the group from the start, Anna Fel-Tukalska – for all the assistance with the samples, mass spectrometry-related discussions, and the great time we spent on the research stay, and Marcel Thiel – for a great help with data analysis, inspiring me to learn Python, and good game recommendations. I want to extend my sincere thanks towards the members of the Laboratory of Mass Spectrometry: prof. Paulina Czaplewska, Ph.D. – who has first introduced me to the world of mass spectrometry and has cultivated my skills and interest in this area ever since, and Katarzyna Macur, Ph.D. – who trained me to operate the LC-MS instrument and helped me with endless problems and questions about protein-based mass spectrometry. I would also like to thank my collaborators: prof. Jacek Wiśniewski, Ph.D. – for all the support and shared knowledge that I received on my research stay under his supervision, and prof. Krzysztof Łukaszuk, M.D., Ph.D. – for the general idea of the presented project and his medical expertise.

This thesis could not have been finished without the continuous support of my family and friends, especially my parents – who have first sparked my fascination in science and encouraged me every step of the way, my husband Michał – who truly is always there for me for better or worse, and my lifelong friend Paweł – who frequently bothered to check on me and shared a discussion on any topic when I needed it.

Funding Acknowledgements

The work presented here received funding from the National Science Centre OPUS 14 financial grant (UMO-2017/27/B/NZ5/02393), the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no 316094 from funds for science in the years 2013-2016 allocated for the internationally co-financed project, and Max-Planck Society for the Advancement of Science. In part, this work is a result of a research stay funded by the Erasmus+ programme. Computational resources used in this project were partly provided by the Informatics Center of the Metropolitan Academic Network (IC MAN-TASK) in Gdańsk.

Table of Contents

Abstract

The recent technology development of instrumentation as well as sample preparation and computational techniques involved in mass spectrometry-based proteomics allows for tailoring the applied methodology specifically for the investigated problem. The aim of the work presented in this thesis was the development of a proper quantitative proteomic workflow for the oocyte quality assessment on the basis of changes in human follicular fluid (hFF) proteome in women undergoing the in vitro fertilization procedure. A considerable number of diversely conducted research studies already focused on exploring the hFF proteome; however, no explicit protein biomarkers were introduced to infertility treatment procedures. The complex data obtained from a large-scale clinical study conducted with sufficient depth of quantitative analysis and a relevant statistical design could point to novel oocyte quality biomarkers. Thus, two distinct label-free quantitative proteomic workflows were created, optimized, and tested for this task in a series of experiments presented in three published studies. The Triple Quad-TOF workflow was designed to be fast and low cost/resource demanding. At the same time the proteome coverage was extended using the SWATH-MS method allowing for the application of multiple fractionation strategies in the spectral library construction without the loss of quantitative measurement accuracy. On the other hand, the Quad-Orbitrap workflow provided comprehensive and sensitive analysis in each single measurement at a higher expense of time and resources. The absolute concentrations of all analyzed proteins were calculated using Total Protein Approach. The utility of both workflows was examined in small-scale clinical studies, which simultaneously generated biomarker candidates of oocyte maturity and competence of blastocyst development. The information obtained in the course of all the conducted experiments allows new insight into the proteome and peptidome landscape of hFF. The good compatibility of the results obtained by both workflows allows to choose a suitable methodology of a future clinical study according to the specific purpose of planned research and available facilities.

Streszczenie

Rozwój technologiczny aparatury, technik przygotowania próbki oraz metod obliczeniowych stosowanych w proteomice opartej na spektrometrii mas pozwala na adaptację wykorzystywanej metodyki do badanego zagadnienia. Celem badań przedstawionych w niniejszej rozprawie było opracowanie proteomicznej metodyki ilościowej odpowiedniej do oceny jakości oocytów na podstawie zmian w proteomie ludzkiego płynu pęcherzykowego kobiet przechodzących procedurę zapłodnienia pozaustrojowego. Znacząca liczba przeprowadzonych do tej pory prac badawczych, wykonanych z użyciem różnorodnych strategii dotyczyła proteomu ludzkiego płynu pęcherzykowego, jednak do procedur leczenia niepłodności nie wprowadzono żadnych sprecyzowanych markerów białkowych. Wielowymiarowe dane uzyskane w wyniku przeprowadzenia szeroko zakrojonego badania klinicznego za pomocą dostatecznie dogłębnej analizy ilościowej i odpowiedniego doboru metod statystycznych mogłyby wskazać nowe markery jakości oocytów. Do tego zadania, w serii badań przedstawionych w trzech opublikowanych pracach opracowano, zoptymalizowano i przetestowano dwa odrębne zestawy metod proteomiki ilościowej bez znakowania. Pierwszy z nich, nazwany Triple Quad-TOF został zaprojektowany tak, aby planowane analizy mogły zostać przeprowadzone szybko i z użyciem niskich kosztów/zasobów. Jednocześnie pokrycie proteomu zostało poszerzone za pomocą metody SWATH-MS pozwalającej na zastosowanie wielu strategii frakcjonowania w budowaniu biblioteki widm bez utraty dokładności pomiaru ilościowego. Drugi opracowany zestaw metod, nazwany Quad-Orbitrap zapewniał dogłębną i czułą analizę w każdym pojedynczym pomiarze przy większym nakładzie czasu i zasobów. Stężenia bezwzględne wszystkich analizowanych białek obliczono przy użyciu metody Total Protein Approach. Przydatność obu zestawów metod została sprawdzona w przeprowadzonych na małą skalę badaniach na próbkach klinicznych, które to badania wskazały potencjalne biomarkery dojrzałości oocytów oraz kompetencji w zakresie rozwoju do stadium blastocysty. Informacje uzyskane w trakcie wszystkich przeprowadzonych eksperymentów dają nowy wgląd w obraz proteomu i peptydomu płynu pęcherzykowego. Dobra zgodność wyników uzyskanych w obu zestawach metod pozwala dobrać odpowiednią metodologię przyszłego badania klinicznego według konkretnego celu planowanych badań i dostępnego zaplecza sprzętowego i materiałowego.

List of Publications

The thesis is based on the following publications:

I. Lewandowska, A.E.; Macur, K.; Czaplewska, P.; Liss, J.; Łukaszuk, K.; Ołdziej, S. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology. J. Proteome Res. **2017**, 16, 3053–3067. doi:10.1021/acs.jproteome.7b00366

II. Lewandowska, A.E.; Macur, K.; Czaplewska, P.; Liss, J.; Łukaszuk, K.; Ołdziej, S. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High PH RP-HPLC Fractiona-tion. J. Proteomics **2019**, 191, 131–142.doi:10.1016/j.jprot.2018.03.010

III. Lewandowska, A.E.; Fel, A.; Thiel, M.; Czaplewska, P.; Łukaszuk, K.; Wiśniewski, J.R.; Ołdziej, S. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid. Int. J. Mol. Sci. **2021**, 22, 7415. doi:10.3390/ijms22147415

List of Abbreviations

ANOVA – analysis of variance

ART – assisted reproductive technology

BMI – body mass index

CDC – Centers for Disease Control and Prevention

CE – capillary electrophoresis

CID – collision-induced dissociation

COH – controlled ovarian hyperstimulation

COC – cumulus-oocyte complex

CV – coefficient of variation

DDA – data dependent acquisition

DIA – data independent acquisition

ECM – extracellular matrix

eFASP – enhanced filter aided sample preparation

ELISA – enzyme-linked immunosorbent assay

ESHRE – European Society of Human Reproduction and Embryology

ESI – electrospray ionization

ETD – electron transfer dissociation

FASP – filter aided sample preparation

FDR – false discovery rate

HAP – high abundant proteins

HCD – higher-energy C-trap dissociation

hCG – human chorionic gonadotropin

hFF – human follicular fluid

HILIC – hydrophilic interaction liquid chromatography

HMWF – high molecular weight fraction

HPLC – high performance liquid chromatography

hSAFF – human small antral follicular fluid

ICSI – Intracytoplasmic Sperm Injection

IgG – immunoglobulin G

ITRAQ – isobaric tags for relative and absolute quantitation

IVF – in vitro fertilization

LC – liquid chromatography

LC-MS – liquid chromatography coupled mass spectrometry

LC-MS/MS – liquid chromatography coupled tandem mass spectrometry

LFQ – label-free quantification

LMWF – low molecular weight fraction

LT – linear ion trap

m/z – mass-to-charge ratio

MALDI – matrix-assisted laser desorption/ionization

MARS-14 – human 14 multiple affinity removal spin cartridge (Agilent Technologies)

MCP – microchannel plate

MED-FASP – multi-enzyme digestion filter aided sample preparation

MI – metaphase I

MII – metaphase II

MS – mass spectrometry

MS/MS – tandem mass spectrometry

MW – molecular weight

PCOS – polycystic ovary syndrome

POR – poor ovarian response

Q, Quad – quadrupole

ROS – reactive oxygen species

RP-HPLC – reversed phase high performance liquid chromatography

SCX – strong cation exchange

SDS – sodium dodecyl sulfate

SWATH-MS – sequential window acquisition of all theoretical mass spectra

TMT – tandem mass tag

TOF – time of flight

TPA – total protein approach

WHO – World Health Organization

XIC – extracted ion chromatogram

Introduction

1. Mass spectrometry-based proteomics

1.1. Proteomics

The main goal of proteomics is the characterization of protein landscape in a given environment, e.g., a cell, a tissue, an organ, whole organism, or a population of organisms (i.e., microorganisms). This may involve protein identification, quantification, characterization of their functions, interaction networks, or modifications[1,2]. There are two main groups of methods for protein identification: based on affinity or sequence. Affinity-based methods make use of antibody-antigen interactions and allow fast and uncomplicated identification and quantification of a target protein, provided the quality of the antibody is sufficient[3]. These methods always require the use of a specific antibody, developed precisely for a given assay. Therefore, only the anticipated set of proteins can be detected using affinity-based methods. In contrast, sequence-based methods allow protein identification from its amino-acid sequence alone, thus enabling untargeted analysis. Sequence-based methods comprise Edman degradation, and mass spectrometry (MS), which becomes a predominant approach to proteomics, due to the multitude of its applications. Ideally, mass-spectrometry-based proteomics would provide information on all proteins present in the sample; this is however limited by the efficiency of sample preparation protocols and by the capabilities of mass spectrometry instruments.

1.2. Bottom-up approach: sample preparation and data analysis

Mass spectrometry allows the measurement of the mass-to-charge ratio (m/z), so it relies heavily on the efficiency of the ionization process. Because of that, the analysis of intact proteins, so called top-down proteomics, is rather difficult and usually allows investigation of only one or few purified protein species at a time. On the other hand, shorter chains of amino-acid residues, peptides, ionize more easily than whole proteins. Protein digestion is an essential step of bottom-up proteomic workflows, which are applicable in a wide range of identification and quantification experiments, especially those involving analysis of multiple proteins at once. Trypsin is the protease most often used for this task, due to its cleavage specificity, although other enzymes, e.g., chymotrypsin, Glu-C, Lys-C, papain, pepsin are also used in particular experiments. Trypsin cleaves the peptide bonds after arginine and lysine residues, which usually creates peptides of a length suitable for mass spectrometry, and more importantly, easily gaining positive charge[1]. Protein digestion is carried out by incubation of the sample with the added enzyme in an optimal temperature and pH. To enable the enzyme

to reach cleavage sites more efficiently, proteins are often denatured, disulfide bonds are reduced, and cysteine residues are alkylated beforehand. Typical in-solution digestion procedure involves reduction of disulfide bonds, alkylation of active sulfhydryl groups, and incubation with a protease conducted sequentially in solution of a single reaction tube[4]. An example of the procedure facilitating conventional protein digestion is Filter-Aided Sample Preparation (FASP), where all the stages are carried out on proteins retained on filter membranes[5]. This results in increased digestion efficiency by allowing the prior use of detergents, such as sodium dodecyl sulfate (SDS), and high concentrations of denaturing agents, such as urea, due to subsequent membrane washing with a suitable digestion buffer. Moreover, a modification of this procedure, Multi-Enzyme Digestion FASP (MED-FASP) comprises of two or three rounds of digestion using different proteases on the same filters (Lys-C followed by trypsin or Lys-C, trypsin, and chymotrypsin)[5].

In data dependent acquisition (DDA) tandem mass spectrometry (MS/MS) measurements, a specified number of the most intense precursor peptide ions are fragmented, and their fragment ion spectra are recorded. The m/z value of the precursor ion and its fragmentation ion spectrum can be later used to deduce the sequence of a given peptide to some extent (e.g., depending on the quality of the spectra and presence of isobaric amino acid residues), which is known as de novo sequencing. A less computationally intensive method, widely used to obtain the information on proteins originally present in the sample, is database search. A protein sequence database is digested and fragmented in silico according to the actual experiment conditions, and the software compares theoretical and experimental spectra according to its scoring algorithm. This method usually involves a decoy search in order to establish a threshold of the percentage of the false positive identifications, which is referred to as false discovery rate (FDR)[6,7].

1.3. Fractionation methods

A disadvantage of the DDA analysis is the limitation of the number of precursor ions analyzed at a time to the most intense ions. It is especially noticeable if the analyzed sample consists of several high abundant proteins (HAP), which conceal the presence of other proteins, such as blood plasma/serum. One solution to this problem is the dynamic exclusion of already fragmented precursors for a limited time or the use of exclusion lists of precursor ion m/z values during MS analysis[8]. Another way involves the application of fractionation strategies during sample preparation. These may involve fractionation of proteins (e.g., affinity chromatography, 1 or 2D gel electrophoresis, isoelectric focusing, liquid chromatography, strong cation exchange, ultrafiltration) or proteolytic peptides (e.g., liquid chromatography, strong cation exchange)[8,9].

1.3.1. Affinity chromatography

Affinity chromatography is particularly beneficial in case of HAP presence, as there are several commercial kits designed for specific biological materials, including human blood serum/plasma. These contain antibodies against a few of the most abundant proteins present in the specific sample, therefore the technique is often referred to as immunodepletion[10]. For example, the Human 14 Multiple Affinity Removal Spin Cartridge (MARS-14) from Agilent Technologies includes antibodies against 14 most abundant human blood plasma/serum proteins. Other simpler solutions are directed towards albumin and/or immunoglobulin depletion using, e.g., Cibacron Blue 3G-A dye (albumin) and Protein A (immunoglobulins)[10,11]. The captured proteins and the depleted sample can then be analyzed separately. An issue with the application of this method is that HAPs often bind other less abundant proteins which are then depleted from the sample as well[10].

1.3.2. Gel electrophoresis-based methods

Gel electrophoresis-based methods include gel electrophoresis in native or denaturing conditions, which allows for separation according to molecular weight (MW; the shape and charge of the protein also play a role in case of native conditions), isoelectric focusing, which segregates proteins according to their isoelectric points, and combination of the two in 2D gel electrophoresis. Protein bands can be later stained for visualization during excision, e.g., by MS-compatible silver or Coomassie stains[4]. An important aspect of the gel-based techniques is that the digestion is conducted in the gel by a modified protocol, and the proteolytic peptides are eluted afterwards[4]. These methods are often used for identification or confirmation of the presence of targeted protein during protein purification. Moreover, 2D gel electrophoresis have been often used for comprehensive proteome characterization[1,12].

1.3.3. Liquid chromatography

Mass spectrometry measurement is often preceded by an on-line high performance liquid chromatography (HPLC) separation (LC-MS and LC-MS/MS, see paragraph 1.4.1.), however, there are many advantages of using LC as a fractionation method prior to MS analysis. LC can be used to fractionate proteins as well as proteolytic peptides after digestion. As the chromatography in the LC-MS analysis is usually conducted in acidic pH to facilitate peptide ionization in positive mode, previous off-line fractionation of the sample in different conditions allows for orthogonal separation[8]. High pH RP-HPLC (reversed phase-HPLC) technique (pH>9) has gained a lot of popularity in in-depth proteomic studies[8,13,14]. Liquid chromatography separates peptides according to their hydrophilic/hydrophobic characteristics in a given environment of mobile (buffer) and stationary

(adsorbent) phase. Another variation of chromatography often used for the purpose of peptide fractionation is strong cation exchange chromatography (SCX). This approach usually utilizes stationary phase with negatively charged functional groups, which bind and separate positively charged analytes according to their net charge[8].

1.3.4. Ultrafiltration

Ultrafiltration divides analytes by their molecular weight on a membrane with pores of specific cutoff. It can be used to isolate and pre-clean proteins from the sample on membranes with low MW cutoff (e.g., 3 kDa). Accordingly, this technique can also be applied to separate the peptides present in the sample. Some peptides can be bound to proteins, therefore specific buffers are used prior ultrafiltration to facilitate their detachment, e.g., containing some percentage of acetonitrile[15,16]. A sequential use of filters with membranes of different cutoff can be used to fractionate proteins according to their MW[17].

1.4. Instrumentation

The mass spectrometry instruments have been improved tremendously through the continuous development of the technology in the 20th and 21st centuries, since the discovery of the MS principles in the 19th century[18]. Nowadays, mass spectrometers break previous limits of resolution, sensitivity, accuracy, and speed making the proteomic analysis of the contents of even a single cell possible[19]. As expected, the performance limits and efficiency of the instrument depend on its components. All mass spectrometers consist of an ion source, a mass analyzer, and a detector. The basic scheme of an MS instrument along with the typical accompanying elements and examples used in bottom-up proteomics is presented in Figure 1.

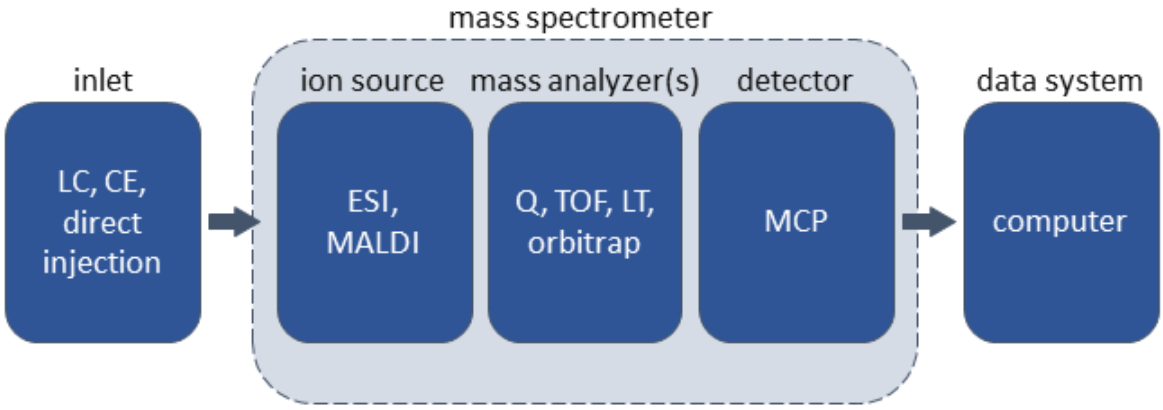


Figure 1. Schematic representation of the components of the mass spectrometry system. The component examples most widely used in bottom-up proteomics are listed in blue rectangles. Abbreviations used: CE – capillary electrophoresis, ESI – electrospray ionization, MALDI – matrix-assisted laser desorption/ionization, Q – quadrupole, TOF – time-of-flight, LT – linear ion trap, MCP – microchannel plate.

1.4.1. Ionization and separation methods

One of the first breakthroughs that led to the widespread use of MS for the analysis of proteins and peptides was the discovery of soft ionization methods: MALDI (matrix-assisted laser desorption/ionization) and ESI (electrospray ionization)[2,18]. MALDI requires prior crystallization of the sample with a specifically designed matrix on a plate. Inside the spectrometer, a laser is used to provide the energy to the matrix compound that first gets desorbed/ionized and facilitates the charge transfer to the analyte. Because of the necessary crystallization step before the analysis, MALDI is not optimal for on-line coupling with other techniques. Therefore, ESI (electrospray ionization) is a technique more popularly used in wide-scale proteomics. In ESI, a sample solution is pumped through a fine needle under high voltage, which creates the solution droplets that in the process of solvent evaporation generate highly charged analyte ions. This type of ion source can be easily coupled with preceding separation techniques, such as LC or capillary electrophoresis (CE) due to the use of liquid samples[20]. This is extremely important in the analysis of complex samples. Liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) is the most popular setup used for proteomics. The efficiency of ionization is proportional to the analyte concentration that reaches the ion source, thus lower flow rates result in higher instrument sensitivity. Because of that, nano-flow-LC (<1 µl/min) is the mainstream solution for LC-MS/MS. The drawbacks of nano-LC are that lower flowrates result in long analysis times and the maintenance of the instrument

and columns are often challenging, which is in general expensive. Micro-flow-LC (<50 µl/min) can be used as a more robust alternative to nano-LC, which overcomes mentioned problems at a cost of lower sensitivity[21].

1.4.2. Mass analyzers and detectors

Mass analyzer is the element that truly determines the performance of the instrument[2]. Quadrupoles (Q, Quad) filter the ions of specific m/z values by creating oscillating electric fields and thus excel in targeted proteomic applications. They can also act as a collision cell for precursor ion fragmentation in triple quadrupole instruments. Time-of-flight (TOF) mass analyzers allow measurement of time in which ions accelerated in electric field travel through the analyzer. TOF analyzers offer high resolution and speed of the measurement. Ion traps, such as linear trap (LT) or orbitrap, are a group of analyzers which hold ions for periods of time. Orbitrap is characterized by a very high resolution, mass accuracy, and sensitivity, and therefore has become a leading solution in the field of proteomics. The longer the ions travel inside the orbitrap electrostatic field, the higher the resolving power of the analysis becomes; therefore, the measurements often take more time than e.g., in TOF analyzers[22]. Mass spectrometers produced today typically contain more than one mass analyzer to provide flexibility of their utilization. In order to allow their application in tandem mass spectrometry, these instruments must also include a collision cell for the fragmentation of precursor ions that is usually specific for the employed mass analyzer. For example, collision-induced dissociation (CID) is used in triple quadrupole instruments, higher-energy C-trap dissociation (HCD) in orbitrap instruments, and electron transfer dissociation (ETD) in ion trap instruments. The last crucial part of the instrument is the detector that records and usually amplifies the obtained signal for later analysis on computational platforms (e.g., microchannel plate). In case of orbitrap instruments, the mass analyzer also acts as the detector.

1.5. Label-free quantification

Mass spectrometry is not an inherently quantitative method due to the “imperfect” ionization process and limits in detection of all peptides present in complex samples. LC-MS provides an additional dimension to the obtained data – retention time, which enabled the invention of most MS quantification strategies. A way to alleviate problems with incomplete ionization and detection is to use stable isotope or chemical labeling, which often allow to conduct measurements on mixed samples and identify relative or absolute differences between experiment conditions based on the applied labeling strategy. This approach is highly precise and not very susceptible to the occurrence of experimental errors. However, labeling is often time and money consuming,

and the number of conditions that can be compared in a single experiment is limited according to the labeling strategy[23]. In contrast, label-free quantification (LFQ) strategies are less accurate and significantly more prone to experimental errors, because all experimental conditions are measured separately, and resulting data is analyzed together to identify differences. At the same time, they provide considerable advantages for large-scale discovery studies involving many samples: (i) any number of conditions can be analyzed in a single experiment, (ii) they are less expensive and time-consuming. Two main approaches to LFQ are: (i) measuring intensity for a given peptide precursor ion/fragment ions, and (ii) counting the number of spectra for a given peptide (spectral counting)[23].

1.5.1. Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH-MS)

One of the label-free relative quantitative methods applied in this work is SWATH-MS (Sequential Window Acquisition of All Theoretical Mass Spectra)[24]. This method utilizes data independent acquisition (DIA), where all precursor ions in a set m/z window are fragmented, and a complicated spectrum of all fragment ions is acquired. The spectra are acquired sequentially in windows that cover the entire precursor m/z range. Ion intensity data for distinct ions in specified retention time range forms extracted ion chromatograms (XIC), which are used to obtain relative quantitative values. The m/z windows can be set with fixed or variable width to better adapt to the ion distribution in a particular analysis[25]. The crucial element in data analysis of SWATH-MS experiment measurements is the spectral library, which enables assignment of recorded fragment ions to a particular peptide sequence, and as a result also to the original protein. Spectral libraries can be created from DDA experiments conducted on the same or similar set of proteins which will be analyzed by SWATH-MS. Therefore, the use of fractionation strategies in library establishment can result in good quality spectra for a greater number of proteins, especially present in low concentrations[26]. Recently, spectral libraries have been also created in silico from existing proteome data[27]. The advantage of the SWATH-MS approach is that a created spectral library can be used for multiple experiments and, conversely, SWATH-MS measurements can be analyzed with any library, thus encouraging the search for new biological data in previously acquired measurements. Due to the use of spectral library, SWATH-MS is a targeted quantitative method in a sense that only peptides present in the library can be identified in the analysis. In practice, hundreds of proteins can be analyzed at once in this experiment, making SWATH-MS a good approach in discovery proteomics[26]. Some programs also facilitate library-free analysis of DIA data, which is significantly more demanding computationally[27].

1.5.2. Total Protein Approach (TPA)

Total Protein Approach (TPA) is a computational approach which enables elucidation of absolute protein concentration values from spectral intensities of large proteomic datasets[5,28]. This method does not require labeling and can be used for proteomic data acquired by any acquisition mode (DDA and DIA)[29]. Concentrations are calculated according to the fraction of combined MS intensity signal of a given protein to the total MS intensity signal registered in the experiment in a following equation:

$$Protein\ concentration(i) = \frac{MS\ signal(i)}{Total\ MS\ signal\ \times MW(i)} \left[\frac{mol}{g\ total\ protein}\right]$$

where MW is the molecular weight of the investigated protein[5]. Therefore, larger, more comprehensive high-quality datasets (depending on high resolution, sensitivity, and mass accuracy of the instrument) present an opportunity for a more precise quantification[28].

2. Assisted Reproductive Technology (ART)

2.1. Infertility

Infertility is a general term for disorders involving reproductive system, distinctively unique as compared to other conditions that it principally affects two people attempting and unable to conceive a child, instead of a single individual. According to the World Health Organization (WHO), infertility is classified as "a disease of the male or female reproductive system defined by the failure to achieve a pregnancy after 12 months or more of regular unprotected sexual intercourse"[30]. The estimated prevalence of infertility worldwide from different reports ranges from 48.5 million couples, through more than 70 million women, rising to 186 million women in developing countries[31–33]. There are several potential causes of infertility, including ovulatory, tubal, uterine, or endocrine disorders in females, and ejaculatory failure or sperm disorders (abnormalities in shape or movement, absent or reduced count) in males[34]. Moreover, the increasing number of factors influencing fertility is associated with civilization development, e.g., lifestyle (alcohol or steroid abuse, obesity, smoking), social and economic (advanced maternal age), or environmental factors (toxins)[34,35]. As a result, the etiology of infertility in a given situation often remains unexplained, hindering the chance of successful treatment. Such cases often lead to mental health problems, relationship issues, and, in extreme instances, social exclusion and ostracism[36].

2.2. Overview of the ART treatment procedures

Assisted reproductive technology (ART) involves multiple procedures aimed specifically at enabling child conception and birth in cases of infertility. It is estimated that until 2013 about 5 million children were born with the ART assistance since the procedure of in vitro fertilization (IVF) was first introduced by Robert G. Edwards, Patrick Steptoe, and Jean Purdy, resulting in the first "test-tube" baby birth in 1978[37]. Because of the versatile character of the offered treatment which can enable pregnancy even when the exact causes of infertility remain unknown, the popularity of ART is constantly on the rise. It is estimated that ART-born individuals account for more than 0.1% of the world population, and this number is expected to rise up to 1.4-3.5% by the year 2100[38]. According to the CDC (Centers for Disease Control and Prevention), ESHRE (European Society of Human Reproduction and Embryology), and WHO definitions, ART concerns the fertility treatment procedures which involve handling both male and female gametes with the intent to combine them outside of the female body[39]. The most popular treatment as a complete set of different ART techniques is the in vitro fertilization procedure. Other examples include such techniques as cryopreservation, reproductive surgeries, or mitochondrial replacement therapy.

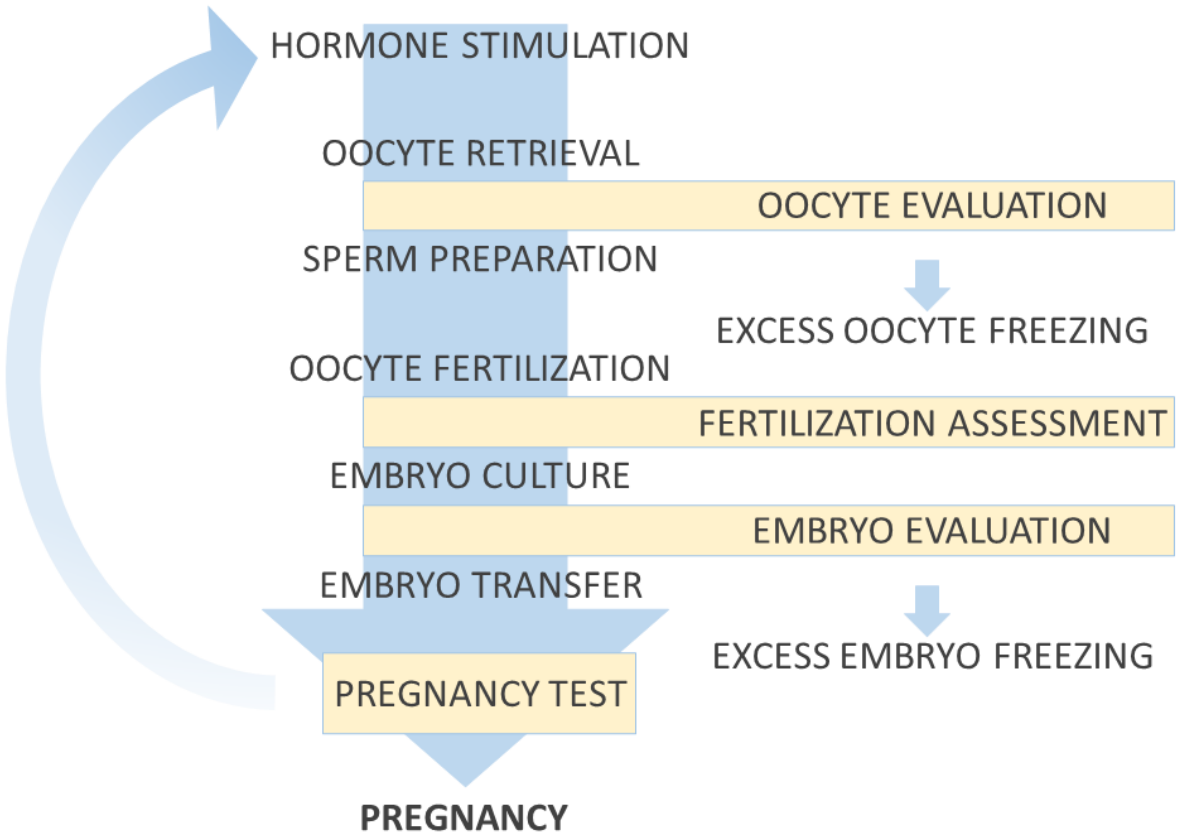


Figure 2. A general course of a single cycle of the in vitro fertilization procedure. The most important quality check stages are highlighted in yellow.

In vitro fertilization is a procedure consisting of several stages (see Figure 2). Typically, a single IVF cycle starts with hormonal stimulation by gonadotropins to trigger the development of multiple ovarian follicles in a single cycle. The oocytes are then harvested along with the contents of mature follicles by transvaginal oocyte retrieval, prepared for fertilization by removing surrounding cumulus cells, and subjected to morphological evaluation. At this time, sperm is also prepared, typically by centrifugation and removal of unnecessary material. Afterwards, fertilization is initiated either by co-incubation, or, especially in cases of reduced sperm count or motility, intracytoplasmic sperm injection (ICSI). By Polish legislation, no more than 6 oocytes can be fertilized in a single cycle[40]. Embryologists assess the presence of two pronuclei in eggs, and fertilized eggs are placed in growth media. Embryos are cultured for 2-5 days, often up until the blastocyst stage, while being regularly evaluated. 1-2 properly developed embryos are chosen for the transfer to the uterus, and the rest of them are subjected to cryopreservation. In the event of the implantation occurrence, the pregnancy is further monitored by a physician.

The success rate of ART procedures, calculated as the cumulative delivery rate per cycle including fresh and frozen embryo cycles, reached 30.8% in Europe in 2017 as reported by ESHRE[41]. Thus, in most cases more cycles of treatment are necessary to achieve pregnancy and deliver a baby. Before 1999, it was more prevalent to transfer 3 or more embryos to maximize the success rate of a single cycle. Since then, this number has been reduced to less than 7% due to complications associated with multiple pregnancies, which often endangered the health of the patient and children[42]. Due to the general course of treatment as presented in Figure 2, ART procedures are often burdened with an ethical and legal issue of embryo overproduction. Later, during embryo culture, zygotes from the same donor show differences in development, which could indicate that oocytes of the same donor may have different fertilization potential[43,44].

2.3. Current methods of oocyte quality evaluation

Precise evaluation of the oocyte quality prior the fertilization event is a promising path to diminish the embryo overproduction problem, increase the IVF success rates, and spare the patients the unnecessary procedures in case of meager chances of success. After their retrieval, oocyte maturity is graded. MI (metaphase I) oocytes may be cultured for a short period of time until they attain maturity, while MII (metaphase II) oocytes are further assessed morphologically by experienced embryologists. Depending on a clinic, a number of features might be analyzed, e.g., shape and volume, meiotic spindles, cytoplasm, perivitelline space, polar body, zona pellucida, or cumulus complex[43,44]. Oocyte-scoring based on some of these parameters has proven to be useful in prognosis of fertilization outcomes determined by embryo cell number, embryo grade, and clinical pregnancy[44]. Morphological

criteria can be evaluated by non-invasive methods, yet this assessment remains to be relative depending on the clinic or embryologist. A more observer-neutral alternatives are molecular markers investigated by genomic, proteomic, or metabolomic techniques. For instance, proteomic analysis of human oocyte has been recently accomplished by high-resolution mass spectrometry[45]. However, no exhaustive tests can be performed on the oocytes used in treatment, as these individual cells are indispensable for fertilization and subsequent growth and development of the embryo.

3. Human follicular fluid (hFF)

3.1. Microenvironment of the oocyte development

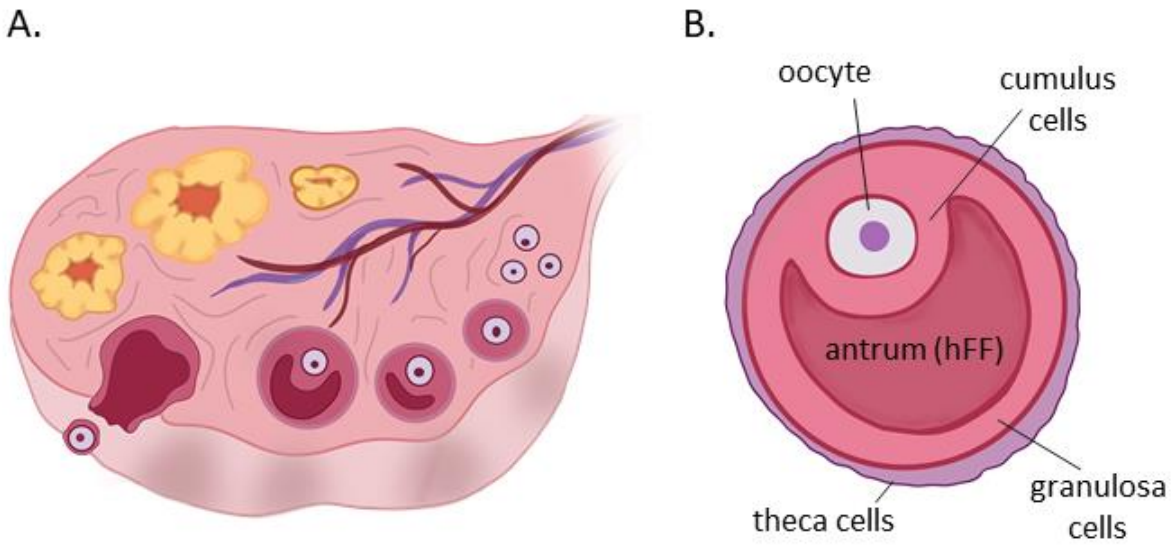


Figure 3. Ovarian follicle development and structure. Schematic representation of the folliculogenesis process in the ovary (A). Components of a mature ovarian follicle (B).

Ovarian follicles develop in a process called folliculogenesis (see Figure 3A), resulting in the full growth and release of a single oocyte (in most cases) per menstruation cycle during ovulation. A mature follicle contains granulosa and theca cells besides the oocyte, and its antrum is filled with human follicular fluid (hFF, see Figure 3B). Layers of granulosa cells are packed around the oocyte to protect it and control the process of maturation. These cells are often referred to as cumulus cells, and with the oocyte comprise the cumulus-oocyte complex (COC). Theca cells form layers around the perimeter of the follicle, controlling its volume and maintaining contact with blood vessels. The correct progression of folliculogenesis and the process of oocyte maturation, oogenesis, demands maintaining a balance between various intra- and extrafollicular factors by these specialized cells.

In consequence, follicular fluid becomes the medium of communication, and its composition may reveal the traces of cell signaling and processes occurring around the oocyte[46]. Therefore, it has been hypothesized that hFF composition could reflect the physiological state of the oocyte, and consequently, also its fertilization potential, understood as quality[46,47]. During the oocyte retrieval procedure after hormonal stimulation (see paragraph 2.2., Figure 2), whole contents of multiple mature follicles are collected separately. After separation of the oocyte from the follicle content, remaining cells and fluid are rarely used in any further medical procedures[47], essentially becoming medical waste. Hence, the hFF analysis enables the search for the oocyte quality biomarkers outside of the precious oocyte cell without any need for additional invasive medical procedures for the patient.

3.2. Proteomics of hFF

Several different hFF components have been studied previously, including hormones, growth factors, interleukins, anti-apoptotic factors, reactive oxygen species (ROS), proteins, peptides, amino-acids, sugars, and prostanoids[47]. Proteins and peptides constitute a wide group of hFF constituents involving several of the mentioned compounds, so they have been regarded as an attractive target of investigation[48–72]. Follicular fluid is created mainly from the fluid selectively absorbed from blood vessels during the follicle development, enriched in compounds secreted by the oocyte and its accompanying cells. Because of that, hFF is highly similar in composition to human blood plasma/serum and its proteomic analysis faces the same problem of substantial differences in ranges of dynamic protein concentrations. Even though the overall protein concentration is high, a few high abundant proteins, such as serum albumin, comprise the vast majority of the protein mass, thus masking the presence of proteins secreted by the oocyte or its accompanying cells[47]. Therefore, proteomic analysis of follicular fluid is often carried out with incorporation of fractionation strategies into sample preparation (see paragraph 1.3., Table 1). Moreover, the use of orbitrap-based spectrometers coupled with nanoflow LC are most widely used in these experiments due to their high-resolution measurements (see paragraph 1.4., Table 1). Bianchi et al. have summed up previous proteomic hFF research in their review published in 2016, where they compiled a list of 617 proteins identified in previous studies[48]. The results of the thorough functional analysis conducted on this set of proteins have shown that hFF proteins are most notably involved in the molecular processes of extracellular matrix (ECM) remodeling and degradation, coagulation, and immunological response and inflammation. Table 1 contains a list of the most comprehensive recent hFF proteomic studies not included in that review.

Table 1. Summary of latest proteomic studies on human follicular fluid with the short description of analyzed samples, main findings, sample preparation workflows, and MS instrument setups.

Study	Participating patients	Samples analyzed	Proteins identified	Findings	Applied sample preparation methods	MS instrument setup
Bayasula et al., 2013[49]	12 patients undergoing IVF/ICSI	2 samples per patient	503 proteins	53 potential marker proteins (comparing fertilized to unfertilized oocytes)	in-solution digestion by trypsin	LTQ Orbitrap, nano-LC
Ambekar et al., 2015[50]	26 PCOS patients, 26 healthy controls	pool samples from each group	770 proteins	186 potential marker proteins (comparing PCOS to healthy controls)	immunodepletion, in-solution digestion by trypsin, ITRAQ labeling, SCX chromatography	LTQ Orbitrap, nano-LC
Regiani et al., 2015[51]	10 endometriosis patients, 10 healthy controls	1-2 samples per patient	537 proteins	33 proteins specific for the control group, 37 for the endometriosis study group, 212 for the endometrioma study group	in-solution digestion by trypsin, high pH RP-HPLC	Q-TOF, nano-LC
Twigt et al., 2015[52]	15 folic acid supplement users, 15 non-users	pool samples from each group	227 proteins	13 potential marker proteins (comparing folic acid supplement users to non-users)	albumin/IgG immunodepletion, in-gel trypsin digestion, TMT labeling, high pH RP-HPLC, isoelectric focusing, HILIC chromatography	LTQ Orbitrap, nano-LC
Wu et al., 2015[53]	6 patients undergoing COH, 6 undergoing natural cycle	samples from dominant follicles	8 of 13 potential marker protein spots	13 potential marker protein spots (comparing COH to natural cycle)	albumin/IgG immunodepletion, 2D-gel electrophoresis, in-gel trypsin digestion, additional western blotting	MALDI-TOF/TOF
Zamah et al., 2015[54]	3 oocyte donors, 3 patients pre-hCG, 3 patients post-hCG	1 sample per patient	742 proteins	17 potential marker proteins (comparing pre-hCG to post-hCG)	immunodepletion, in-solution digestion by trypsin, ITRAQ labeling, high pH RP-HPLC	LTQ Orbitrap, nano-LC

Chen et al., 2016[55] (peptidomics)	50 patients undergoing IVF/ICSI	1-2 samples per patient	7102 peptides from 159 proteins	53 potential marker peptides (comparing fertilized to unfertilized oocytes), 7 peptides from 7 proteins identified	ultrafiltration (filtrate analysis)	Q-Orbitrap, nano-LC
Oh et al., 2017[56]	6 poor ovarian responders, 7 healthy controls	samples from dominant follicles	1079 proteins	131 potential marker proteins (comparing poor ovarian responders to controls)	immunodepletion, FASP with trypsin, TMT labeling, high pH RP-HPLC	Q-Orbitrap, nano-LC
Shen et al., 2017[57]	10 patients undergoing IVF/ICSI	a pool sample	219 proteins	description of identified proteins	in-solution digestion by trypsin, high pH RP-HPLC, additional western blotting	MALDI-TOF/TOF, LC spotting system
Lehmann et al., 2018[58]	3 patients undergoing IVF/ICSI	1 sample per patient	1392 proteins	comparison of sample preparation/data analysis methods	FASP, eFASP, in-solution digestion by trypsin	Q-Orbitrap, nano-LC
Poulsen et al., 2019[59]	25 patients undergoing IVF/ICSI	1 sample per patient	400 proteins	40 potential marker proteins (different concentrations across ovulation)	in-solution trypsin digestion	Q-Orbitrap, nano-LC
Zhang et al., 2019[60]	9 overweight/ obese PCOS patients, 9 normal-weight PCOS patients, 9 healthy controls	3 pool samples from each group	1153 proteins	41 potential marker proteins (comparing overweight/ obese PCOS to control), 19 potential marker proteins (comparing normal-weight PCOS to control)	immunodepletion, in-solution digestion by trypsin, TMT labeling, high pH RP-HPLC, additional ELISA tests	Q-Orbitrap, nano-LC
Pla et al., 2021[61]	31 patients undergoing fertility preservation procedures	1-3 samples per patient (2-3 for quantification experiments)	2461 proteins	100 potential marker proteins (comparing oocytes capable of maturing to MII to not capable)	immunodepletion, in-solution digestion by trypsin, SCX, high pH RP-HPLC, no fractionation for quantification	Q-Orbitrap, Triple Quad, nano-LC

So far, up to 2461 proteins have been identified in a single study regarding the analysis of the human small antral follicular fluid (hSAFF) retrieved in a fertility preservation procedure from cancer patients[61]. The volume of hSAFF is lower than hFF in mature follicles as it has not yet absorbed as much fluid from blood vessels, and the protein concentration is higher, which is mostly essential for identification of proteins secreted by the oocyte or the accompanying follicle cells. Other presented research usually led to identification of few hundreds of proteins. Proteomic studies of hFF often concern investigation of reproduction disorders (e.g., endometriosis[51], polycystic ovary syndrome (PCOS)[50,60], poor ovarian response (POR)[56]) or other personal factors which may influence fertility (e.g., folic acid supplement consumption[52], obesity[60]). Other experiments concentrate on examination of oocyte development[59], maturation[54,61], or fertilization[49,55], focusing on elucidation of the oocyte quality markers.

3.3. Identification of the oocyte quality protein biomarkers

Critical issues of the oocyte quality research based on hFF analysis are associated with careful experiment planning and interpretation of the results. In order to assess the harvested oocytes individually, corresponding hFF samples also must be analyzed separately to discern their divergent characteristics[47]. The physiological state of the patient is also responsible for general features in the whole set of retrieved follicles. Individual traits, e.g., age, body mass index (BMI), presence of any disorders, as well as exercise, diet or smoking habits may all influence the reproductive performance[73], which might be reflected in the composition of follicular fluid. For this reason, multiple samples from separate follicles of single patients should be incorporated in a study to be able to account for individual patient features, and thus, identify markers associated with quality of single oocytes. Moreover, a crucial part of data interpretation is the establishment of clinical properties associated with “good quality” of the oocyte. A successful IVF treatment concludes with the delivery of a healthy baby; however, it consists of multiple stages involving a number of factors (see Figure 2). The most crucial points other than the quality of the oocyte include, e.g., sperm quality, endometrium receptivity in the uterus, and nine months of physiological events which take place during the pregnancy. Although there have been efforts to correlate hFF composition directly with the final result of the IVF procedure[62], these outcomes should be interpreted with caution, considering the additional aspects mentioned above.

So far, few studies were based on the strategy to examine two or more hFF samples from single donors (see Table 1). Bayasula et al. analyzed samples associated with fertilized (resulting in pregnancy) and unfertilized oocytes from the same patient[49]. Pla et al. have compared hSAFF accompanying oocytes which were capable of maturing to MII stage in vitro to those which

did not mature analyzing multiple follicles from single patients[61]. Both discussed studies attempted to examine the oocyte quality by investigation of the oocyte fertilization competence[49] or maturation potential[61]. There were also other studies aimed similarly at uncovering the process of oocyte development and maturation by studying changes in hFF proteome during ovulation[59] or after hCG administration[54]. Moreover, Chen et al. searched for endogenous peptide markers of oocyte fertilization competence in the only conducted peptidomics study of hFF so far[55].

The described approaches applied in hFF proteomics studies allowed to increase the knowledge on the process of oogenesis and the fertilization event. However, they did not yet result in a satisfactory answer to the problems of embryo overproduction and insufficient success rates of the ART procedures. For this reason, there is a need for a large-scale clinical study incorporating multiple samples from individual oocyte donors along with the clinical data on oocyte maturity and embryo culture outcomes to provide the oocyte quality markers, which could be used to improve existing ART procedures. The emergence of new technologies in proteomics allows for such discovery-based research. The choice of specific experimental methods and analytical design, however, remains crucial for the unbiased interpretation of the research outcomes.

Aim

The aim of the presented work was to develop a suitable workflow for a quantitative proteomic analysis of human follicular fluid. Additional essential characteristics of the selected workflow included: (i) suitability for a large-scale clinical study, and (ii) applicability in identification of biomarkers of oocyte quality. The completion of the main aim involved optimization of sample preparation procedures, LC-MS/MS methodology, and data analysis strategy.

Additional purposes of this research comprised testing devised workflow(s) in a clinical setting, comparison of the compatibility of the results obtained using different strategies, and overall characterization of the human follicular fluid proteome and peptidome.

Results and Discussion

This section includes the general summary description and discussion of the outcomes obtained in the course of this work. Details on experimental procedures and complete presentation of the results including supplementary data can be found in the published works (see List of publications, page 7). Throughout this section, particular findings will be referenced to sections in respective publications using the order presented on the page 7 (I, II, III).

1. Development of suitable workflow for hFF proteome analysis in clinical studies

Optimization of the methodology of hFF proteome analysis fitting the aim of this work involved testing multiple factors associated with sample preparation (digestion, fractionation, sample clean-up), MS analysis (instrument setup, parameters, quantitative method), and data analysis (database search, quantification, data processing, statistics). In the course of these experiments, two label-free quantitative workflows were established, each of them centered around the employed LC-MS/MS instrumentation setup.

1.1. Triple Quad-TOF workflow

The first developed workflow was based on the use of TripleTOF 5600+ mass spectrometer (SCIEX) coupled with microflow LC. This instrument, especially in combination with micro-LC, allows for high-resolution measurements in short times (see Introduction 1.4.1.-1.4.2. paragraphs). Therefore, the development of this workflow was aimed at increasing the proteomic coverage of the analysis while maintaining its robustness. The use of SWATH-MS (see Introduction 1.5.1. paragraph) was an adequate quantitative solution in this case, as it allowed for flexibility in sample preparation procedures for the creation of the spectral library[26]. At the same time, the preparation of actual samples measured quantitatively could remain uncomplicated, so that the measured concentrations would not be disrupted by the preparation protocols, e.g., extensive fractionation. A simplified course of the optimization of the Triple Quad-TOF workflow over the experiments presented in publications I-III is demonstrated in Table 2. In order to quantitatively analyze both proteins (high molecular weight fraction, HMWF) and peptides (low molecular weight fraction, LMWF) in Publications I-II, hFF samples were subjected to ultrafiltration on 10 kDa cutoff membranes (see Publication I, Figure 1). The experiments presented in Publication II involved optional protein (HMWF) and peptide (LMWF) fractionation by high pH RP-HPLC to assess its effect on spectral library construction (see Publication

II, Results 3.8. and 3.10. paragraphs). In Publication III, a few sample processing techniques (digestion, fractionation) were tested for their efficiency in hFF proteome identification and quantification (see Publication III, Results 2.3. paragraph and Figure 2). Moreover, the parameters of LC-MS/MS analysis were adjusted, and the use of variable width transmission windows was introduced (see Publication III, Materials and Methods).

Table 2. Optimization of the Triple Quad-TOF workflow described in publications I-III.

		Publication I	**Publication II**	**Publication III**
Workflow stages				
Protein fractionation for spectral library construction		immunodepletion, ultrafiltration	high pH RP-HPLC, immunodepletion, ultrafiltration	high pH RP-HPLC, immunodepletion
Protein fractionation for quantitative measurements		ultrafiltration		none
Protein digestion		in-solution digestion		FASP (trypsin)
Peptide clean-up		commercial C18 spin tips		STAGE Tips (C18)
LC method parameters		30 min, 10-90% buffer B		30 min, 8-40% buffer B
DDA method parameters	MS	100-2000 Da, 50 ms		400-1000 Da, 250 ms
	MS/MS	100-2000 Da, 40 ms		100-1500 Da, 100 ms
	Cycle time	1.1 s		2.3 s
SWATH-MS method parameters	MS	100-2000 Da, 50 ms		400-1000 Da, 50 ms
	MS/MS	100-2000 Da, 40 ms		100-1500 Da, 40 ms
	Cycle time	1.1 s		
	transmission windows	25 fixed width windows over 400-1000 Da range		25 variable width windows over 400-1000 Da range
Main results				
Total number of proteins identified in all experiments		158	400	1182
Number of proteins identified in HMWF / LMWF		103/91	302/161	1177/14
Number of proteins identified in the unfractionated pool sample		-	85	129
Number of proteins quantified		74 (pool sample) / 72 (clinical samples)	39-79 (pool sample)	129 (pool sample) / 215 (clinical samples)
Number of proteins quantified with CV < 20% in the pool sample		40	4-17	98

CV – coefficient of variation

A total of 103 HMWF proteins were identified in the pilot study (see Publication I, Figure 3) using only a simple fractionation procedure involving immunodepletion. The addition of high pH

RP-HPLC in sample preparation procedure significantly increased this number to 302, whereas the number of proteins identified without it remained similar at 104 proteins (see Publication II, Figure 1). The adjustments of sample processing protocol and DDA method parameters in Publication III allowed to enhance the identification capabilities of the presented workflow, which is noticeable by comparing the number of proteins identified in the unfractionated pool sample (85 in Publication II to 129 in Publication III). The total number of proteins identified without extensive fractionation was also higher at 259 proteins (see Publication III, Results 2.3. paragraph). Here, several methods of sample preparation were tested (fractionation, digestion, comparison in Publication III, Figure 2A). In summary, the lowest number of identifications were found in HAPs fraction resulting from ultrafiltration (86) and in-solution digested sample (93). Nevertheless, more proteins were identified in the HAPs fraction than targeted 14 proteins, suggesting a critical impact of protein binding in this type of fractionation[10]. The presence of similar numbers of identifications in an unfractionated sample digested in-solution in Publications II and III suggest that the change to FASP in sample processing protocol was the most influential in the result improvement. Similar numbers were established for unfractionated (129), HMWF after ultrafiltration (133), and immunodepleted samples (131) digested by FASP with trypsin. The result of the average number in immunodepleted sample is rather underwhelming; however, more than half of the immunodepleted sample identifications were not detected in the unfractionated sample (see Publication III, Supporting Material 2, Table S6). The highest numbers of proteins were detected after MED-FASP digestion with two (trypsin and chymotrypsin, 139) or three enzymes (LysC, trypsin, and chymotrypsin, 154). Nevertheless, the most remarkable expansion of the identified proteins set was caused by enhanced protocol of high pH RP-HPLC fractionation. In contrast to 30-min gradients used in Publication II, 120-min gradients were used in Publication III. This one-time resource cost enabled identification of 1151 proteins in these experiments alone (fractionation of previously unfractionated sample: 664 proteins and immunodepleted sample: 958 proteins), resulting in total of 1177 HMWF proteins identified in this study (see Publication III, Results 2.3. paragraph). In this case, a significant increase in identifications after two-step fractionation (including immunodepletion) suggests that the instrument sensitivity or resolution in short-gradient single measurements (without high pH RP-HPLC fractionation) is not sufficient to detect proteins present at such low concentrations.

Regarding the efficiency of SWATH-MS quantification, 74 proteins were analyzed in the pool sample in the pilot study and more than half of them (40) exhibited low CV values (<20%) (Publication I, Supporting Information Part 3). Surprisingly, the addition of extensive fractionation experiments in the spectral libraries worsened quantification in almost all instances resulting in analysis of 39 to 79 proteins depending on the library (Publication II, Figure 3), with only 4-17 proteins measured with low CV values (publication II, Figure 4). One explanation of this issue is that a multitude of spectral

information in the library, possibly associated with different retention times, overburdened the quantification capabilities of the method. Therefore, both DDA and SWATH-MS method parameters were further adjusted, i.e., the scan MS and MS/MS m/z ranges were limited to encompass only the length typically populated by ions, the DDA scan times were increased to allow the collection of high-quality spectra, and variable width windows constructed on the basis of ion equalized frequency[25] were introduced (see Publication III, Materials and Methods). Peptide standards (iRT kit, Biognosys) were spiked in all samples to mitigate the errors associated with retention time differences. A few sample processing protocols for quantification were compared, and FASP with trypsin on unfractionated sample was selected for its best performance (Publication III, Figure 2D). These changes along with the new extended spectral library allowed for quantification of 129 proteins in the pool sample with very good quality (more than 75% proteins with low CV values). At this point, it is worth to mention that 215 proteins were quantified using the same spectral library in 20 clinical samples (across 60 measurements). This observation allows for a hypothesis that the quantification capability of the developed workflow might further improve with a greater number of analyzed samples, which is highly desirable in a large-scale clinical study. Publication III focused on the optimization of the protocol for the analysis of hFF proteome, thus the samples analyzed quantitatively were not subjected to ultrafiltration to decrease the possibility of errors induced by sample processing[23], which also resulted in a notable increase of the number of proteins analyzed with low CV values (see Publication III, Figure 2D). Moreover, TPA was employed in data analysis to obtain meaningful absolute protein concentrations. In this case, an increase in number of quantified proteins associated with larger number of samples should also result in more accurate absolute quantification[28].

Quantitative peptidomics was not the main focus of the presented studies, however it should be mentioned that Publications I-II involved SWATH-MS quantification of peptides in the isolated LMWF. In the first study, 91 proteins were identified in the spectral library allowing for quantification of 43 peptides from 23 proteins (see Publication I, Results Low Molecular Fraction paragraphs). On the other hand, 29 proteins were identified in Publication II in experiments not including further fractionation (see Publication II, Figure 1), and only 14 were identified in Publication III (see Publication III, Figure 2A). The reason for these differences is most likely associated with the number of performed measurements. The first spectral library (Publication I) was constructed on the basis of 92 separate measurements, whereas the uncomplicated library in Publication II was based on 8 measurements, and only 3 LMWF measurements were performed in Publication III. Due to the absence of any protease inhibitors addition in the presented experiments, the LMWF fraction composition was highly unstable[74], hence a larger number of separate experiments were related to a higher probability of discovery of new proteins. This observation may also be associated with the fact that peptides are easily bound

by proteins (especially such as highly abundant in hFF serum albumin), and this process might be difficult to control in experimental conditions causing differences in qualitative and quantitative output of each single analysis[75]. Thus, the use of high pH RP-HPLC fractionation in Publication II caused a direct increase of identified proteins to 154 in all experiments. Similarly, the most comprehensive spectral library containing all LMWF measurements enabled quantification of 106 peptides from 53 proteins in contrast to 38 peptides from 14 proteins quantified in the simplest library (see Publication II, Figure 5). However, this improvement was not associated with the increased quality of quantification, which remained extremely poor as reflected by high CV values (see Publication II, Figure 6). A probable solution to this problem could be the addition of protease inhibitors to hFF after collection, as well as strict care taken to preserve the same conditions during sample storage and processing[74].

1.2. Quad-Orbitrap workflow

The other developed workflow was centered around the Q Exactive HF-X (Thermo Fisher Scientific) mass spectrometer coupled with nanoflow LC. This instrument enables measurements of very high-resolution and sensitivity, the features enhanced even more by low flowrates utilized in nano-LC (see Introduction 1.4.1.-1.4.2. paragraphs). As was mentioned before, such instrument setup has been employed most frequently in hFF proteomics (see Introduction 3.2. paragraph, Table 1)[49,50,52,54–56,58–61]. To be able to best exploit the advantages of this mass spectrometer, the MED-FASP method[5] with three consecutive digestions (LysC, trypsin, chymotrypsin) was applied for sample processing (see Publication III, Materials and Methods) to increase the proteomic coverage in each single sample. Peptides resulting from each of the digestion stages were analyzed separately in longer LC gradients (95 min). Thus, the analysis of each sample consumed significantly more time and resources than in the case of Triple Quad-Orbitrap workflow (see Results 1.1. paragraph, Table 2 and Publication III, Table 3), but at the same time produced an individual, precise, and comprehensive outcome. These measurements acquired in DDA mode were used also for protein quantification further assisted by TPA[5]. Due to the limited access to this instrument, less experiments than in the case of the Triple Quad-TOF workflow were conducted, and only simple fractionation procedures (immunodepletion, ultrafiltration) were performed to analyze their effect on the proteomic outcomes. The experiment involving the most popular digestion method, in-solution digestion by trypsin, was also carried out to better compare the optimized workflow to other hFF proteomic studies[49,50,54,58–61].

The results obtained using the Quad-Orbitrap workflow are summed up in Figure 1 of Publication III (Publication III, Results 2.2. paragraph). A total of 942 proteins were identified in all conducted experiments. Two digestion methods were compared (MED-FASP with three enzymes

and in-solution digestion by trypsin), and MED-FASP resulted in a significantly more identified (565 to 380) and quantified proteins (438 to 265) as well as higher quality of quantification (e.g., 107 to 30 proteins quantified by 5 or more peptides with less than 10% CV values, see Publication III, Figure 1E). Fractionation by ultrafiltration generated slightly better results in HMWF than the unfractionated sample (e.g., 601 identified, 451 quantified, 115 proteins quantified by 5 or more peptides with less than 10% CV values). However, the comparison of concentrations obtained in the unfractionated sample and including ultrafiltration shows already substantiate discrepancies in the case of medium abundant proteins, and even higher for low abundant proteins (Publication III, Figure 1B). This was rather unexpected due to the fact that 10 kDa cutoff membranes were used for ultrafiltration, while 30 kDa membranes were used in FASP (see Publication III, Materials and Methods). Apparently, even this step of additional sample processing, which potentially does not target the analyzed proteins directly, may significantly influence the outcome. The analysis of LMWF by the Quad-Orbitrap workflow yielded 157 identifications with as much as 69 proteins not detected in other experiments. Immunodepletion resulted in the highest number of 665 identified proteins with 198 proteins found only in this experiment, whereas the HAPs fraction contained the lowest number (among HMWF experiments) at 324 identified proteins, which exceeded the anticipated 14 proteins to an even greater extent than in the case of the Triple Quad-TOF workflow. These comprehensive outcomes achieved only after a simple one-step fractionation procedure demonstrate the great potential of the utilized instrument and the developed workflow in achieving a comprehensive view of the proteomic landscape despite the hFF material analysis issue of considerable differences between dynamic protein concentrations (see Introduction 3.2. paragraph).

1.3. Compatibility of developed workflows

A complete separation of all analytical stages between the developed workflows (sample processing, LC-MS/MS measurements, data analysis) as well as different intentions for their optimization do not allow for their direct comparison in terms of efficiency and performance (see Publication III, Table III and Discussion). However, the use of TPA computational approach in both workflows resulted in absolute protein concentration outcomes, which can be compared directly to assess their compatibility in the discovery of meaningful biological information. The agreement of the developed workflows was tested on the shared fraction of 124 proteins quantified in unfractionated sample (see Figure 4A) and is presented on a scatterplot of median concentrations (see Figure 4B). Moreover, a very high Pearson correlation coefficient of 0.86 was obtained for compared median concentrations (see Publication III, Results 2.5.1. paragraph). These results demonstrate good compatibility between the devised workflows, and therefore support

the further choice of one workflow based on the situation, e.g., instrument and resource availability, number of samples, or targeted set of proteins. Further experiments on workflow compatibility were performed on a small set of clinical samples and are presented in the Results "Pilot clinical studies" section.

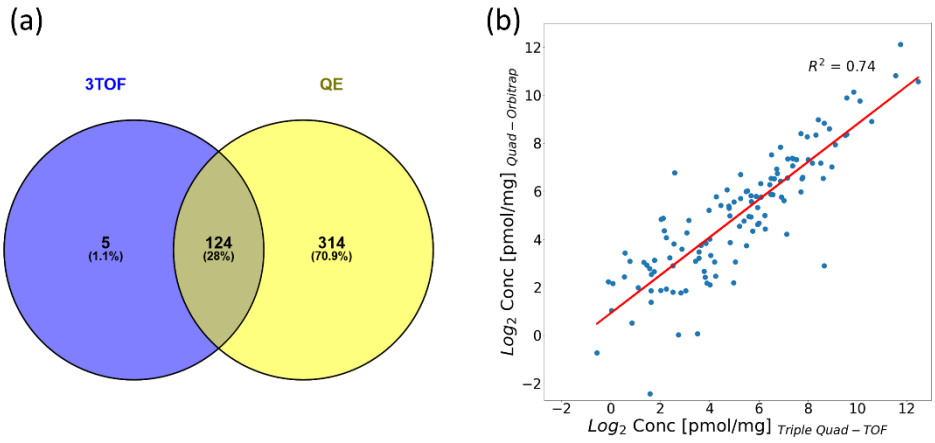


Figure 4. General comparison of compatibility of quantification capabilities of tested workflows. (A) Venn diagram illustrating numbers of quantified proteins by the Triple Quad-TOF (3TOF) and Quad-Orbitrap (QE) workflows. (B) Scatterplot of log2-transformed median concentration values from the Triple Quad-TOF and Quad-Orbitrap workflows. Adapted from Publication III (Figure 3).

2. Pilot clinical studies: data analysis strategy and biomarker candidates

In Publications I and III, trial clinical studies were performed on small numbers of samples obtained from patients undergoing fertility treatment at the INVICTA clinic. The purpose of these analyses was to test the performance of the developed methodology in a real clinical setting as well as to establish further procedures of statistical analysis, and to determine biomarker candidates which could be further verified in a large scale clinical study planned in the future.

2.1. Clinical study (Publication I)

In the first pilot study, samples from three individual follicles of four donors (12 samples in total) without any further clinical information on embryo culture outcomes were examined (see Publication I, Material and Methods). Due to the small number of included samples, the main purpose was to identify differences in protein concentrations between donors, and most importantly between the follicles of a single donor. Because of the reported differences in oocyte fertilization potential between oocytes from the same donor[47], the proteins present at different concentrations in these

oocytes' vicinity might be associated with the physiological state of the oocyte, whereas differences between donors most likely point to personal traits (see Introduction 3.3. paragraph). Therefore, two groups of t-tests were performed: (i) between each pair of donors (including all samples from all respective follicles) and (ii) between each pair of follicles. The results for HMWF are shown in Table 2 of Publication I with 17 different proteins found at significantly different concentration (p-value < 0.05, fold change > 2) for each group of comparisons (I, ii). These outcomes were also compared to previous hFF proteomic studies to search for similarities in results and are summed up in Table 3 of Publication I. To select the potential candidates for biomarkers reflecting inter-follicle differences in single patients, 6 proteins detected in both groups of comparisons were excluded, resulting in remaining 11 proteins divided into (i) proteins reported as significant in literature and (ii) proteins identified in other hFF proteomic publications, not reported as significant in conducted experiments, listed here in Table 4 (a modified section of Table 3 from Publication I). These candidates must be further validated on much larger numbers of clinical samples, yet they constitute an initial glimpse into a set of hFF proteins that might be associated with oocyte quality.

Table 3. List of proteins detected at significantly different concentrations in follicles of single patients including relevant literature reports and the fraction where a protein has been found. Adapted from Publication I (Table 3).

Group	Proteins	Literature reports as significant	Fraction	
			HMWF	LMWF
I. Proteins present at different concentrations among oocytes, reported as important in other studies	Antithrombin-III	[50,64,65,68]	✓	✓
	Complement factor I	[51,65–67]	✓	
	Leucine-rich alpha-2-glycoprotein	[53,64]	✓	
	Complement factor B	[51,64]	✓	✓
	Alpha-2-antiplasmin	[62,64]	✓	✓
	Afamin	[62,63]	✓	
	Apolipoprotein D	[51]	✓	
	Inter-alpha-trypsin inhibitor heavy chain H4	[51]	✓	✓
II. Proteins present at different concentrations among oocytes, identified in other studies	Insulin-like growth factor-binding protein complex acid labile subunit	-	✓	
	Heparin cofactor 2	-	✓	✓
	Alpha-2-macroglobulin	-	✓	

In the LMWF quantitative clinical study conducted on the same samples, 21 peptides originating from 12 proteins were distinguished as differential between any two single follicles (see Publication I, Table 4). One of these peptides, the sequence IHWESASLL from complement component C3 was also determined as a potential biomarker in a peptidomic study by Chen et al.[55]. No specific

differences between follicles of the same donor were identified. However, these results should be interpreted with extreme caution as mentioned before in the case of LMWF analysis (Results 1.1. paragraph).

2.2. Clinical study (Publication III)

Samples from twenty follicles of four donors were analyzed in Publication III by both developed workflows taking into account the clinical information on retrieved oocyte maturity and outcomes of embryo culture (5 individual follicles per donor including all considered clinical outcomes, see Publication III, Materials and Methods and Table S9). The numbers of quantified proteins were: 215 by the Triple Quad-TOF workflow, 455 by the Quad-Orbitrap workflow, and 199 shared by both workflows (see Publication III, Tables S10-S11). After checking the consistency of obtained data by multiscatter plots and inter-workflow comparisons of correlation, one of the samples was excluded from the study as a potential outlier (see Publication III, 2.5.2. paragraph, Figures S2-S4, Tables S12-S13). The agreement of both workflows was further analyzed by calculating Pearson correlation coefficient values for (i) individual clinical samples and (ii) single proteins (Publication III, Tables S12-S13). All samples (after the exclusion of the outlier) presented high values at 0.75, however only about 20% of proteins shown values higher than 0.7, irrespective of their general concentration (see Publication III, Figure S4). This result shows that both workflows are more consistent in detecting differences between the samples but differ in the elucidation of the exact protein concentration. The samples were divided into study groups according to obtained clinical information on the retrieved oocyte status considering two criteria: (i) oocyte maturity and (ii) development of blastocyst from fertilized oocyte (see Publication III, Table S9). To be able to discern the effect associated with these events on the hFF proteome from individual traits of the donors, a two-way ANOVA (analysis of variance) was applied in the statistical analysis of the results. Two rounds of analysis were carried out with one factor grouping samples from the same donor: the first grouping the mature (n=14) and immature (n=5) oocytes, and the second grouping remaining mature oocyte-associated samples into developed (n=8) and not developed (n=6) blastocyst as the other factor. The results are listed in Table S14 and thoroughly described in 2.5.2. paragraph of Publication III, therefore the following is the summary of the most important observations. As expected, the factor related to individual features of the patients had the predominant impact on the differences between the samples with almost half and one third of the proteins displaying significant concentration changes at 5% and 1% FDR, respectively (see Publication III, Figure S5). More proteins associated with the oocyte and blastocyst status were appointed to be significant by the Quad-Orbitrap workflow with 49 and 45 proteins, respectively, in comparison to 10 and 7 proteins appointed by the Triple Quad-TOF workflow

(compare the interaction networks in Publication III, Figure S6). All of the Triple Quad-TOF workflow-appointed proteins were also quantified by the Quad-Orbitrap workflow, while about 45-65% of the Quad-Orbitrap workflow-appointed proteins were quantified by the Triple Quad-TOF workflow. In general, most of the designated proteins presented only minimal fold changes. Only one protein was appointed by both workflows in each of the comparisons: hepatocyte growth factor-like protein for oocyte maturity (showing opposing directions of fold change), and carboxypeptidase B2 for blastocyst development (present at lower concentrations in the developed blastocyst group). In order to test the compatibility of both workflows on statistical analysis results and further refine the lists of biomarker candidates, ratios of median fold changes of 199 commonly quantified proteins were calculated for both comparisons (see Publication III, Figure 4A,B). This analysis also revealed a good agreement between both workflows demonstrated by small fold change differences; however, this outcome could arise from generally minimal fold changes. The lists of proteins appointed as significant by each workflow were filtered using ≤ 20% fold change ratio resulting in the sets of 20 and 22 proteins associated with oocyte maturity and blastocyst development, respectively. These lists were compared with previous literature reports (see Publication III, Table S15)[48,49,54,55,59,61,62,64,66,69–72], and are presented here as interaction networks in Figure 5 adapted from Figure 4C,D of Publication III. These biomarker candidates also need to be validated on a larger set of samples. Interestingly, 7 of the previously determined proteins (Publication I, see Table 3) were also found to be significant in this experiment. However, in this case, the data analysis methodology was selected to include the clinical data; therefore, it could be applied in the same form to a large-scale clinical study.

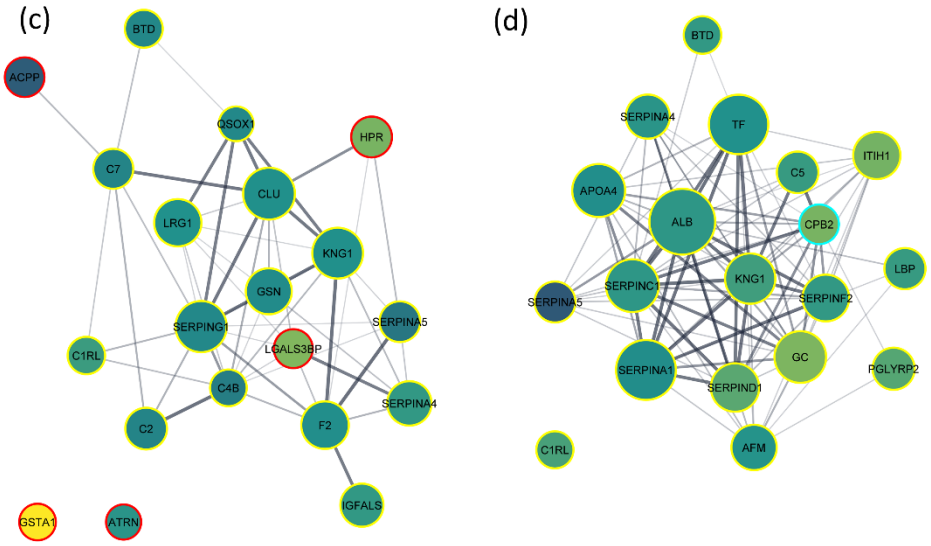


Figure 5. Interaction networks for statistically significant proteins related to (a) oocyte maturity and (b) blastocyst development status with less than 20%-fold change ratio difference between quantitative methods. Node edge color designates the quantitative method used to establish statistical significance (yellow—Quad-Orbitrap; red—Triple Quad-TOF; and both workflows—cyan). Fill color relates to the mean fold change of both methods, from 0.5 and below (yellow) through 1 (aquamarine) to 2 and above (purple). Node size represents mean log10 median abundance TPA concentrations of both methods in the test group (either mature oocyte or developed blastocyst). Adapted from Publication III (Figure 4C,D).

3. HFF proteome and peptidome characterization

Besides the development and tests of the proteomic workflows suitable for hFF analysis, the characterization of obtained sets of identified proteins was also carried out as a part of the study. In the Publication I, 103 proteins were identified in HMWF. This is not a particularly comprehensive result in comparison to other hFF studies (see Table 1); however, it was expected from a trial study using only simple fractionation strategies (immunodepletion, ultrafiltration). 91 proteins were identified in LMWF and most of them were not found in HMWF (55). Moreover, 14 of the HMWF identified proteins (a comparatively large percentage of the identifications) and as much as 45 of the LMWF proteins have not been reported in hFF studies before at the time of publication (see Publication I, Table 1, Figure 3). This outcome hinted at the potential presence of yet undiscovered proteome, which might play a role in elucidation of oocyte quality markers. It is especially applicable to peptidomic research, which has usually been omitted in respect to hFF analysis (with the exception

of the work of Chen et al.[55]). A functional analysis of identified proteins has been conducted and is described in detail in Publication I, "Functional Analysis" paragraph (see also Publication I, Figures 2, S69). After the addition of high pH RP-HPLC to the fractionation scheme in Publication II, the number of identifications increased to 302 in HMWF and 161 in LMWF (see Publication II, Figures 1, 2). This result was also anticipated as compared to the literature[51,52,54,56,57,60,61]. A similar functional analysis of this set of proteins is described in Publication II, 3.5. paragraph (see also Publication II, Figures S1-S3). A number of various experiments conducted by two different workflows in the Publication III resulted in the most inclusive identification lists with 1177 and 873 (HMWF) proteins identified using the Triple Quad-TOF and the Quad-Orbitrap workflows, respectively. In this study, the research focus was shifted away from the peptidomic fraction, yet few such experiments were also carried out resulting in 14 and 157 proteins identified by the Triple Quad-TOF, and the Quad-Orbitrap workflows, respectively. All sets of identified proteins were compared to the most comprehensive proteomic studies of hFF[48,54,56,59–61] as well as to other relevant materials, which might contribute to hFF final composition, i.e., blood plasma/serum[76], granulosa cells[77], and the oocyte[45] (see Publication III, Table S8). The summary of these results is also presented here in Table 4 adapted from Table 2 of Publication III. The comparison of protein identifications revealed the expected sources contributing mostly to hFF composition: plasma/serum, granulosa cells, and oocyte with about 80, 50, and 25% overlap, respectively. The number of identifications shared with other comprehensive hFF studies depended mostly on a total number of reported identifications (see Table 1).

Table 4. Numbers of proteins identified in this study by Quad-Orbitrap (QE) or Triple Quad-TOF (3TOF) workflows, which were also reported in proteomic studies of hFF or proximate biological materials (plasma, oocyte, granulosa cell). Adapted from Publication III (Table 2).

Resource	All identified proteins		All proteins identified in HMW fraction		Proteins identified only in HMW fraction		All proteins identified in LMW fraction		Proteins identified only in LMW fraction	
	QE (942)	3TOF (1182)	QE (873)	3TOF (1177)	QE (785)	3TOF (1168)	QE (157)	3TOF (14)	QE (69)	3TOF (5)
Plasma Proteome Database[76]	773	975	723	975	644	966	129	9	50	0
Human oocyte[45] (oocyte specific)	226 (22)	301 (18)	211 (22)	301 (18)	183 (18)	294 (17)	43 (4)	7 (1)	15	0
Human granulosa cell[77]	436	599	390	599	346	591	90	8	46	0
HFF (Zamah et al., 2015)[54]	545	610	542	610	470	602	75	8	3	0
HFF (Bianchi et al., 2016)[48]	357	368	352	368	284	360	73	8	5	0
HFF (Oh et al., 2017)[56]	521	534	518	534	445	525	76	9	3	0
HFF (Poulsen et al., 2019)[59]	336	330	333	330	269	321	67	9	3	0
HFF (Zhang et al., 2019)[60]	567	647	540	647	463	640	104	7	27	0
HFF from hSAF (Pla et al., 2020)[61]	829	987	794	987	708	978	121	9	35	0
Unique	28	40	20	35	20	35	8	5	8	5

The purpose of functional analysis conducted in Publication III was to use the obtained absolute concentrations to see how groups of proteins contribute to total protein content in hFF. Therefore, the HMWF proteins quantified in unfractionated sample by the sensitive Quad-Orbitrap workflow were divided into concentration groups as described in Publication III, 2.2.2. paragraph. In addition, proteins identified only after immunodepletion were also included in this analysis to study even lower abundant proteome, concealed by HAPs. These results are presented in Table S4 of Publication III. A similar analysis was carried out for LMWF proteins detected by the Quad-Orbitrap workflow, divided into all LMWF-identified proteins and proteins uniquely identified in LMWF along with interaction networks shown in Figure S1A,B (see Publication III, Table S5 and 2.2.2. paragraph).

Conclusions and Future Perspectives

Two distinct label-free quantitative proteomic workflows for hFF analysis associated with different LC-MS instrument setups have been developed in the course of this work. The Triple Quad-TOF workflow has been optimized across three presented publications (Publication I-III) to provide sufficient proteome coverage (with the potential of moderate increase along with the number of samples) at a relatively low resource/time cost. The Quad-Orbitrap workflow offers very high resolution and sensitivity in each single measurement at higher resource/time cost (Publication III). Both workflows have a high degree of compatibility in obtained results; therefore, the choice of a suitable methodology for a large-scale clinical study directed at elucidation of oocyte quality biomarkers should be dictated by the available resources and anticipated depth of the analysis. The relevant workflow could be optimized even further to lower the time/resource cost, e.g., by the use of shorter LC gradients (especially in the case of Triple Quad-TOF[78]), by decreasing digestion time[79] or application of modifications of FASP procedure[80], and by elimination of the alkylation step in sample preparation[81]. The quality of quantification could also be improved at the cost of lower proteome coverage by increasing the number of peptides required for analysis or implementing the MaxLFQ algorithm in the case of the Quad-Orbitrap workflow[82]. Both workflows have been successfully tested in small-scale clinical studies, and potential candidates for proteins associated with oocyte maturity and blastocyst development have been determined and could be further validated in future studies. The data analysis protocols described here can also be implemented for the analysis of large number of clinical samples to enable identification of oocyte quality biomarkers. The conducted trial peptidomic experiments demonstrate a great potential of peptidomics in discovery of new, previously omitted biological information in hFF, which could be followed in future experiments. Sets of proteins identified and quantified across all presented publications have been characterized by functional analysis and discussed against relevant literature to provide a comprehensive view of the hFF proteomic and peptidomic landscape. Moreover, mass spectrometry data obtained in both proteomic and peptidomic experiments, including quantitative experiments, could be reanalyzed by searching against emerging databases containing possible protein variants as annotated in open reading frames of human genome, such as OpenProt[83]. This analysis might potentially uncover more information on the processes occurring in the environment around the oocyte along with the novel attractive targets for oocyte quality biomarkers.

References

(1) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198. https://doi.org/10.1038/nature01511.

(2) Han, X.; Aslanian, A.; Yates, J. R. Mass Spectrometry for Proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12* (5), 483–490. https://doi.org/10.1016/j.cbpa.2008.07.024.

(3) Noor, Z.; Ahn, S. B.; Baker, M. S.; Ranganathan, S.; Mohamedali, A. Mass Spectrometry–Based Protein Identification in Proteomics—a Review. *Brief. Bioinform.* **2021**, *22* (2), 1620–1638. https://doi.org/10.1093/bib/bbz163.

(4) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al* **2009**, *CHAPTER*, Unit10.25. https://doi.org/10.1002/0471142727.mb1025s88.

(5) Wiśniewski, J. R.; Rakus, D. Multi-Enzyme Digestion FASP and the 'Total Protein Approach'-Based Absolute Quantification of the Escherichia Coli Proteome. *J. Proteomics* **2014**, *109*, 322–331. https://doi.org/10.1016/j.jprot.2014.07.012.

(6) Sinitcyn, P.; Rudolph, J. D.; Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.* **2018**, *1* (1), 207–234. https://doi.org/10.1146/annurev-biodatasci-080917-013516.

(7) Aggarwal, S.; Yadav, A. K. False Discovery Rate Estimation in Proteomics. In *Statistical Analysis in Proteomics*; Jung, K., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp 119–128. https://doi.org/10.1007/978-1-4939-3106-4_7.

(8) Manadas, B.; Mendes, V. M.; English, J.; Dunn, M. J. Peptide Fractionation in Proteomics Approaches. *Expert Rev. Proteomics* **2010**, *7* (5), 655–663. https://doi.org/10.1586/epr.10.46.

(9) Mostovenko, E.; Hassan, C.; Rattke, J.; Deelder, A. M.; van Veelen, P. A.; Palmblad, M. Comparison of Peptide and Protein Fractionation Methods in Proteomics. *EuPA Open Proteomics* **2013**, *1*, 30–37. https://doi.org/10.1016/j.euprot.2013.09.001.

(10) Wu, C.; Duan, J.; Liu, T.; Smith, R. D.; Qian, W.-J. Contributions of Immunoaffinity Chromatography to Deep Proteome Profiling of Human Biofluids. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **2016**, *1021*, 57–68. https://doi.org/10.1016/j.jchromb.2016.01.015.

(11) Li, C.; Lee, K. H. Affinity Depletion of Albumin from Human Cerebrospinal Fluid Using Cibacron-Blue-3G-A-Derivatized Photopatterned Copolymer in a Microfluidic Device. *Anal. Biochem.* **2004**, *333* (2), 381–388. https://doi.org/10.1016/j.ab.2004.06.007.

(12) Magdeldin, S.; Enany, S.; Yoshida, Y.; Xu, B.; Zhang, Y.; Zureena, Z.; Lokamani, I.; Yaoita, E.; Yamamoto, T. Basics and Recent Advances of Two Dimensional- Polyacrylamide Gel Electrophoresis. *Clin. Proteomics* **2014**, *11* (1), 16. https://doi.org/10.1186/1559-0275-11-16.

(13) Batth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-PH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13* (12), 6176–6186. https://doi.org/10.1021/pr500893m.

(14) Han, D.; Jin, J.; Woo, J.; Min, H.; Kim, Y. Proteomic Analysis of Mouse Astrocytes and Their Secretome by a Combination of FASP and StageTip-Based, High PH, Reversed-Phase Fractionation. *Proteomics* **2014**, *14* (13–14), 1604–1609. https://doi.org/10.1002/pmic.201300495.

(15) Greening, D. W.; Simpson, R. J. A Centrifugal Ultrafiltration Strategy for Isolating the Low-Molecular Weight (≤25K) Component of Human Plasma Proteome. *J. Proteomics* **2010**, *73* (3), 637–648. https://doi.org/10.1016/j.jprot.2009.09.013.

(16) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. Characterization of the Low Molecular Weight Human Serum Proteome*S. *Mol. Cell. Proteomics* **2003**, *2* (10), 1096–1103. https://doi.org/10.1074/mcp.M300031-MCP200.

(17) Vollet Marson, G.; Belleville, M.-P.; Lacour, S.; Dupas Hubinger, M. Membrane Fractionation of Protein Hydrolysates from By-Products: Recovery of Valuable Compounds from Spent Yeasts. *Membranes* **2021**, *11* (1), 23. https://doi.org/10.3390/membranes11010023.

(18) Griffiths, J. A Brief History of Mass Spectrometry. *Anal. Chem.* **2008**, *80* (15), 5678–5683. https://doi.org/10.1021/ac8013065.

(19) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity during Cell Differentiation. *Genome Biol.* **2018**, *19* (1), 161. https://doi.org/10.1186/s13059-018-1547-5.

(20) Karpievitch, Y. V.; Polpitiya, A. D.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. *Ann. Appl. Stat.* **2010**, *4* (4), 1797–1823. https://doi.org/10.1214/10-AOAS341.

(21) Bian, Y.; Zheng, R.; Bayer, F. P.; Wong, C.; Chang, Y.-C.; Meng, C.; Zolg, D. P.; Reinecke, M.; Zecha, J.; Wiechmann, S.; Heinzlmeir, S.; Scherr, J.; Hemmer, B.; Baynham, M.; Gingras, A.-C.; Boychenko, O.; Kuster, B. Robust, Reproducible and Quantitative Analysis of Thousands of Proteomes by Micro-Flow LC–MS/MS. *Nat. Commun.* **2020**, *11* (1), 157. https://doi.org/10.1038/s41467-019-13973-x.

(22) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296. https://doi.org/10.1021/ac4001223.

(23) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative Mass Spectrometry in Proteomics: A Critical Review. *Anal. Bioanal. Chem.* **2007**, *389* (4), 1017–1031. https://doi.org/10.1007/s00216-007-1486-6.

(24) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics MCP* **2012**, *11* (6), O111.016717. https://doi.org/10.1074/mcp.O111.016717.

(25) Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* **2015**, *14* (10), 4359–4371. https://doi.org/10.1021/acs.jproteome.5b00543.

(26) Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14* (8), e8126. https://doi.org/10.15252/msb.20178126.

(27) Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; Prianichnikov, N.; Yılmaz, Ş.; Rudolph, J. D.; Tenzer, S.; Perez-Riverol, Y.; Nagaraj, N.; Humphrey, S. J.; Cox, J. MaxDIA Enables Library-Based and Library-Free Data-Independent Acquisition Proteomics. *Nat. Biotechnol.* **2021**, 1–11. https://doi.org/10.1038/s41587-021-00968-7.

(28) Extensive Quantitative Remodeling of the Proteome between Normal Colon Tissue and Adenocarcinoma. *Mol. Syst. Biol.* **2012**, *8* (1), 611. https://doi.org/10.1038/msb.2012.44.

(29) He, B.; Shi, J.; Wang, X.; Jiang, H.; Zhu, H.-J. Label-Free Absolute Protein Quantification with Data-Independent Acquisition. *J. Proteomics* **2019**, *200*, 51–59. https://doi.org/10.1016/j.jprot.2019.03.005.

(30) ICD-11 https://icd.who.int/en (accessed 2021 -12 -02).

(31) Mascarenhas, M. N.; Flaxman, S. R.; Boerma, T.; Vanderpoel, S.; Stevens, G. A. National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys. *PLOS Med.* **2012**, *9* (12), e1001356. https://doi.org/10.1371/journal.pmed.1001356.

(32) Boivin, J.; Bunting, L.; Collins, J. A.; Nygren, K. G. International Estimates of Infertility Prevalence and Treatment-Seeking: Potential Need and Demand for Infertility Medical Care. *Hum. Reprod.* **2007**, *22* (6), 1506–1512. https://doi.org/10.1093/humrep/dem046.

(33) WHO | Infecundity, infertility, and childlessness in developing countries. Demographic and Health Surveys (DHS) Comparative reports No. 9 https://www.who.int/reproductivehealth/publications/infertility/DHS_9/en/ (accessed 2021 -12 -02).

(34) Overview | Fertility problems: assessment and treatment | Guidance | NICE https://www.nice.org.uk/guidance/cg156 (accessed 2021 -12 -02).

(35) Segal, T. R.; Giudice, L. C. Before the Beginning: Environmental Exposures and Reproductive and Obstetrical Outcomes. *Fertil. Steril.* **2019**, *112* (4), 613–621. https://doi.org/10.1016/j.fertnstert.2019.08.001.

(36) Ombelet, W.; Cooke, I.; Dyer, S.; Serour, G.; Devroey, P. Infertility and the Provision of Infertility Medical Services in Developing Countries. *Hum. Reprod. Update* **2008**, *14* (6), 605–621. https://doi.org/10.1093/humupd/dmn042.

(37) Adamson, G. D.; Tabangin, M.; Macaluso, M.; Mouzon, J. de. The Number of Babies Born Globally after Treatment with the Assisted Reproductive Technologies (ART). *Fertil. Steril.* **2013**, *100* (3), S42. https://doi.org/10.1016/j.fertnstert.2013.07.1807.

(38) Faddy, M. J.; Gosden, M. D.; Gosden, R. G. A Demographic Projection of the Contribution of Assisted Reproductive Technologies to World Population Growth. *Reprod. Biomed. Online* **2018**, *36* (4), 455–458. https://doi.org/10.1016/j.rbmo.2018.01.006.

(39) Zegers-Hochschild, F.; Adamson, G. D.; de Mouzon, J.; Ishihara, O.; Mansour, R.; Nygren, K.; Sullivan, E.; Vanderpoel, S. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) Revised Glossary of ART Terminology, 2009∗. *Fertil. Steril.* **2009**, *92* (5), 1520–1524. https://doi.org/10.1016/j.fertnstert.2009.09.009.

(40) Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 13 lutego 2020 r. w sprawie ogłoszenia jednolitego tekstu ustawy o leczeniu niepłodności http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200000442 (accessed 2021 -12 -02).

(41) The European IVF-Monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE); Wyns, C.; De Geyter, C.; Calhaz-Jorge, C.; Kupka, M. S.; Motrenko, T.; Smeenk, J.; Bergh, C.; Tandler-Schneider, A.; Rugescu, I. A.; Vidakovic, S.; Goossens, V. ART in Europe, 2017: Results Generated from European Registries by ESHRE†. *Hum. Reprod. Open* **2021**, *2021* (3), hoab026. https://doi.org/10.1093/hropen/hoab026.

(42) De Geyter, C.; Wyns, C.; Calhaz-Jorge, C.; de Mouzon, J.; Ferraretti, A. P.; Kupka, M.; Nyboe Andersen, A.; Nygren, K. G.; Goossens, V. 20 Years of the European IVF-Monitoring Consortium Registry: What Have We Learned? A Comparison with Registries from Two Other Regions. *Hum. Reprod.* **2020**, *35* (12), 2832–2849. https://doi.org/10.1093/humrep/deaa250.

(43) Ozturk, S. Selection of Competent Oocytes by Morphological Criteria for Assisted Reproductive Technologies. *Mol. Reprod. Dev.* **2020**, *87* (10), 1021–1036. https://doi.org/10.1002/mrd.23420.

(44) Lazzaroni-Tealdi, E.; Barad, D. H.; Albertini, D. F.; Yu, Y.; Kushnir, V. A.; Russell, H.; Wu, Y.-G.; Gleicher, N. Oocyte Scoring Enhances Embryo-Scoring in Predicting Pregnancy Chances with IVF

Where It Counts Most. *PLOS ONE* **2015**, *10* (12), e0143632. https://doi.org/10.1371/journal.pone.0143632.

(45) Virant-Klun, I.; Leicht, S.; Hughes, C.; Krijgsveld, J. Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. *Mol. Cell. Proteomics* **2016**, *15* (8), 2616–2627. https://doi.org/10.1074/mcp.M115.056887.

(46) Basuino, L.; Silveira, C. F. Human Follicular Fluid and Effects on Reproduction. *JBRA Assist. Reprod.* **2016**, *20* (1), 38–40. https://doi.org/10.5935/1518-0557.20160009.

(47) Revelli, A.; Piane, L. D.; Casano, S.; Molinari, E.; Massobrio, M.; Rinaudo, P. Follicular Fluid Content and Oocyte Quality: From Single Biochemical Markers to Metabolomics. *Reprod. Biol. Endocrinol.* **2009**, *7* (1), 40. https://doi.org/10.1186/1477-7827-7-40.

(48) Bianchi, L.; Gagliardi, A.; Landi, C.; Focarelli, R.; Leo, V. D.; Luddi, A.; Bini, L.; Piomboni, P. Protein Pathways Working in Human Follicular Fluid: The Future for Tailored IVF? *Expert Rev. Mol. Med.* **2016**, *18*. https://doi.org/10.1017/erm.2016.4.

(49) Bayasula; Iwase, A.; Kobayashi, H.; Goto, M.; Nakahara, T.; Nakamura, T.; Kondo, M.; Nagatomo, Y.; Kotani, T.; Kikkawa, F. A Proteomic Analysis of Human Follicular Fluid: Comparison between Fertilized Oocytes and Non-Fertilized Oocytes in the Same Patient. *J. Assist. Reprod. Genet.* **2013**, *30* (9), 1231–1238. https://doi.org/10.1007/s10815-013-0004-3.

(50) Ambekar, A. S.; Kelkar, D. S.; Pinto, S. M.; Sharma, R.; Hinduja, I.; Zaveri, K.; Pandey, A.; Prasad, T. S. K.; Gowda, H.; Mukherjee, S. Proteomics of Follicular Fluid From Women With Polycystic Ovary Syndrome Suggests Molecular Defects in Follicular Development. *J. Clin. Endocrinol. Metab.* **2015**, *100* (2), 744–753. https://doi.org/10.1210/jc.2014-2086.

(51) Regiani, T.; Cordeiro, F.; Do, L.; Costa, V.; Salgueiro, J.; Cardozo, K.; Carvalho, V.; Perkel, K.; Zylbersztejn, D.; Cedenho, A.; Guimarães, E.; Turco, E. Follicular Fluid Alterations in Endometriosis: Label-Free Proteomics by MSE as a Functional Tool for Endometriosis. *Syst. Biol. Reprod. Med.* **2015**, *61*, 1939–6376. https://doi.org/10.3109/19396368.2015.1037025.

(52) Twigt, J. M.; Bezstarosti, K.; Demmers, J.; Lindemans, J.; Laven, J. S. E.; Steegers-Theunissen, R. P. Preconception Folic Acid Use Influences the Follicle Fluid Proteome. *Eur. J. Clin. Invest.* **2015**, *45* (8), 833–841. https://doi.org/10.1111/eci.12478.

(53) Wu, Y.-T.; Wu, Y.; Zhang, J.-Y.; Hou, N.-N.; Liu, A.-X.; Pan, J.-X.; Lu, J.-Y.; Sheng, J.-Z.; Huang, H.-F. Preliminary Proteomic Analysis on the Alterations in Follicular Fluid Proteins from Women Undergoing Natural Cycles or Controlled Ovarian Hyperstimulation. *J. Assist. Reprod. Genet.* **2015**, *32* (3), 417–427. https://doi.org/10.1007/s10815-014-0419-5.

(54) Zamah, A. M.; Hassis, M. E.; Albertolle, M. E.; Williams, K. E. Proteomic Analysis of Human Follicular Fluid from Fertile Women. *Clin. Proteomics* **2015**, *12* (1), 5. https://doi.org/10.1186/s12014-015-9077-6.

(55) Chen, F.; Spiessens, C.; D'Hooghe, T.; Peeraer, K.; Carpentier, S. Follicular Fluid Biomarkers for Human in Vitro Fertilization Outcome: Proof of Principle. *Proteome Sci.* **2016**, *14*, 17. https://doi.org/10.1186/s12953-016-0106-9.

(56) Oh, J. W.; Kim, S. K.; Cho, K.-C.; Kim, M.-S.; Suh, C. S.; Lee, J. R.; Kim, K. P. Proteomic Analysis of Human Follicular Fluid in Poor Ovarian Responders during in Vitro Fertilization. *PROTEOMICS* **2017**, *17* (6), 1600333. https://doi.org/10.1002/pmic.201600333.

(57) Shen, X.; Liu, X.; Zhu, P.; Zhang, Y.; Wang, J.; Wang, Y.; Wang, W.; Liu, J.; Li, N.; Liu, F. Proteomic Analysis of Human Follicular Fluid Associated with Successful in Vitro Fertilization. *Reprod. Biol. Endocrinol. RBE* **2017**, *15* (1), 58. https://doi.org/10.1186/s12958-017-0277-y.

(58) Lehmann, R.; Schmidt, A.; Pastuschek, J.; Müller, M. M.; Fritzsche, A.; Dieterle, S.; Greb, R. R.; Markert, U. R.; Slevogt, H. Comparison of Sample Preparation Techniques and Data Analysis for the LC-MS/MS-Based Identification of Proteins in Human Follicular Fluid. *Am. J. Reprod. Immunol.* **2018**, *80* (2), e12994. https://doi.org/10.1111/aji.12994.

(59) Poulsen, L. la C.; Pla, I.; Sanchez, A.; Grøndahl, M. L.; Marko-Varga, G.; Yding Andersen, C.; Englund, A. L. M.; Malm, J. Progressive Changes in Human Follicular Fluid Composition over the Course of Ovulation: Quantitative Proteomic Analyses. *Mol. Cell. Endocrinol.* **2019**, *495*, 110522. https://doi.org/10.1016/j.mce.2019.110522.

(60) Zhang, X.; Xu, X.; Li, P.; Zhou, F.; Kong, L.; Qiu, J.; Yuan, Z.; Tan, J. TMT Based Proteomic Analysis of Human Follicular Fluid From Overweight/Obese and Normal-Weight Patients With Polycystic Ovary Syndrome. *Front. Endocrinol.* **2019**, *10*, 821. https://doi.org/10.3389/fendo.2019.00821.

(61) Pla, I.; Sanchez, A.; Pors, S. E.; Pawlowski, K.; Appelqvist, R.; Sahlin, K. B.; Poulsen, L. L. C.; Marko-Varga, G.; Andersen, C. Y.; Malm, J. Proteome of Fluid from Human Ovarian Small Antral Follicles Reveals Insights in Folliculogenesis and Oocyte Maturation. *Hum. Reprod.* **2021**, *36* (3), 756–770. https://doi.org/10.1093/humrep/deaa335.

(62) Kushnir, M. M.; Naessén, T.; Wanggren, K.; Rockwood, A. L.; Crockett, D. K.; Bergquist, J. Protein and Steroid Profiles in Follicular Fluid after Ovarian Hyperstimulation as Potential Biomarkers of IVF Outcome. *J. Proteome Res.* **2012**, *11* (10), 5090–5100. https://doi.org/10.1021/pr300535g.

(63) Hanrieder, J.; Zuberovic, A.; Bergquist, J. Surface Modified Capillary Electrophoresis Combined with in Solution Isoelectric Focusing and MALDI-TOF/TOF MS: A Gel-Free Multidimensional Electrophoresis Approach for Proteomic Profiling—Exemplified on Human Follicular Fluid. *J. Chromatogr. A* **2009**, *1216* (17), 3621–3628. https://doi.org/10.1016/j.chroma.2008.12.026.

(64) Bianchi, L.; Gagliardi, A.; Campanella, G.; Landi, C.; Capaldo, A.; Carleo, A.; Armini, A.; De Leo, V.; Piomboni, P.; Focarelli, R.; Bini, L. A Methodological and Functional Proteomic Approach of Human Follicular Fluid En Route for Oocyte Quality Evaluation. *J. Proteomics* **2013**, *90*, 61–76. https://doi.org/10.1016/j.jprot.2013.02.025.

(65) Lo Turco, E. G.; Cordeiro, F. B.; Lopes, P. H. de C.; Gozzo, F. C.; Pilau, E. J.; Soler, T. B.; da Silva, B. F.; Del Giudice, P. T.; Bertolla, R. P.; Fraietta, R.; Cedenho, A. P. Proteomic Analysis of Follicular Fluid from Women with and without Endometriosis: New Therapeutic Targets and Biomarkers. *Mol. Reprod. Dev.* **2013**, *80* (6), 441–450. https://doi.org/10.1002/mrd.22180.

(66) Yoo, S. W.; Bolbot, T.; Koulova, A.; Sneeringer, R.; Humm, K.; Dagon, Y.; Usheva, A. Complement Factors Are Secreted in Human Follicular Fluid by Granulosa Cells and Are Possible Oocyte Maturation Factors. *J. Obstet. Gynaecol. Res.* **2013**, *39* (2), 522–527. https://doi.org/10.1111/j.1447-0756.2012.01985.x.

(67) Jarkovska, K.; Kupcova Skalnikova, H.; Halada, P.; Hrabakova, R.; Moos, J.; Rezabek, K.; Gadher, S. J.; Kovarova, H. Development of Ovarian Hyperstimulation Syndrome: Interrogation of Key Proteins and Biological Processes in Human Follicular Fluid of Women Undergoing in Vitro Fertilization. *Mol. Hum. Reprod.* **2011**, *17* (11), 679–692. https://doi.org/10.1093/molehr/gar047.

(68) Kim, Y.-S.; Kim, M.-S.; Lee, S.-H.; Choi, B.-C.; Lim, J.-M.; Cha, K. Y.; Baek, K.-H. Proteomic Analysis of Recurrent Spontaneous Abortion: Identification of an Inadequately Expressed Set of Proteins in Human Follicular Fluid. *Proteomics* **2006**, *6* (11), 3445–3454. https://doi.org/10.1002/pmic.200500775.

(69) Zakerkish, F.; Brännström, M.; Carlsohn, E.; Sihlbom, C.; van der Post, S.; Thoroddsen, A. Proteomic Analysis of Follicular Fluid during Human Ovulation. *Acta Obstet. Gynecol. Scand.* **2020**, *99* (7), 917–924. https://doi.org/10.1111/aogs.13805.

(70) Carbone, M. C.; Tatone, C.; Monache, S. D.; Marci, R.; Caserta, D.; Colonna, R.; Amicarelli, F. Antioxidant Enzymatic Defences in Human Follicular Fluid: Characterization and Age-dependent Changes. *Mol. Hum. Reprod.* **2003**, *9* (11), 639–643. https://doi.org/10.1093/molehr/gag090.

(71) Kleinman, D.; Insler, V.; Leiberman, J. R.; Glezerman, M.; Albotiano, S.; Potashnik, G.; Meizner, I. Acid Phosphatase Levels in Follicular Fluids Following Induction of Ovulation in in Vitro Fertilization Patients. *J. Vitro Fertil. Embryo Transf. IVF* **1987**, *4* (3), 181–184. https://doi.org/10.1007/BF01555468.

(72) Estes, S. J.; Ye, B.; Qiu, W.; Cramer, D.; Hornstein, M. D.; Missmer, S. A. A Proteomic Analysis of IVF Follicular Fluid in Women ≤32 Years Old. *Fertil. Steril.* **2009**, *92* (5), 1569–1578. https://doi.org/10.1016/j.fertnstert.2008.08.120.

(73) Homan, G. F.; Davies, M.; Norman, R. The Impact of Lifestyle Factors on Reproductive Performance in the General Population and Those Undergoing Infertility Treatment: A Review. *Hum. Reprod. Update* **2007**, *13* (3), 209–223. https://doi.org/10.1093/humupd/dml056.

(74) Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B. Current Peptidomics: Applications, Purification, Identification, Quantification, and

Functional Analysis. *Proteomics* **2015**, *15* (0), 1026–1038. https://doi.org/10.1002/pmic.201400310.

(75) Araki, Y.; Nonaka, D.; Tajima, A.; Maruyama, M.; Nitto, T.; Ishikawa, H.; Yoshitake, H.; Yoshida, E.; Kuronaka, N.; Asada, K.; Yanagida, M.; Nojima, M.; Yoshida, K.; Takamori, K.; Hashiguchi, T.; Maruyama, I.; Lee, L.-J.; Tanaka, K. Quantitative Peptidomic Analysis by a Newly Developed One-Step Direct Transfer Technology without Depletion of Major Blood Proteins: Its Potential Utility for Monitoring of Pathophysiological Status in Pregnancy-Induced Hypertension. *P*roteomics **2011**, *11* (13), 2727–2737. https://doi.org/10.1002/pmic.201000753.

(76) Nanjappa, V.; Thomas, J. K.; Marimuthu, A.; Muthusamy, B.; Radhakrishnan, A.; Sharma, R.; Ahmad Khan, A.; Balakrishnan, L.; Sahasrabuddhe, N. A.; Kumar, S.; Jhaveri, B. N.; Sheth, K. V.; Kumar Khatana, R.; Shaw, P. G.; Srikanth, S. M.; Mathur, P. P.; Shankar, S.; Nagaraja, D.; Christopher, R.; Mathivanan, S.; Raju, R.; Sirdeshmukh, R.; Chatterjee, A.; Simpson, R. J.; Harsha, H. C.; Pandey, A.; Prasad, T. S. K. Plasma Proteome Database as a Resource for Proteomics Research: 2014 Update. *Nucleic Acids Res.* **2014**, *42* (D1), D959–D965. https://doi.org/10.1093/nar/gkt1251.

(77) Bagnjuk, K.; Stöckl, J. B.; Fröhlich, T.; Arnold, G. J.; Behr, R.; Berg, U.; Berg, D.; Kunz, L.; Bishop, C.; Xu, J.; Mayerhofer, A. Necroptosis in Primate Luteolysis: A Role for Ceramide. *Cell Death Discov.* **2019**, *5* (1), 1–14. https://doi.org/10.1038/s41420-019-0149-7.

(78) Sun, R.; Hunter, C.; Chen, C.; Ge, W.; Morrice, N.; Liang, S.; Zhu, T.; Yuan, C.; Ruan, G.; Zhang, Q.; Cai, X.; Yu, X.; Chen, L.; Dai, S.; Luan, Z.; Aebersold, R.; Zhu, Y.; Guo, T. Accelerated Protein Biomarker Discovery from FFPE Tissue Samples Using Single-Shot, Short Gradient Microflow SWATH MS. *J. Proteome Res.* **2020**, *19* (7), 2732–2741. https://doi.org/10.1021/acs.jproteome.9b00671.

(79) Wiśniewski, J. R. Filter-Aided Sample Preparation for Proteome Analysis. In *Microbial Proteomics: Methods and Protocols*; Becher, D., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2018; pp 3–10. https://doi.org/10.1007/978-1-4939-8695-8_1.

(80) Carvalho, L. B.; Capelo-Martínez, J.-L.; Lodeiro, C.; Wiśniewski, J. R.; Santos, H. M. Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics. *Anal. Chem.* **2020**, *92* (13), 9164–9171. https://doi.org/10.1021/acs.analchem.0c01470.

(81) Wiśniewski, J. R.; Zettl, K.; Pilch, M.; Rysiewicz, B.; Sadok, I. "Shotgun" Proteomic Analyses without Alkylation of Cysteine. *Anal. Chim. Acta* **2020**, *1100*, 131–137. https://doi.org/10.1016/j.aca.2019.12.007.

(82) Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed

MaxLFQ. *Mol. Cell. Proteomics MCP* **2014**, *13* (9), 2513–2526. https://doi.org/10.1074/mcp.M113.031591.

(83) Brunet, M. A.; Lucier, J.-F.; Levesque, M.; Leblanc, S.; Jacques, J.-F.; Al-Saedi, H. R. H.; Guilloy, N.; Grenier, F.; Avino, M.; Fournier, I.; Salzet, M.; Ouangraoua, A.; Scott, M. S.; Boisvert, F.-M.; Roucou, X. OpenProt 2021: Deeper Functional Annotation of the Coding Potential of Eukaryotic Genomes. *Nucleic Acids Res.* **2021**, *49* (D1), D380–D388. https://doi.org/10.1093/nar/gkaa1036.

Attachment 1

This attachment contains the full text of the publications, which the presented thesis was based on, put in the order presented in the List of publications (page 7).

Journal of proteome research

Article

pubs.acs.org/jpr

Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC−MS and SWATH Methodology

Aleksandra E. Lewandowska,[*,†] Katarzyna Macur,[†] Paulina Czaplewska,[†] Joanna Liss,[‡] Krzysztof Łukaszuk,[‡,§] and Stanisław Ołdziej[*,†]

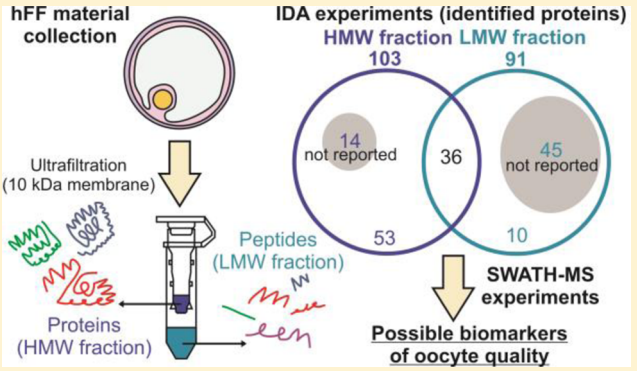
[†]Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Abrahama 58, 80-307 Gdańsk, Gdańsk, Poland

[‡]INVICTA Fertility and Reproductive Center, Trzy Lipy 3, 80-172 Gdańsk, Gdańsk, Poland

[§]Department of Obstetrics and Gynecological Nursing, Faculty of Health Sciences, Medical University of Gdańsk, Dębinki 7, 80-211 Gdańsk, Gdańsk, Poland

S *Supporting Information*

ABSTRACT: Human follicular fluid (hFF) is a natural environment of oocyte maturation, and some components of hFF could be used to judge oocyte capability for fertilization and further development. In our pilot small-scale study three samples from four donors (12 samples in total) were analyzed to determine which hFF proteins/peptides could be used to differentiate individual oocytes and which are patient-specific. Ultrafiltration was used to fractionate hFF to high-molecular-weight (HMW) proteome (>10 kDa) and low-molecular-weight (LMW) peptidome (<10 kDa) fractions. HMW and LMW compositions were analyzed using LC−MS in SWATH data acquisition and processing methodology. In total we were able to identify 158 proteins, from which 59 were never reported before as hFF components. 55 (45 not reported before) proteins were found by analyzing LMW fraction, 67 (14 not reported before) were found by analyzing HMW fraction, and 36 were identified in both fractions of hFF. We were able to perform quantitative analysis for 72 proteins from HMW fraction of hFF. We found that concentrations of 11 proteins varied substantially among hFF samples from single donors, and those proteins are promising targets to identify biomarkers useful in oocyte quality assessment.



KEYWORDS: *LC−MS, SWATH, human follicular fluid, proteome, peptidome, oocyte quality control*

INTRODUCTION

The oocyte in the follicle is immersed in follicular fluid (FF) during its growth and development in the ovary. This unique microenvironment surrounding the oocyte provides many important hormones (FSH, LH, GH, inhibin, activin, estrogens, and androgens), pro-apoptotic factors (TNF and Fas-ligands), proteins, peptides, amino acids, and nucleotides[1] but mostly is a filtrate of blood modified by substance uptake and secretion by granulosa and theca cells present in a follicle.[2] Because of the close relation to the maturing oocyte, it has been hypothesized that the composition of FF may reflect the quality of a given oocyte and its potential for embryonic development and a live birth following a successful pregnancy.[1] The significance of this assumption has been especially evident in the field of assisted reproductive technology (ART), especially the in vitro fertilization (IVF) technique. Currently, the oocyte selection in IVF procedures is mostly based on morphology, a relatively quick and simple but very subjective method.[3] Some other techniques based on, for example, gene expression analysis in granulosa cells or polar body biopsy are complicated, time-consuming, and require special laboratory equipment.[4,5] Without a reliable tool allowing the oocyte quality assessment prior to fertilization, most IVF procedures result in embryo overproduction that is an immediate reason for embryo storage difficulties, ethical issues, and legal problems in some countries.[1] Follicular fluid composition analysis could provide a unique opportunity to gain insight into the folliculogenesis process and the state of the oocyte. It can be easily performed during the IVF procedure, presenting the possibility of noninvasive means of the oocyte quality assessment. To analyze the content of FF, a quick, precise, and cheap methodology is needed. Follicular fluid proteomics has been

Received: June 1, 2017
Published: June 28, 2017

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

developing very intensively and is oriented toward noninvasive methods and identification of biomarkers of oocyte quality.[6,7]

A number of human follicular fluid (hFF) components have been studied in an attempt to determine the biomarkers of oocyte quality involving hormones, growth factors, antiapoptotic factors, proteins, peptides, amino acids, and sugars.[1] A great deal of attention has been given to the proteome of human follicular fluid. Combined proteomics research of hFF resulted in discovery of over 1100 distinct proteins so far.[6−27] As in human serum, follicular fluid contains an abundance of various proteins that are present at very different levels of concentrations. Therefore, fractionation is equally necessary, as in the case of more popular serum studies, where fractionation methods are emphasized in a number of recent proteomics studies[28−30] and in a review summarizing development of such methods in serum proteomics.[31] Several fractionation methods have been utilized in hFF proteome studies, which, coupled to mass spectrometry analysis, resulted in identification of a number of proteins varying from tens to over 750 in a single study.[6,8−16] Most successful approaches (in the sense of number of identified proteins) engage immunodepletion of most abundant proteins followed by HPLC fractionation (e.g., in alkaline pH[6] or by strong cation exchange chromatography (SCX)[8]).

A vast majority of proteomics studies on hFF relied only on the investigation of protein components of molecular weight 3 kDa and higher, leaving the lower-molecular-weight compounds originating from proteins (peptides) unexamined. Analysis of endogenous peptides is a very popular approach in studies of biological fluids, such as serum, plasma, cerebrospinal fluid, or urine.[32−34] This strategy allows for evaluation of dynamic protein breakdown or cleavage processes, which lead to the occurrence of peptides preserving a part of original protein's amino acid sequence. These peptides might constitute excellent biomarkers of, for example, a disease state or, in this case, the oocyte quality; however, so far hFF peptidome did not bring too much scientific attention, resulting in only one study.[7]

As described above, proteomics research on hFF is focused on the determination of biomarkers that could be used to describe oocyte quality. Up to now, results of studies that tried to correlate hFF proteome composition with the oocyte fate were rather inconclusive.[1,9,35] In those studies, many proteins were selected as possible biomarkers, but none of them has predicting power expected for medical diagnostic purposes. Among many studies on hFF proteome, in only one of them[9] did authors individually analyze follicular fluid samples drawn from the same donor. Bayasula and coworkers[9] performed a retrospective study using two samples per patient: one sample of hFF fluid from the oocyte that has led to healthy birth and another sample from an oocyte whose development stopped shortly after fertilization. Taking into account the size of follicles considered in this study as well as the number of considered follicles (from 6 up to 28 per patient), it should be assumed (this information is not provided in the described work) that samples were accumulated for an extended period of time. Extended time of samples' collection could be the reason why such promising studies as this performed by Bayasula and coworkers[9] may not provide proper candidates for biomarkers of oocyte quality. Patient (donor) aging is one of the major concerns in infertility treatment; therefore, extending studies over some period of time (as was done in the case of Bayasula et al.) could lead to change oocyte characteristics. In a typical IVF procedure more than one oocyte is harvested; in this work, we will focus on identification of differences in hFF composition where samples come from follicles of the same donor and are taken at the same time. Using samples coming from the same donor and harvested at the same time has the advantage that it eliminates differences between samples that are associated with patient age, health condition, race, life style, and so on; therefore, identification of differences between oocytes should be more reliable and easier to observe. The work was designed as a small-scale pilot study with only 12 samples (four donors with three samples each) and is focused on selection of proteins/peptides whose concentrations vary between hFF samples from the same donor. Some studies on hFF proteome composition using mass spectrometry were performed on only one to three samples, and each sample was a pool from more than one donor.[6,12,15] On the contrary, studies on composition of hFF from individual follicles involve from 3[6] to 58[35] patients and as many samples. Taking into account MS proteomics studies related to hFF composition published so far, our studies are relatively small (number of donors)/mediocre (number of samples). However, using such small number of samples allows us to conduct more detailed studies, and, as a result, this could allow us to narrow down a list of possible protein/peptide candidates for biomarkers of oocyte quality. Statistical significance of obtained results could be verified in additional studies on a larger number of samples but focused on a shorter list of molecular targets, which we plan to do in the future. Such approach of small steps seem to be more sound from s methodological (development of targeted analytical methodology on a limited number of molecules) and economical point of view (lower quantity of more focused measurements).

We investigated the proteome, as well as the peptidome, of human follicular fluid to obtain a more complete image of its composition. We used ultrafiltration to separate the two analyzed fractions as a prefractionation scheme (see Figure 1).To obtain quantitative results for both protein and peptide fractions, we employed a micro liquid chromatography−tandem mass spectrometry (microLC−MS/MS) technique in a novel data acquisition mode abbreviated as SWATH (Sequential Windowed Acquisition of All Theoretical Fragment Ion Spectra).[36] The aim of this work is to perform preliminary investigation of hFF proteome composition using sets of samples of hFF coming from a single donor to identify potential biomarkers describing oocyte quality. Moreover, we simultaneously analyzed compositions of proteome as well as peptidome of hFF to widen the range of possible biomarkers. To our knowledge, our work is the first one that involves simultaneous analysis of peptidome and proteome of hFF performed on several samples from the same donor.

MATERIALS AND METHODS

Research Approval

Proteomics analysis described herein is part of the "Identification of biomarkers of early embryonic development and pregnancy" project that has been approved by the Independent Bioethics Commission at the Medical University of Gdansk (decision 62/2016). Each couple undergoing IVF treatment has signed a written informed consent regarding the treatment and all included procedures. The obtained written consents also include agreement for publication of treatment-related data as long as patient anonymity is maintained.

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

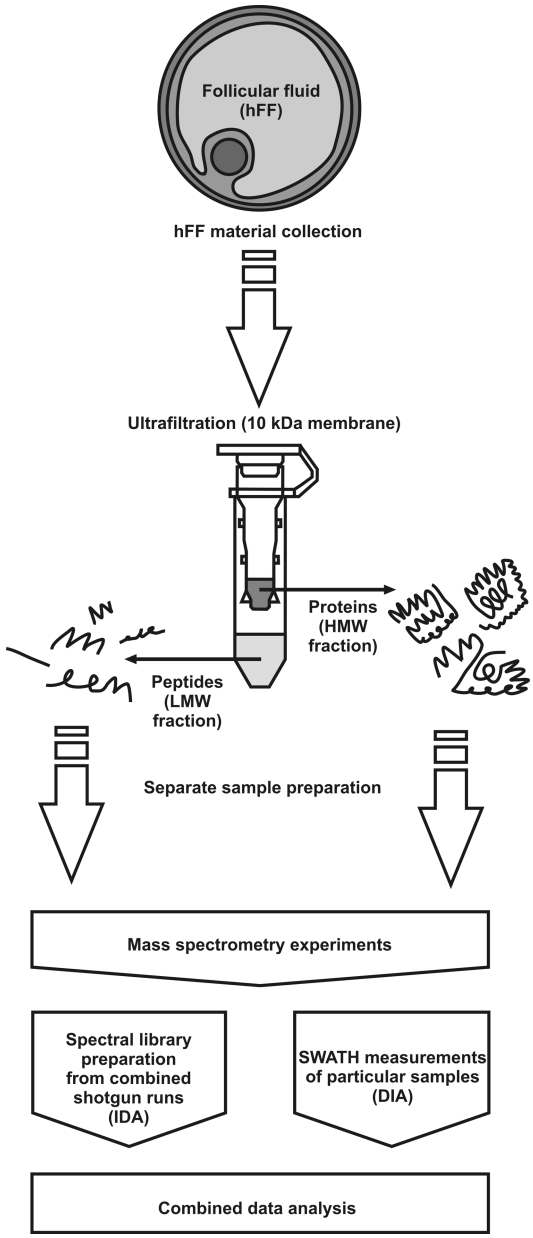


Figure 1. Overall workflow of the study. Collected separate hFF samples were subjected to ultrafiltration to divide the material into two fractions: high-molecular-weight fraction (HMW, >10 kDa) containing whole proteins and low-molecular-weight fraction (LMW, <10 kDa) containing endogenous peptides (fragments of proteins). Both fractions were separately prepared and analyzed by mass spectrometry. MS analysis was divided into IDA (Information Dependent Acquisition) measurements for spectral library construction and SWATH measurements of individual samples. The results of both types of measurements were later combined in data analysis step to obtain final quantitative results.

Human Follicular Fluid Samples Collection

Four women undergoing IVF procedure due to male factor infertility took part in this study. All patients underwent hormonal stimulation as part of IVF treatment.

Women participating in the study underwent ovarian stimulation according to a long agonist protocol starting from oral contraceptive (OC) pills (Ovulastan, Adamed, Czosnow, Poland) from second to fifth day of the cycle. Fourteen days after the beginning of OC, Triptorelin acetate 0.1 mg (Gonapeptyl, Ferring, Saint-Prex, Switzerland) was administered. Fourteen days later (i.e., 7 days after the end of OC), administration of urinary gonadotropins (Menopur, Ferring, Saint-Prex, Switzerland) for ovarian stimulation was begun, and their dosage was based on patients antimüllerian hormone (AMH) level (from 150 to 225 IU daily).[37] Follicular growth was monitored on day 8 (and later if necessary) using transvaginal ultrasound and assays evaluating serum oestradiol (E2) and progesterone (P) levels. Oocyte pick-up was performed 36 h after the trigger administration of hCG (Choragon, Ferring, Kiel, Germany).

Each follicular fluid with cumulus complex was retrieved separately with needle flushing between each follicle puncture. Follicular fluid volume as well as a presence of oocyte was verified. Quality and maturity of the oocyte was determined and recorded, and the oocyte was moved to a drop of culture media.[38]

Follicular fluid was centrifuged; then, supernatant was stored in three tubes and granulosa cells in one additional tube. Only follicular fluids from follicles containing oocytes of visible good quality were considered in this study. Our pilot study covered four patients, and three hFF samples from individual follicles of each patient were obtained, resulting in 12 samples in total. Samples are further referred to using abbreviations: P1−P4, patient number 1−4; F1−F3, follicle number 1−3. In addition to samples coming from individual oocytes, we used a pool sample coming from several patients not involved in individual analyzes as a reference material. All samples qualified for the study were free of visible blood contamination. Stored material was frozen at −20 °C until analysis.

Sample Preparation

Each separate follicular fluid sample was centrifuged at 1000*g* for 10 min, and 400 μL of each supernatant was separated from the cell pellets. 100 μL of acetonitrile was added to each sample and briefly vortexed to break protein−protein interactions.[39,40] Then, each sample was subjected to ultrafiltration by centrifugation on the 10 kDa Amicon membrane (Merck-Millipore). Both filtrates (LMWF, low-molecular-weight fraction) and retentates (HMWF, high-molecular-weight fraction) were collected for further analysis.

Low-Molecular-Weight Fraction

Prior to desalting, filtrates were evaporated under reduced pressure to dryness in SpeedVac. Peptides were then dissolved in 50 μL of 0.5% trifluoroacetic acid in water and filtered on 0.22 μm membranes (Agilent Technologies) by centrifugation for 1 min at 16 000*g*. Samples were later desalted on C18 Pierce Spin Tips (Thermo Fisher Scientific) according to the manufacturer protocol with the following exception: The whole procedure was repeated three times with increasing concentration of acetonitrile in the elution step (30, 50, and 80% acetonitrile in water with the addition of 0.1% formic acid). Resulting eluates were combined and subjected to LC−MS/MS analysis.

High-Molecular-Weight Fraction

Fractionation. Clinical samples from individual follicles from patients taking part in the study were not subjected to any fractionation procedures. We employed fractionation only in the pool samples of hFF to obtain more protein identifications in the spectral library further used for SWATH data analysis. Several of the pool samples were depleted of 14 most abundant human serum proteins (albumin, IgG, antitrypsin, IgA,

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) prior to initial preparation step using a commercially available Multiple Affinity Removal Spin Cartridge Human 14 (MARS-14) kit (Agilent Technologies, Santa Clara, CA) according to manufacturer's protocol. Samples were then subjected to ultrafiltration, as described in the Sample Preparation section and processed further accordingly as other individual clinical HMWF samples used for quantification.

Protein Digestion and Desalting. Protein digestion was performed according to a protocol proposed by Gundry et al.[41] Samples of concentrated protein residue (HMWF) were diluted to approximate protein concentration of 1 mg/mL in NH_4HCO_3. Diluted protein solutions were subjected to protein digestion. First, proteins were reduced by dithiotreitol (DTT) in final concentration in solution of 10 mM for 30 min in 56 °C. Then, proteins were alkylated by iodoacetamide in the final concentration in solution of 20 mM for 30 min in room temperature in darkness. Afterward, trypsin solution in 50 mM acetic acid was added in enzyme−substrate ratio 1:50, and the samples were incubated in 37 °C for 19 h. Finally, digestion was stopped by the addition of 5% trifluoroacetic acid in 50% acetonitrile/water by decreasing solution pH to 3. Solutions were then evaporated to dryness in SpeedVac. Peptides were dissolved in 50 μL of 0.5% trifluoroacetic acid in water and desalted on C18 Pierce Spin Tips, as described in the Low-Molecular-Weight Fraction section and later subjected to LC−MS/MS measurements.

LC−MS/MS Measurements

Chromatography. LC separation was performed on the Eksigent microLC (Ekspert MicroLC 200 Plus System, Eksigent, Redwood City, CA) using for the digested HMWF the ChromXP C18CL column (3 μm, 120 Å, 150 × 0.5 mm) and for the nondigested LMWF the ChromXP C8CL column (3 μm, 120 Å, 150 × 0.3 mm). Samples were loaded onto the column using the CTC Pal Autosampler (CTC Analytics AG, Zwinger, Switzerland), with each injection of a 5 μL volume. The solvents A and B composed of 0.1% (v/v) formic acid in water and acetonitrile, respectively. Loaded material was separated on the column using a gradient program that lasted 30 min, and it was divided into following parts: (i) 0−2 min − 10% solvent B, (ii) 2−23 min − 10−90% solvent B, (iii) 23−28 min − 90% solvent B, and (iv) 28.1−30 min − 10% solvent B.

Mass Spectrometry. Eluate from the column was analyzed in a positive ion mode on a TripleTOF 5600$^+$ hybrid mass spectrometer with DuoSpray Ion Source (AB SCIEX, Framingham, MA). The microLC−MS/MS system was controlled by the AB SCIEX Analyst TF 1.6 software.

Shotgun Mass Spectrometry Experiments. The data-dependent experiments (IDA) were conducted for all of the investigated samples. The TOF MS survey scan was performed in the m/z range of 100−2000 with the accumulation time of 50 ms. The top 10 precursor ions, with the charge states from +2 to +5, were then selected for collision-induced dissociation (CID) fragmentation. Product ion spectra were collected in the m/z range of 100−2000 with the accumulation time of 40 ms. This resulted in duty cycle time of 1.11 s. All IDA runs were executed with rolling collision energy, and the precursor ions were excluded from reselection for 5 s after two occurrences.

SWATH Mass Spectrometry Experiments. SWATH-MS analyses, according to a method developed by Ruedi Aebersold's group,[36] of the investigated HMWF follicular fluid samples were performed in a looped product ion mode. A set of overlapping 25 transmission windows, each 25 Da wide, was constructed and covered the precursor mass range of 400−1000 m/z. The SWATH product ion scans were acquired in the range of 100−2000 m/z. The collision energy for each window was calculated for a +2- to +5-charged ions centered upon the window with a spread of 2. The SWATH-MS1 survey scan was acquired in high sensitivity mode in the range of 100−2000 Da in the beginning of each cycle with the accumulation time of 50 ms, and it was followed by 40 ms accumulation time high-sensitivity product ion scans, which resulted in the total cycle time of 1.11 s.

Data Analysis

Protein database search was conducted in the ProteinPilot 4.5 Software (AB SCIEX) using the Paragon algorithm. The search of combined HMWF IDA runs was performed against SwissProt *Homo sapiens* database (ver. 05.09.2016, 20 200 entries), with an automated false discovery rate and the following parameters: instrument TripleTOF 5600, alkylation of cysteines by iodoacetamide, and trypsin enzyme digestion; ID focus on amino acid substitutions and biological modifications; search effort "thorough ID"; and detected protein threshold [unused protein score (Conf)]>0.05 (10%). All identified hits from database search were manually inspected for presence of amino acids substitutions or modifications to lower the chances of possible false protein identification. Runs of LMWF samples were processed accordingly with changes in parameters: no modification of cysteines and no enzyme specified in digestion. Only the protein identifications with 1% FDR in HMWF and peptide identifications with 1% FDR in LMWF were considered valid.

Quantitative analysis was performed by loading the result from the ProteinPilot database search file to the MS/MS ALL with SWATH Acquisition MicroApp 2.01 in PeakView 2.2 (SCIEX) to automatically create a spectral library file. The assumptions employed during the spectral library creation included: a maximum of 10 peptides per protein and 10 transitions per peptide; peptide modification was allowed (but shared peptides were excluded); only peptides of at least 95% confidence and an extraction window of 5 min and 0.02 Da width were used. The peptide profiles and transitions were manually inspected to ensure that PeakView extraction was accurate and that ion transitions correlated with each other. Proteins have been quantified using at least two peptides with a 95% identification certainty. The created spectral libraries were used for targeted data extraction from SWATH-MS experiments and further processing in MS/MS ALL with SWATH Acquisition MicroApp 2.01 in PeakView 2.2 software. Statistical analysis of processed data was performed in MarkerView 1.2.1 Software (SCIEX). For each hFF sample three biological and three technical replicates were analyzed and the areas under XICs for technical replicates were averaged. Quantitative data were normalized using Total Area Sums (TAS) approach. The coefficient of variation (CV) was calculated using SWATH data from parallel measurements of four pool reference samples to monitor measurement variations for each protein (in HMWF) and peptide (in LMWF). The t tests were performed twice between groups composed of all samples for each individual oocyte and groups composed of all samples from each patient, and the peptides/proteins with $p < 0.05$ were considered as statistically significant. Gene ontology analysis was carried out

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Table 1. List of Proteins Previously Not Reported in hFF Literature with Indication of the Presence in Human Plasma Proteome Database

N	% confidence	Uniprot ID	name	peptides (95% confidence)	plasma proteome database ID
			High-Molecular-Weight Fraction		
1	100	P06309	Ig kappa chain V−II region GM607 (Fragment)	5	
2	100	P01772	Ig heavy chain V−III region KOL	3	
3	100	P06314	Ig kappa chain V−IV region B17	8	
4	100	P18887	DNA repair protein XRCC1	2	HPRD_01909
5	99	Q5T9C2	Protein FAM102A	2	HPRD_13262
6	99	P16499	Rod cGMP-specific 3′,5′-cyclic phosphodiesterase subunit alpha[a]	1	HPRD_01570
7	99	O00522	Krev interaction trapped protein 1	3	HPRD_05020
8	99	P15538	Cytochrome P450 11B1, mitochondrial[a]	1	HPRD_01943
9	99	P01743	Ig heavy chain V−I region HG3	2	
10	99	Q7Z5N4	Protein sidekick-1[a]	1	
11	99	Q7RTS5	Otopetrin-3[a]	1	
12	99	Q14DG7	Transmembrane protein 132B[a]	1	
13	99	Q5SY13	Putative uncharacterized protein encoded by COL5A1-AS1	2	
14	99	P04208	Ig lambda chain V−I region WAH	2	
			Low-Molecular-Weight Fraction		
1	100	Q7Z5P9	Mucin-19	2	
2	100	Q9NXV6	CDKN2A-interacting protein	1	
3	100	Q6ZU65	Ubinuclein-2	1	
4	100	Q8NFD5	AT-rich interactive domain-containing protein 1B	1	HPRD_10660
5	99	Q9UPA5	Protein bassoon	1	HPRD_04933
6	99	P31942	Heterogeneous nuclear ribonucleoprotein H3	1	HPRD_03818
7	99	O43448	Voltage-gated potassium channel subunit beta-3	1	HPRD_04983
8	99	Q07954	Prolow-density lipoprotein receptor-related protein 1	1	HPRD_00138
9	99	P43146	Netrin receptor DCC	1	HPRD_00391
10	99	Q8ND83	SLAIN motif-containing protein 1	1	
11	99	A6NGC4	TLC domain-containing protein 2	1	
12	99	Q69YQ0	Cytospin-A	1	HPRD_11080
13	99	Q08378	Golgin subfamily A member 3	1	HPRD_03990
14	99	P50406	5-hydroxytryptamine receptor 6	1	HPRD_03066
15	99	Q14210	Lymphocyte antigen 6D[b]	1	HPRD_09370
16	99	Q96PX6	Coiled-coil domain-containing protein 85A	1	
17	99	O60732	Melanoma-associated antigen C1[b]	1	HPRD_02201
18	99	P23921	Ribonucleoside-diphosphate reductase large subunit	1	HPRD_01588
19	99	Q6ZUJ8	Phosphoinositide 3-kinase adapter protein 1	1	
20	99	Q06710	Paired box protein Pax-8	1	HPRD_01335
21	99	Q9HBL0	Tensin-1	1	HPRD_02512
22	99	Q69YH5	Cell division cycle-associated protein 2	1	
23	99	P23769	Endothelial transcription factor GATA-2	1	HPRD_00673
24	99	Q9Y2P0	Zinc finger protein 835	1	
25	99	Q13671	Ras and Rab interactor 1	1	HPRD_05813
26	99	Q99808	Equilibrative nucleoside transporter 1	1	HPRD_03724
27	99	Q8N693	Homeobox protein ESX1	1	HPRD_06503
28	99	P00966	Argininosuccinate synthase	1	HPRD_04590
29	99	Q02878	60S ribosomal protein L6	1	HPRD_04745
30	99	Q86YV0	RAS protein activator like-3	1	
31	99	Q9H201	Epsin-3	1	HPRD_06272
32	99	Q96FN5	Kinesin-like protein KIF12	1	HPRD_13914
33	99	Q9NYY3	Serine/threonine-protein kinase PLK2	1	HPRD_06118
34	99	Q9H7T3	Uncharacterized protein C10orf95	1	
35	99	Q96PJ5	Fc receptor-like protein 4	1	
36	99	Q8TE59	A disintegrin and metalloproteinase with thrombospondin motifs 19	1	HPRD_06334
37	99	P09496	Clathrin light chain A	1	HPRD_00351
38	99	Q9Y3D2	Methionine-R-sulfoxide reductase B2, mitochondrial	1	HPRD_17606
39	99	Q9NUB1	Acetyl-coenzyme A synthetase 2-like, mitochondrial	1	HPRD_12413
40	99	O60315	Zinc finger E-box-binding homeobox 2	1	HPRD_05780
41	99	Q8WW01	tRNA-splicing endonuclease subunit Sen15	1	HPRD_12292
42	99	P57053	Histone H2B type F−S	1	
43	99	Q6ZMS7	Protein ZNF783	1	

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Table 1. continued

N	% confidence	Uniprot ID	name	peptides (95% confidence)	plasma proteome database ID
			Low-Molecular-Weight Fraction		
44	99	Q8N976	Putative uncharacterized protein FLJ38264	2	
45	98	Q93084	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	1	HPRD_03568

[a]Proteins identified basing on a single peptide evidence resulting from enzymatic digestion of HMWF (see Supporting Information Part 1: Figures S2, S4, and S7−S9). [b]Proteins identified basing on one endogenous peptide, recognized with a single amino acid substitution (see Supporting Information Part 1: Figures S29 and S32).

using the PANTHER database.[42,43] The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[44] via the PRIDE[45] partner repository with the data set identifier PXD006550.

RESULTS AND DISCUSSION

We divided the human follicular fluid material into two fractions: proteomic and peptidomic, which were later processed and analyzed separately (see Figure 1. and Material and Methods section). Therefore, the results from both fractions are described individually in the following paragraphs divided into qualitative and quantitative parts of the experiment.

Qualitative Analysis

High-Molecular-Weight Fraction. In our study we attempted proteomics analysis using a microHPLC coupled to a hybrid triple quadrupole TOF spectrometer. This system has already been employed in proteomics research of other biological materials, for example, cerebrospinal fluid, tear proteins, or *Saccharomyces cerevisiae* proteome.[46−48] In our study we obtained 103 protein identifications at 99% or higher level of confidence from a single combined database search of all analyzed HMWF samples. The full list of identified proteins is presented in the Supporting Information Part 2. Using a very similar protocol (immunodepletion followed by 50 kDa membrane ultrafiltration fractionation with measurements on a nano-HPLC coupled to quadrupole TOF spectrometer), Kushnir et al.[11] reported the identification of 75 proteins in hFF samples. Similarly, in other studies that used simple one-step separation protocols (capillary electrophoresis, electrophoresis) the number of identified proteins in hFF samples usually did not exceed 100.[12,13] It could be concluded that the number of proteins identified in this work is similar to numbers obtained in various studies using simple one-step separation protocols.[11−13] More advanced and complicated multistep separation procedures, for example, combination of immunodepletion and 2D-PAGE/HPLC/SCX separation followed by analysis by nano-LC coupled to LTQ Orbitrap, seem to result in most single-study protein discovery rate, as reported by Ambekar et al. or Zamah et al. (482, 742, and 770 protein identifications).[6,8,15] It is evident that fractionation plays the most important role in the number of distinct protein identifications in the investigated material. The number of proteins identified in our study is comparable to other studies that used simple separation protocols,[11−13] but those numbers are rather small in comparison with the results of more complicated separation schemes.[6,8,15] Another factor that has an impact on the number of proteins reported is statistical significance of presented results. For example, Zamah et al.[6] presented results that were obtained using a 5% FDR threshold in database search; in this case, a more restricted FDR threshold value (for example, 1% as implemented in this study) would reduce the number of hits. In the same study authors reported proteins even when ProteinPilot confidence score value was very low (far below 2.0, which is the value used in this study). From data provided by authors we were not able (lack of specified data) to assess the influence of changing of FDR threshold on number of identified proteins, but using a higher ProteinPilot confidence cutoff (>2.0) eliminates about 50 proteins from 742 reported.[6] Another issue is the number of peptides that were used for protein identification, for example, in the work of Ambekar et al.[15] 204 proteins out of 482 (>42%) were identified based on only one peptide found in the analysis. For comparison, in this study we report 9 such proteins out of 103 identified ((<9%), see Supporting Information Part 1).

In summary, the number of proteins identified in this study is comparable to that of other studies that used a simple one-step separation procedure[11−13] but does not seem to be very impressive as compared with studies with more sophisticated fractionation protocols.[6,8,15] However, it should be kept in mind that not only fractionation protocols but also data processing and data interpretation affects the number of identified proteins and does not allow for a simple comparison of the numbers of identified proteins obtained in different laboratories.

To enable determination of proteins newly reported in our experiment, we conducted a thorough literature search of publications reporting protein discovery and analysis in hFF from January 2000 to September 2016 by searching for the term "human follicular fluid" against the PubMed database. While the number of identifications in extensively studied HMWF is not remarkably high in the field of hFF proteome research, we managed to find 14 proteins in HMWF not reported in previous publications related to research of the hFF proteome (Table 1), for example, DNA repair protein XRCC1 (P18887) or Ig kappa chain V−IV region B17 (P06314). Five of those newly reported for the hFF proteome proteins were previously identified in plasma (see Table 1). The remaining nine proteins mostly belong to immunoglobulin chain type group. Three proteins identified in this study among the already reported as present in hFF have been previously reported in only a single work. Literature search as well as results of our work clearly indicate that full characterization of hFF proteome is far from completion. The application of various techniques for protein detection as well as various fractionation schemes leads to finding previously undescribed proteins. We assume that currently there is no workflow allowing a chance for the discovery of all follicular proteins at once, which favors employing different strategies.

Low-Molecular-Weight Fraction. Peptidomics is a new approach to research of biological material, emerging from proteomics.[32−34,49] It involves analysis of endogenous peptides, which result from protein cleavage or breakdown, in contrast with protein digests prepared during proteomic sample preparation. Such peptides may display biological functions,

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

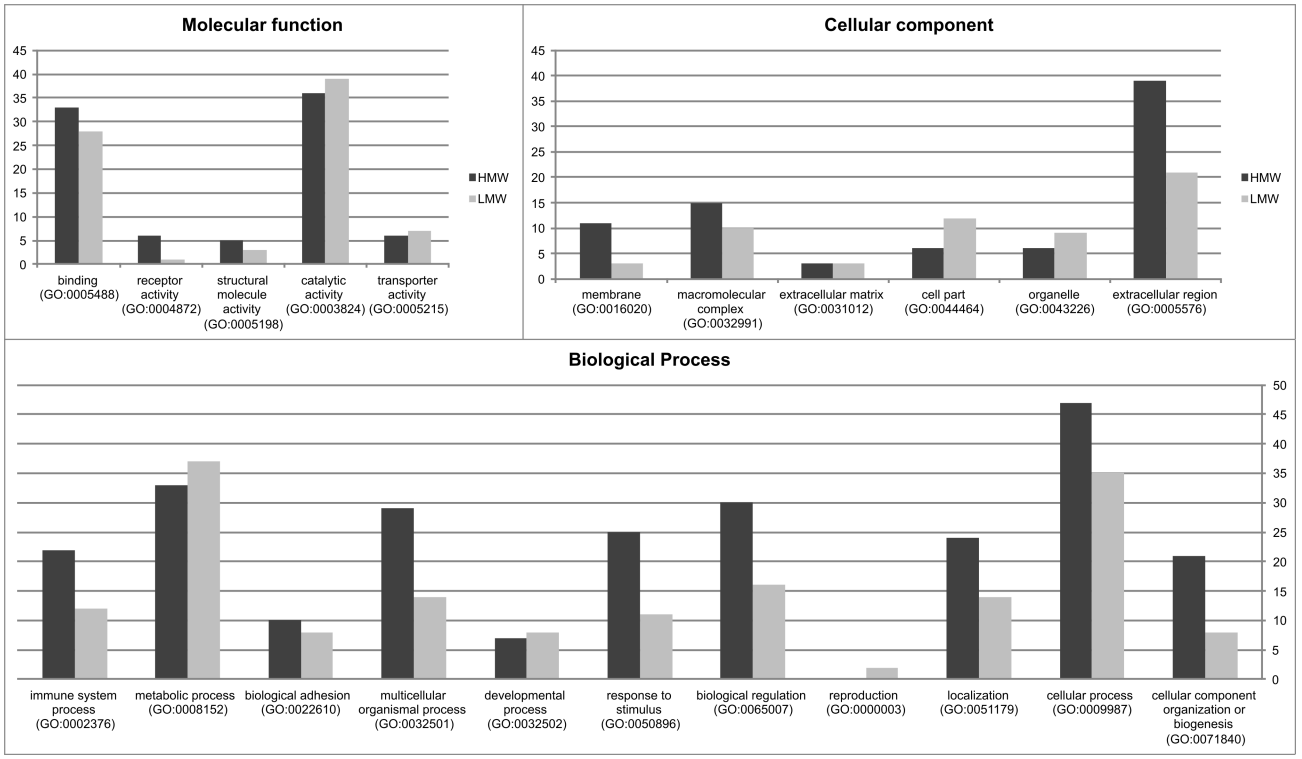


Figure 2. Gene ontology classification of proteins in separate fractions according to molecular function, cellular component, and biological process.

or their presence may indicate a particular state of the organism (e.g., disease).[33,50] Peptidomic analyses have been widely propagated in human biological fluids, such as serum or plasma, urine, saliva, or cerebrospinal fluid, giving insight into peptide and protein composition of these materials and mining potential disease biomarkers.[33,50] In the case of hFF, so far only a recently published work of Chen et al. concerns the peptidomic approach in human follicular fluid research field.[7] His group studied LMWF of hFF resulting from ultrafiltration (3 kDa membrane) on a nano-LC coupled to hybrid quadrupole orbitrap mass spectrometer. In their experiment 159 proteins were identified from endogenous peptides. In our study we used the microLC system coupled to a TripleTOF 5600+ hybrid mass spectrometer. We also conducted different sample preparation, that is, peptide extraction with the use of acetonitrile, and in contrast with Chen et al. we did not condense desalted filtrate fractions by lyophilization and resuspension in a small solvent volume. Using an approach relatively similar to that of Chen and coworkers (see Material and Methods section) we were able to identify with high confidence 91 proteins in the LMWF of human follicular fluid (Supporting Information Part 2), from which 27 proteins have also been identified by Chen et al.[7] Again, differences in results may be a consequence of the application of divergent sample preparation techniques, LC−MS/MS setups, variances in database search setup, statistical data processing, or even the number of processed samples. In particular, the differences in final sample volumes analyzed by LC−MS/MS, that is, emerging from application of lyophilization and subsequent resuspension in a small solvent volume, could be crucial to measured ranges of concentration and, as a result, to number of identified proteins. In our study we were able to identify 45 proteins in LMWF that were never reported before in hFF proteome research (see Table 1). Thirty of these unreported proteins have already been identified in human plasma, which may strengthen our discovery, as a broad part of human plasma proteome is able to cross the blood−follicle barrier, causing both proteomes to be comparable.[1,6] Results presented by Chen and coworkers and results from our study clearly indicate that analysis of LMWF could give valuable information about protein content of hFF.

Functional Analysis. We conducted a general functional analysis of the proteins identified in our study as described in the Material and Methods section. As expected, the majority of the proteins in both fractions localize in the extracellular region (see Figure 2). However, some of the proteins were also found in cellular components, cell part or organelles: 6 proteins in HMWF and 12 proteins in LMWF. The presence of these proteins in hFF material may have two distinct explanations: (1) It is an artifact of sample retrieval and preparation procedures or (2) it results from natural catabolic processes or cell breakdown. During the oocyte retrieval procedure, a small amount of blood may be aspirated along with hFF. However, the amount may be so small that only a few proteins from disrupted blood cells are present in the sample. Cells present inside the follicle may also be fractured during this procedure. Moreover, out of six HMWF proteins present in cell components, four are types of keratin, whose presence can be easily introduced into the sample during its processing or preparation. The other explanation of the obtained results is based on processes occurring naturally in the follicle. In the course of folliculogenesis, some granulosa or theca cells may undergo apoptosis, which could result in protein remains in hFF. Regardless of their genesis, such proteins should not be directly taken into consideration as reliable biomarkers because of the impossibility to exclude the sample processing factor. In our opinion, comparison of this fraction proteins from differently developed oocytes, not only mature but also

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

immature ones, and in particular oocytes that underwent atresia could confirm or exclude the idea of apoptosis contribution in this phenomena. The number of such proteins is higher in the LMWF. It is possible that some peptides may be transported outside the cell or simply that proteins after cell breakdown may be subjected to unspecific cleavage and remain in hFF as short peptides. However, in some cases such transport seems highly improbable, especially in the case of nuclear proteins, that is, histone H2B type F−S. Proteins localized in the cell mostly have structural molecule activity and take part in diverse biological processes.

In general, HMWF and LMWF proteins mainly exhibit binding and catalytic activity (see Figure 2). Dominant biological processes for identified proteins are cellular-, metabolic-, and regulation-related. Enzyme modulator protein class was dominantly assigned for proteins in both fractions (see Figure S-69 in Supporting Information Part 1). As expected, because of the similarities between proteomic composition of blood serum and hFF, the most important pathways assigned for identified proteins were blood coagulation and plasminogen-activating cascade.

Summary of Qualitative Analysis. In separate fraction analysis, we managed to identify 103 proteins in HMWF and 91 proteins in LMWF. In all, we reported 158 distinct protein identifications, as 36 of them have been present in both fractions (see Figure 3). The number of proteins found in a previously extensively analyzed HMWF is not significant in comparison with other complex studies.[6,8] The reason for that is primarily a choice of particular fractionation scheme. However, despite a relatively low number of protein identifications, we found 59 proteins that have not been previously reported in hFF research (Table 1). Among those there were 14 proteins identified in HMWF and 45 proteins identified in LMWF (see Figure 3). In essence, about half of protein identifications in LMWF were not reported previously in hFF proteome research, whereas only a small part of HMWF contained such nonreported proteins. Therefore, it is clear that our preliminary fractionation to LMWFs and HMWFs by ultrafiltration is a promising approach to comprehensive analysis of hFF proteome. Additionally, such peptidomic workflow provides information on natural protein cleavage or breakdown, allowing insight into processes occurring in human follicular fluid. The fact that we managed to identify previously not reported hFF proteins also in HMWF proves that currently there is no single technique for identification of whole hFF proteome. Different sample preparation workflows and types of LC−MS/MS measurement result in identification of distinct sets of hFF proteins, allowing a chance of discovery of new hFF proteins even by using workflows with potentially less resolving power.

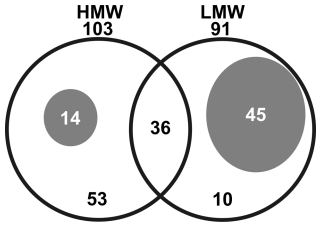


Figure 3. Venn diagram displaying distribution of proteins identified in both fractions of hFF. Gray ellipsoids in each fraction depict fractions of proteins never previously reported in previous hFF literature.

Quantitative Analysis

High-Molecular-Weight Fraction. Some groups reporting protein identifications in hFF also performed quantitative analysis of discovered proteins by mass spectrometric methods.[6,8,9,11,16,17] In our research we used the SWATH-MS technique (Sequential Windowed Acquisition of All Theoretical Fragment Ion Spectra). This method has never been previously used in hFF proteome research. Nevertheless, its popularity in proteomics has been growing recently. SWATH-MS has already been used in analysis of various human body fluids, for example, plasma, urine, or cervicovaginal fluid.[51−53] We quantified 72 of 103 identified HMWF proteins (data in Supporting Information Part 3). There were at least two tryptic peptides identified at 95% or more confidence for each quantified protein. Moreover, we only chose proteins with stable narrow XICs for quantitative data processing. 48 quantified proteins differed significantly between 12 measured follicles, and 17 proteins differed significantly among 4 patients (at least 2-fold change) (see Supporting Information Part 3). This relation illustrates a potential diversity in proteomic composition of individual follicles, whether they come from the same patient or from different patients. Table 2 contains summarized data of 17 proteins at highly varying levels among individual follicles of a single patient and 17 proteins at highly varying levels between patients (at least 2-fold change).

To have a closer look into their potential biomarker capability, we compared the results of our study with the studies of other groups considering identification or quantification of those proteins. We divided the proteins into eight groups: (I) proteins present at different concentrations among oocytes and reported as important in other studies; (II) proteins present at different concentrations among oocytes and identified in other studies; (III) proteins present at different concentrations among patients and reported as important in other studies; (IV) proteins present at different concentrations among patients and identified in other studies; (V) proteins present at different concentrations among patients and unreported in other studies; (VI) proteins identified in LMWF and reported as important in other studies; (VII) proteins identified in LMWF also identified in other studies; and (VIII) proteins not reported in other studies. Summarized results of our research are shown in Table 3. The most important finding is the identification of several proteins/peptides whose concentrations vary substantially among oocytes coming from the same patient (see Table 3: Groups I and II). Moreover, most of those proteins in previous proteomics studies are highlighted as promising biomarkers to assess oocytes' quality (Table 3: Group I). Proteins listed in Group I, but also proteins listed in the Group II should be considered as primary targets in searching biomarkers describing oocytes' quality. Taking into account the limitations of our study (a small number of patients participating in the study), it is necessary to conduct more extensive studies that will allow eventual correlation of concentration changes of proteins from Groups I and II with further oocyte fate; however, our results at least allow us to narrow down a list of potential proteins that could be promising biomarkers of oocyte quality. Taking into account the simplified fractionation scheme developed in this work, our procedure of protein quantification will allow for fast measurements of many samples at low cost in

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Table 2. List of Proteins Detected at Statistically Significant Different Concentrations by *t*-Tests in HMWF among the Samples of (1) Individual Follicles of a Single Patient and (2) Patients (at least 2-Fold Change)

index	peak name	group	*p* value	fold change	group 1	group 2
		Proteins Significantly Different between Separate Follicles of Individual Patients				
1	P43652	Afamin	0.03953	0.35	P3 F1	P3 F2
1	P43652	Afamin	0.02602	0.36	P3 F1	P3 F3
2	P08697	Alpha-2-antiplasmin	0.02334	0.32	P3 F1	P3 F3
3	P01023	Alpha-2-macroglobulin	0.04073	2.05	P1 F1	P1 F2
4	P01008	Antithrombin-III	0.0282	0.39	P3 F1	P3 F3
5	P06727	Apolipoprotein A-IV	0.01365	0.22	P3 F1	P3 F2
5	P06727	Apolipoprotein A-IV	0.03744	0.13	P3 F1	P3 F3
6	P05090	Apolipoprotein D	0.00755	0.12	P1 F1	P1 F3
7	P00751	Complement factor B	0.04061	0.37	P3 F1	P3 F3
8	P05156	Complement factor I	0.0371	2.13	P2 F2	P2 F3
9	P68871	Hemoglobin subunit beta	0.00109	0.46	P1 F1	P1 F3
10	P05546	Heparin cofactor 2	0.00477	0.37	P3 F1	P3 F2
11	P04196	Histidine-rich glycoprotein	1.77×10^{-6}	0.46	P4 F1	P4 F3
12	P01876	Ig alpha-1 chain C region	0.00445	2.37	P3 F1	P3 F2
12	P01876	Ig alpha-1 chain C region	0.0098	2.42	P3 F1	P3 F3
13	P35858	Insulin-like growth factor-binding protein complex acid labile subunit	0.04533	0.24	P3 F1	P3 F3
14	Q14624	Interalpha-trypsin inhibitor heavy chain H4	0.00682	0.45	P3 F1	P3 F2
14	Q14624	Interalpha-trypsin inhibitor heavy chain H4	0.04015	0.29	P3 F1	P3 F3
15	P35527	Keratin, type I cytoskeletal 9	0.04977	0.43	P3 F1	P3 F2
15	P35527	Keratin, type I cytoskeletal 9	0.04215	0.38	P3 F1	P3 F3
16	P04264	Keratin, type II cytoskeletal 1	0.03025	2.52	P4 F1	P4 F3
17	P02750	Leucine-rich alpha-2-glycoprotein	0.00238	0.28	P3 F1	P3 F2
17	P02750	Leucine-rich alpha-2-glycoprotein	0.00584	3.85	P3 F2	P3 F3
		Proteins Significantly Different between Patients				
1	P06727	Apolipoprotein A-IV	0.03848	0.42	P2	P3
1	P06727	Apolipoprotein A-IV	0.01128	3.35	P3	P4
2	P01031	Complement C5	0.04277	0.44	P1	P3
3	P02671	Fibrinogen alpha chain	3.28×10^{-14}	2.36	P2	P4
4	P00738	Haptoglobin	7.53×10^{-28}	0.37	P1	P2
4	P00738	Haptoglobin	4.42×10^{-30}	3.30	P2	P3
4	P00738	Haptoglobin	1.37×10^{-33}	4.06	P2	P4
5	P68871	Hemoglobin subunit beta	1.52×10^{-2}	4.66	P2	P4
6	P04196	Histidine-rich glycoprotein	4.89×10^{-8}	0.44	P3	P4
7	P01876	Ig alpha-1 chain C region	6.71×10^{-7}	0.14	P2	P4
7	P01876	Ig alpha-1 chain C region	4.38×10^{-6}	0.20	P2	P3
7	P01876	Ig alpha-1 chain C region	0.00161	0.41	P1	P4
7	P01876	Ig alpha-1 chain C region	0.01201	2.30	P1	P2
8	P01857	Ig gamma-1 chain C region	5.99×10^{-24}	0.33	P1	P2
9	P01860	Ig gamma-3 chain C region	1.53×10^{-11}	0.49	P1	P2
10	P01861	Ig gamma-4 chain C region	3.30×10^{-5}	2.17	P1	P3
11	P01834	Ig kappa chain C region	4.07×10^{-16}	0.47	P1	P4
12	P29622	Kallistatin	4.94×10^{-2}	3.96	P1	P4
13	P35527	Keratin, type I cytoskeletal 9	0.00076	2.58	P3	P4
14	P04264	Keratin, type II cytoskeletal 1	0.03822	0.37	P2	P3
14	P04264	Keratin, type II cytoskeletal 1	3.15×10^{-2}	2.91	P3	P4
15	O00522	Krev interaction trapped protein 1	0.01089	0.36	P1	P3
15	O00522	Krev interaction trapped protein 1	0.04608	0.50	P2	P3
15	O00522	Krev interaction trapped protein 1	0.00677	3.02	P3	P4
16	P36955	Pigment epithelium-derived factor	0.03596	2.20	P3	P4
17	P05543	Thyroxine-binding globulin	4.71×10^{-2}	0.48	P1	P3

the future, which are very important factors in practical applications.

Additionally, we identified several proteins whose concentrations vary substantially among patients but not when considering oocytes from a single patient (see Table 3: Groups III and IV). Previous studies suggested that proteins listed in Group III of Table 3 possessed some potential to become biomarkers of oocytes quality. However, results of our study show that concentration changes of those proteins are a feature of an individual patient rather than a property that differentiates oocytes. It seems that proteins listed in Group III could rather be used as biomarkers associated with a physiological or pathological state of the patient, which does not exclude them

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Table 3. List of Most Relevant Proteins Detected and Quantified in the Study Divided in Eight Groups by Following Factors: (1) Found at Differing Concentrations in Oocytes of a Single Patients or among Patients in HMWF Quantification or Detected in LMWF and (2) Reported as Significantly Differently Abundant in Other Studies on hFF, Only Detected in Other Studies on hFF, or Not Reported Previously

		references		fraction	
group	proteins	significant as differential	identified	HMW	LMW
I. Proteins present at different concentrations among oocytes, reported as important in other studies	Antithrombin-III	8, 13, 18, 27	6, 10−14, 17, 22, 24, 27	×	×
	Complement factor I	16, 18−20	6, 10, 14−20	×	
	Leucine-rich alpha-2-glycoprotein	13, 26	6, 7, 10, 12−17, 21, 26	×	
	Complement factor B	13, 16	6, 7, 10, 12, 13, 15−17, 22	×	×
	Alpha-2-antiplasmin	11, 13	6, 8, 11−15, 20, 22	×	×
	Afamin	11, 12	6, 10−12, 14−17	×	
	Apolipoprotein D	16	7, 10, 11, 14−16	×	
	Interalpha-trypsin inhibitor heavy chain H4	16	6, 14−17	×	×
II. Proteins present at different concentrations among oocytes, identified in other studies	Insulin-like growth factor-binding protein complex acid labile subunit		11, 14, 15, 21, 22	×	
	Heparin cofactor 2		6, 11, 14, 22	×	×
	Alpha-2-macroglobulin		14	×	
III. Proteins present at different concentrations among patients, reported as important in other studies	Haptoglobin	8, 11, 13, 26	6−8, 10−16, 22, 23, 26	×	
	Histidine-rich glycoprotein	6, 11, 17	6, 7, 11, 14−17, 22, 24	×	×
	Apolipoprotein A-IV	13, 20	6, 7, 10, 12−18, 20−22, 24	×	×
	Ig alpha-1 chain C region	13, 18	6, 7, 12−14, 16, 18, 20, 22	×	
	Hemoglobin subunit beta	11, 13	6, 7, 11, 13−16	×	×
	Fibrinogen alpha chain	11	6, 10−12, 14, 15, 21, 22	×	×
	Ig gamma-1 chain C region	13	6, 7, 12−14, 16, 22	×	
	Complement C5	16	6, 11, 14−17	×	×
	Pigment epithelium-derived factor	13	6, 13, 14, 20, 21	×	×
	Keratin, type I cytoskeletal 9	13	6, 13, 21	×	
	Ig gamma-3 chain C region	16	12, 14, 16	×	
	Keratin, type II cytoskeletal 1	13	6, 13	×	
IV. Proteins present at different concentrations among patients, identified in other studies	Ig kappa chain C region		6, 7, 12, 14, 16, 18, 22	×	
	Ig gamma-4 chain C region		7, 12, 14, 16, 22	×	
	Thyroxine-binding globulin		6, 14, 16, 17	×	
	Kallistatin		6, 14	×	
V. Proteins present at different concentrations among patients, unreported in other studies	Krev interaction trapped protein 1			×	
VI. Proteins identified in LMW fraction, reported as important in other studies	Complement C 1s subcomponent	19	6, 10, 14, 15, 17, 19, 22		×
	Complement factor D	19	6, 14, 15, 19		×
	L-lactate dehydrogenase B chain	16	6, 15−17		×
	L-lactate dehydrogenase A chain	16	6, 15, 16		×
	Tissue factor pathway inhibitor	25	6, 7, 25		×
VII. Proteins identified in LMW fraction, identified in other studies	Glia-derived nexin		6, 7, 14−17		×
	Biotinidase		6, 14, 15, 17		×
	Selenoprotein P		6, 15, 17, 21		×
	Collagen alpha-1(I) chain		6, 15, 17		×
	Serglycin		6		×
VIII. Proteins not reported in other studies	see Table 1				

from potential diagnostic application, but they are not necessarily directly related to IVF procedures.

Aside from proteins listed in Table 3 in Groups I−V identified in HMWF (some of them also in LMWF), we identified a substantial number of proteins only at the peptidome (in LMWF) level (see Table 3: Groups VI−VIII). Some of those proteins were identified before in studies on hFF proteome/peptidome composition (see Table 3: Groups VI−VII), and some of them (Group VI) were highlighted as possible biomarkers in previous studies. It is remarkable that

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Table 4. List of Peptides Detected at Statistically Significant Different Concentrations by *t* tests in LMW Fraction among the Samples of Single Follicles of Patients (at Least 2-Fold Change)

N	peptide	charge	*m/z*	ret. time	protein	Uniprot ID	*p* value	fold change	group 1	group 2
1	SSKITHRIHWESASLLR	+4	506.0314	1.70	Complement C3	P01024	0.04904	11.46	P2 F2	P3 F1
2	TLDPERLG	+2	450.7429	7.46	Complement C3	P01024	2.45×10^{-5}	0.28	P2 F2	P4 F2
2	TLDPERLG	+2	450.7429	7.46	Complement C3	P01024	0.02019	4.52	P1 F3	P2 F2
3	MKPVPDLVPGNFK	+3	481.2672	8.72	Fibrinogen alpha chain	P02671	0.04202	0.33	P2 F2	P4 F2
4	MKPVPDLVPGNF	+2	657.3497	9.05	Fibrinogen alpha chain	P02671	0.0185	4.04	P1 F3	P2 F2
5	RPPGFSPF	+2	452.7374	9.04	Kininogen-1	P01042	0.01993	0.31	P2 F2	P4 F2
5	RPPGFSPF	+2	452.7374	9.04	Kininogen-1	P01042	0.03403	3.86	P1 F3	P2 F2
6	RPP[Oxi]GFSPFR	+2	538.7855	1.55	Kininogen-1	P01042	0.00722	0.49	P2 F2	P4 F2
7	RHDWGHEKQR	+3	450.2254	1.56	Kininogen-1	P01042	0.04265	0.29	P2 F2	P4 F2
8	HTFMGVVSLGSPSGEVSHPRKT	+5	462.8390	8.59	Alpha-2-HS-glycoprotein	P02765	0.00393	0.23	P2 F2	P4 F2
8	HTFMGVVSLGSPSGEVSHPRKT	+5	462.8390	8.59	Alpha-2-HS-glycoprotein	P02765	0.0078	0.28	P1 F3	P4 F2
8	HTFMGVVSLGSPSGEVSHPRKT	+5	462.8390	8.59	Alpha-2-HS-glycoprotein	P02765	0.04127	0.29	P2 F2	P3 F1
9	HTFM[Oxi]GVVSLGSPSGEVSHPRKT	+4	582.2957	8.55	Alpha-2-HS-glycoprotein	P02765	0.02042	0.16	P2 F2	P3 F1
9	HTFM[Oxi]GVVSLGSPSGEVSHPRKT	+4	582.2957	8.55	Alpha-2-HS-glycoprotein	P02765	0.00167	0.17	P2 F2	P4 F2
9	HTFM[Oxi]GVVSLGSPSGEVSHPRKT	+4	582.2957	8.55	Alpha-2-HS-glycoprotein	P02765	0.03856	0.26	P1 F3	P3 F1
9	HTFM[Oxi]GVVSLGSPSGEVSHPRKT	+4	582.2957	8.55	Alpha-2-HS-glycoprotein	P02765	0.0055	0.28	P1 F3	P4 F2
10	[PGQ]-QGVNDNEEGFFSARGHRPLDK	+4	589.7821	6.01	Fibrinogen beta chain	P02675	0.02244	0.34	P1 F3	P4 F2
11	LAPLAEDVRGNLR	+3	475.2720	6.54	Apolipoprotein A-IV	P06727	0.00043	0.06	P2 F2	P4 F2
11	LAPLAEDVRGNLR	+3	475.2720	6.54	Apolipoprotein A-IV	P06727	0.0185	0.13	P2 F2	P3 F1
11	LAPLAEDVRGNLR	+3	475.2720	6.54	Apolipoprotein A-IV	P06727	0.01157	0.34	P1 F3	P4 F2
11	LAPLAEDVRGNLR	+3	475.2720	6.54	Apolipoprotein A-IV	P06727	0.01019	5.30	P1 F3	P2 F2
12	SLAELGGHLDQQVEEFR	+3	643.3218	10.14	Apolipoprotein A-IV	P06727	0.0196	0.30	P2 F2	P4 F2
13	NGFKSHALQLNNRQIR	+4	474.7632	1.57	Complement C4−B	P0C0L5	0.02964	5.09	P2 F2	P4 F2
14	NGFKSHALQLNNRQI	+3	580.6482	1.55	Complement C4−B	P0C0L5	0.04748	0.44	P2 F2	P4 F2
15	DAPLQPVTPLQLFEGRRN	+3	684.3726	10.02	Complement C4−B	P0C0L5	0.04488	0.37	P1 F3	P3 F1
16	NGFKSHALQLNNRQIR	+3	632.6819	1.55	Complement C4−B	P0C0L5	0.04052	2.89	P1 F3	P2 F2
16	NGFKSHALQLNNRQIR	+3	632.6819	1.55	Complement C4−B	P0C0L5	0.00687	6.94	P1 F3	P4 F2
17	LMLNPENL	+2	472.2495	1.76	Golgin subfamily A member 3	Q08378	0.00054	9.37	P1 F3	P4 F2
17	LMLNPENL	+2	472.2495	1.76	Golgin subfamily A member 3	Q08378	6.39×10^{-6}	18.65	P2 F2	P4 F2
17	LMLNPENL	+2	472.2495	1.76	Golgin subfamily A member 3	Q08378	0.00018	39.07	P1 F3	P3 F1
17	LMLNPENL	+2	472.2495	1.76	Golgin subfamily A member 3	Q08378	3.22×10^{-6}	77.74	P2 F2	P3 F1
18	TPKNPWSMD	+2	538.2475	10.62	Argininosuccinate synthase	P00966	0.02617	5.02	P1 F3	P3 F1
18	TPKNPWSMD	+2	538.2475	10.62	Argininosuccinate synthase	P00966	0.00377	57.32	P1 F3	P4 F2
19	LCPAILAPSL	+2	499.2912	17.05	RAS protein activator like-3	Q86YV0	0.0271	0.36	P2 F2	P3 F1
20	PEPAKSAPAPKKGS	+3	455.5893	1.57	Histone H2B type F−S	P57053	0.00694	3.80	P1 F3	P2 F2
21	MISAHASNL	+2	472.2369	1.87	Putative uncharacterized protein FLJ38264	Q8N976	0.03062	0.46	P1 F3	P2 F2
21	MISAHASNL	+2	472.2369	1.87	Putative uncharacterized protein FLJ38264	Q8N976	0.03299	2.88	P2 F2	P4 F2
21	MISAHASNL	+2	472.2369	1.87	Putative uncharacterized protein FLJ38264	Q8N976	0.00453	3.77	P2 F2	P3 F1

using a very simple fractionation technique we were able to identify a substantial number of proteins never reported before as hFF components in LMWF of hFF (see Table 1). All of those newly discovered in hFF proteins could potentially become important diagnostic biomarkers; however, more extensive studies should be conducted. A high number of newly discovered proteins identified in our study in the LMWF of hFF is probably associated with the fact that this work is only the second after the recent Chen et al. publication[7] dealing with the hFF peptidome. It is still an open question why so many of proteins could be identified on the peptidome level but could not be found as intact proteins. One obvious reason for such

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

situation is a low concentration of intact proteins in hFF, and such low concentration could be associated with low expression level or quick degradation/proteolysis. In our opinion, analysis of hFF peptidome composition could lead to a better understanding of oocyte condition and physiology, but larger scale studies using a wide range of techniques are necessary.

Low-Molecular-Weight Fraction. As mentioned before, the LMWF was usually discarded during hFF proteomics analyses. Nonetheless, Chen et al. focused their interest in the role of endogenous peptides in hFF as potential biomarkers of in vitro treatment outcome.[7] His group studied relative concentrations of identified peptides in filtrates of hFF (3 kDa membrane) by ion abundance quantification without labeling. Out of 16 454 detected peptides, 53 peptides were distinguished as important to fertilization. Seven of those peptides were successfully assigned to their original proteins, such as serum albumin, insulin-like growth factor binding protein-5, or alpha-2-antiplasmin.

In our attempt of peptide quantification, we analyzed hFF material individually collected from four follicles of four individual patients. Our approach was different from the group of Chen et al. as we analyzed only peptides that we were able to assign to specific proteins. However, as one endogenous peptide originating from a protein may constitute a biomarker itself, we considered single peptides in our analysis, unlike in the HMWF quantification. We quantified 43 distinct peptides originating from 23 proteins (see Supporting Information Part 4). Twenty-two of those peptides were detected in significantly different ($p < 0.05$) concentrations among the samples and identified as a part of sequences of 13 individual proteins. Twenty-one of peptides at significantly varying levels of concentrations were reported at least at 2-fold changes, and those peptides originated from 12 proteins (see Table 4). We were able to quantify one peptide present in hFF samples, which was distinguished by Chen et al. as a potential biomarker. This peptide's sequence is IHWESASLL, and it is a part of complement component C3 protein. Complement component C3 protein has already been reported numerous times in hFF research literature as having function in follicular development.[13,19,26] In our experiment, however, the calculated statistical significance of concentration changes in this peptide was above the set significance threshold. Five of proteins identified in this part of the experiment were not previously reported in hFF literature. The rest of these proteins, similarly to complement component C3 were already identified and detected as differentiating in numerous hFF studies concerning mostly follicle development and reproductive diseases. Among 12 identified proteins, 3 were assigned as present in cell components (see Figure 2), and all of them were not reported before in hFF research: arginosuccinate synthase and RAS protein activator like-3, which localize in cytoplasm, and histone H2B type F−S, which localizes in the nucleus. Peptides originating from proteins localized in the cytoplasm could naturally appear in hFF, especially since these proteins were also reported to act outside the cell. Histone H2B is much less likely to get through to the outside of the cell. Still, the results for these three proteins should be interpreted keeping in mind that their presence may be uncorrelated with natural events but with preparation procedures (see subsection of Quantitative Analysis: High-Molecular-Weight Fraction).

Even though this method can give additional information on hFF proteomic composition and endogenous peptides may constitute excellent biomarkers of follicular (and oocyte) diversity, it is crucial to keep in mind the limitations of this approach. Peptides can be easily bound by various proteins, especially when present in abundance in hFF serum albumin.[54] This event may cause differences in quantitative response in sample repetitions due to peptide retaining on proteins during sample preparation. Observed changes in the measured relative concentration of the particular peptide may not reflect a real situation in the sample; therefore, results of quantitative analysis of LMWF should be evaluated in consideration with extreme caution. The presence of peptides (protein fragments) and changes of peptide concentrations could be used to describe the condition or a physiological/pathological state of the oocyte. However, as it was described previously in this section, analyzing and further using especially quantitative data as a diagnostic tool requires a deeper understanding of protein breakdown chains existing in hFF and also possible secretion routes of peptides by oocyte. In particular, the detection of peptides directly secreted by oocyte could be very valuable to assess oocyte condition and quality; however, it would be extremely difficult to prove the origin of such peptides. Taking all of this into account, the results from quantitative analysis of LMWF of hFF could be used as biomarkers but only with additional information that shows mechanisms that lead to peptide appearance; unfortunately, our knowledge about the dynamics of proteome hFF−oocyte system is so far very limited.

■ CONCLUSIONS

The results of our work show that a very simple fractionation technique of ultrafiltration and simultaneous analysis of LMWFs and HMWFs opens new possibilities in the analysis of the hFF proteome composition. We believe that a simple, cheap, and fast fractionation technique such as the one used in our studies has better potential in possible diagnostic applications than slow, more expensive, and time-consuming fractionation procedures used so far in similar studies.[6,8] Combination of such simple fractionation methods with the label-free SWATH technique (first time used to analyze hFF in this study) to perform quantitative analysis of hFF proteome/peptidome composition could lead to the development of a very efficient diagnostic procedure. Importantly, our study is the first one that attempts to identify biomarkers that differentiate oocytes coming from a single donor. We were able to identify several proteins whose concentrations vary between oocytes (see Table 3: Groups I and II), and those proteins should be considered as a primary target in research leading to the development of a diagnostic tool to perform oocyte quality check and screening before fertilization, which, in consequence, would help to solve ethical and medical issues related to the IVF procedures. The presented results were obtained on a small number of samples (only 12 samples from four donors) and are rather focused on developing procedures and identifying possible molecular targets for a further more extensive search. We plan to perform a large-scale study that will involve 50−70 patients and 100−400 hFF samples (two to five oocytes from a single patient). It should be mentioned that so far even large-scale studies (as large as those which we are planning) did not lead to the establishment of biomarkers of oocyte quality. Therefore, we would like to combine our further proteomics/peptidomics studies with steroid hormone profiling in hFF. Moreover, because hFF composition is strongly coupled to blood serum proteins, simultaneously we would like to find possible correlation between concentrations of

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

proteins in serum and hFF. Such number of samples as well as the introduction of serum proteins and steroid hormones analysis should provide statistically significant data, allowing the identification of biomarkers of oocyte quality that could be used in clinical practice. Moreover, in our further larger scale research we would like to confirm our finding that concentrations of some proteins are characteristic of a patient and reflect the patient's physiological or pathological state (Table 3: Groups III–V).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.7b00366.

Supporting Information Part 1: Precursor and fragment ion spectra of single peptide identifications for identified proteins, supplementary gene ontology classification results, and volcano plots for HMWF and LMWF proteins changes in follicles of the same patient and between patients. (PDF)

Supporting Information Part 2: Identification information on peptides and assigned proteins found in LMWF and HMWF of human follicular fluid. (ZIP)

Supporting Information Part 3: Information on transitions, peptides and proteins quantified in HMWF of hFF in SWATH-MS experiments along with statistical analysis. (XLSX)

Supporting Information Part 4: Information on transitions, peptides, and proteins quantified in LMWF of hFF in SWATH-MS experiments along with statistical analysis. (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

*A.L.: Tel: +48 58 523 64 34. E-mail: aleksandra.lewandowska@biotech.ug.edu.pl.

*S.O.: Tel: +48 58 523 64 28. E-mail: stanislaw.oldziej@biotech.ug.edu.pl.

ORCID

Aleksandra E. Lewandowska: 0000-0002-7750-7585

Notes

The authors declare no competing financial interest.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium[44] via the PRIDE[45] partner repository with the data set identifier PXD006550.

■ ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no 316094. The scientific work was financed from funds for science in the years 2013–2016 allocated for the internationally cofinanced project.

■ REFERENCES

(1) Revelli, A.; Delle Piane, L.; Casano, S.; Molinari, E.; Massobrio, M.; Rinaudo, P. Follicular fluid content and oocyte quality: from single biochemical markers to metabolomics. *Reprod. Biol. Endocrinol.* **2009**, *7*, 40.

(2) Fortune, J. E. Ovarian Follicular Growth and Development in Mammals. *Biol. Reprod.* **1994**, *50*, 225–232.

(3) Balaban, B.; Urman, B. Effect of oocyte morphology on embryo development and implantation. *Reprod. BioMed. Online* **2006**, *12* (5), 608–615.

(4) Dawson, A.; Griesinger, G.; Diedrich, K. Screening oocytes by polar body biopsy. *Reprod. BioMed. Online* **2006**, *13* (1), 104–109.

(5) Patrizio, P.; Fragouli, E.; Bianchi, V.; Borini, A.; Wells, D. Molecular methods for selection of the ideal oocyte. *Reprod. BioMed. Online* **2007**, *15* (3), 346–353.

(6) Zamah, A. M.; Hassis, M. E.; Albertolle, M. E.; Williams, K. E. Proteomic analysis of human follicular fluid from fertile women. *Clin. Proteomics* **2015**, *12* (1), 1–12.

(7) Chen, F.; Spiessens, C.; D'Hooghe, T.; Peeraer, K.; Carpentier, S. Follicular fluid biomarkers for human in vitro fertilization outcome: Proof of principle. *Proteome Sci.* **2016**, *14* (1), 17.

(8) Ambekar, A. S.; Kelkar, D. S.; Pinto, S. M.; Sharma, R.; Hinduja, I.; Zaveri, K.; Pandey, A.; Prasad, T. S. K.; Gowda, H.; Mukherjee, S. Proteomics of follicular fluid from women with polycystic ovary syndrome suggests molecular defects in follicular development. *J. Clin. Endocrinol. Metab.* **2015**, *100* (2), 744–753.

(9) Bayasula; Iwase, A.; Kobayashi, H.; Goto, M.; Nakahara, T.; Nakamura, T.; Kondo, M.; Nagatomo, Y.; Kotani, T.; Kikkawa, F. A proteomic analysis of human follicular fluid: Comparison between fertilized oocytes and non-fertilized oocytes in the same patient. *J. Assist. Reprod. Genet.* **2013**, *30* (9), 1231–1238.

(10) Angelucci, S.; Ciavardelli, D.; Di Giuseppe, F.; Eleuterio, E.; Sulpizio, M.; Tiboni, G. M.; Giampietro, F.; Palumbo, P.; Di Ilio, C. Proteome analysis of human follicular fluid. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764* (11), 1775–1785.

(11) Kushnir, M. M.; Naessén, T.; Wanggren, K.; Rockwood, A. L.; Crockett, D. K.; Bergquist, J. Protein and steroid profiles in follicular fluid after ovarian hyperstimulation as potential biomarkers of IVF outcome. *J. Proteome Res.* **2012**, *11* (10), 5090–5100.

(12) Hanrieder, J.; Zuberovic, A.; Bergquist, J. Surface modified capillary electrophoresis combined with in solution isoelectric focusing and MALDI-TOF/TOF MS: A gel-free multidimensional electrophoresis approach for proteomic profiling-Exemplified on human follicular fluid. *J. Chromatogr. A* **2009**, *1216* (17), 3621–3628.

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

(14) Twigt, J.; Steegers-Theunissen, R. P.; Bezstarosti, K.; Demmers, J. A. A. Proteomic analysis of the microenvironment of developing oocytes. *Proteomics* **2012**, *12* (9), 1463–1471.

(15) Ambekar, A. S.; Nirujogi, R. S.; Srikanth, S. M.; Chavan, S.; Kelkar, D. S.; Hinduja, I.; Zaveri, K.; Prasad, T. S. K.; Harsha, H. C.; Pandey, A.; et al. Proteomic analysis of human follicular fluid: A new perspective towards understanding folliculogenesis. *J. Proteomics* **2013**, *87*, 68–77.

(16) Regiani, T.; Cordeiro, F. B.; da Costa, L.; Salgueiro, J.; Cardozo, K.; Carvalho, V. M.; Perkel, K. J.; Zylbersztejn, D. S.; Cedenho, A. P.; Lo Turco, E. G. Follicular fluid alterations in endometriosis: label-free proteomics by MS(E) as a functional tool for endometriosis. *Syst. Biol. Reprod. Med.* **2015**, *61* (5), 263–276.

(17) Twigt, J. M.; Bezstarosti, K.; Demmers, J.; Lindemans, J.; Laven, J. S. E.; Steegers-Theunissen, R. P. Preconception folic acid use influences the follicle fluid proteome. *Eur. J. Clin. Invest.* **2015**, *45* (8), 833–841.

(18) Lo Turco, E. G.; Cordeiro, F. B.; de Carvalho Lopes, P. H.; Gozzo, F. C.; Pilau, E. J.; Soler, T. B.; da Silva, B. F.; Del Giudice, P. T.; Bertolla, R. P.; Fraietta, R.; et al. Proteomic analysis of follicular fluid from women with and without endometriosis: New therapeutic targets and biomarkers. *Mol. Reprod. Dev.* **2013**, *80* (6), 441–450.

(19) Yoo, S. W.; Bolbot, T.; Koulova, A.; Sneeringer, R.; Humm, K.; Dagon, Y.; Usheva, A. Complement factors are secreted in human follicular fluid by granulosa cells and are possible oocyte maturation factors. *J. Obstet. Gynaecol. Res.* **2013**, *39* (2), 522–527.

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053–3067

(20) Jarkovska, K.; Kupcova Skalnikova, H.; Halada, P.; Hrabakova, R.; Moos, J.; Rezabek, K.; Gadher, S. J.; Kovarova, H. Development of ovarian hyperstimulation syndrome: interrogation of key proteins and biological processes in human follicular fluid of women undergoing in vitro fertilization. *Mol. Hum. Reprod.* **2011**, *17* (11), 679−692.
(21) Yoo, S. W.; Savchev, S.; Sergott, L.; Rezai, T.; Lopez, M. F.; Von Wald, T.; Eaton, J. L.; Reindollar, R.; Usheva, A. A Large Network of Interconnected Signaling Pathways in Human Ovarian Follicles is Supported by the Gene Expression Activity of the Granulosa Cells. *Reprod. Sci.* **2011**, *18* (5), 476−484.
(22) Hanrieder, J.; Nyakas, A.; Naessén, T.; Bergquist, J. Proteomic Analysis of Human Follicular Fluid Using an Alternative Bottom-Up Approach. *J. Proteome Res.* **2008**, *7* (1), 443−449.
(23) Schweigert, F. J.; Gericke, B.; Wolfram, W.; Kaisers, U.; Dudenhausen, J. W. Peptide and protein profiles in serum and follicular fluid of women undergoing IVF. *Hum. Reprod.* **2006**, *21* (11), 2960−2968.
(24) Jarkovska, K.; Martinkova, J.; Liskova, L.; Halada, P.; Moos, J.; Rezabek, K.; Gadher, S. J.; Kovarova, H. Proteome mining of human follicular fluid reveals a crucial role of complement cascade and key biological pathways in women undergoing in vitro fertilization. *J. Proteome Res.* **2010**, *9* (3), 1289−1301.
(25) Thyzel, E.; Siegling, S.; Götting, C.; Tinneberg, H.-R.; Brinkmann, T.; Kleesiek, K. Quantification of tissue factor pathway inhibitor in human seminal plasma and in human follicular fluid. *Thromb. Res.* **2003**, *109* (5−6), 329−332.
(26) Wu, Y.-T.; Wu, Y.; Zhang, J.-Y.; Hou, N.-N.; Liu, A.-X.; Pan, J.-X.; Lu, J.-Y.; Sheng, J.-Z.; Huang, H.-F. Preliminary proteomic analysis on the alterations in follicular fluid proteins from women undergoing natural cycles or controlled ovarian hyperstimulation. *J. Assist. Reprod. Genet.* **2015**, *32* (3), 417−427.
(27) Kim, Y.-S.; Kim, M.-S.; Lee, S.-H.; Choi, B.-C.; Lim, J.-M.; Cha, K. Y.; Baek, K.-H. Proteomic analysis of recurrent spontaneous abortion: Identification of an inadequately expressed set of proteins in human follicular fluid. *Proteomics* **2006**, *6* (11), 3445−3454.
(28) Timms, J. F.; Arslan-Low, E.; Kabir, M.; Worthington, J.; Camuzeaux, S.; Sinclair, J.; Szaub, J.; Afrough, B.; Podust, V. N.; Fourkala, E.-O.; et al. Discovery of serum biomarkers of ovarian cancer using complementary proteomic profiling strategies. *Proteomics: Clin. Appl.* **2014**, *8* (11−12), 982−993.
(29) Oonk, S.; Spitali, P.; Hiller, M.; Switzar, L.; Dalebout, H.; Calissano, M.; Lochmüller, H.; Aartsma-Rus, A.; Hoen, P. A. C.; van der Burgt, Y. E. M. Comparative mass spectrometric and immuno-assay-based proteome analysis in serum of Duchenne muscular dystrophy patients. *Proteomics: Clin. Appl.* **2016**, *10* (3), 290−299.
(30) Ortea, I.; Roschitzki, B.; López-Rodríguez, R.; Tomero, E. G.; Ovalles, J. G.; López-Longo, J.; de la Torre, I.; González-Alvaro, I.; Gómez-Reino, J. J.; González, A. Independent Candidate Serum Protein Biomarkers of Response to Adalimumab and to Infliximab in Rheumatoid Arthritis: An Exploratory Study. *PLoS One* **2016**, *11* (4), e0153140.
(31) Pernemalm, M.; Lehtiö, J. Mass spectrometry-based plasma proteomics: state of the art and future outlook. *Expert Rev. Proteomics* **2014**, *11* (4), 431−448.
(32) Tammen, H.; Schulte, I.; Hess, R.; Menzel, C.; Kellmann, M.; Mohring, T.; Schulz-Knappe, P. Peptidomic analysis of human blood specimens: Comparison between plasma specimens and serum by differential peptide display. *Proteomics* **2005**, *5* (13), 3414−3422.
(33) Schuhmann, M. U.; Zucht, H. D.; Nassimi, R.; Heine, G.; Schneekloth, C. G.; Stuerenburg, H. J.; Selle, H. Peptide screening of cerebrospinal fluid in patients with glioblastoma multiforme. *Eur. J. Surg. Oncol.* **2010**, *36* (2), 201−207.
(34) Jürgens, M.; Appel, A.; Heine, G.; Neitz, S.; Menzel, C.; Tammen, H.; Zucht, H.-D. Towards characterization of the human urinary peptidome. *Comb. Chem. High Throughput Screening* **2005**, *8* (8), 757−765.
(35) Wallace, M.; Cottell, E.; Gibney, M. J.; McAuliffe, F. M.; Wingfield, M.; Brennan, L. An investigation into the relationship between the metabolic profile of follicular fluid, oocyte developmental potential, and implantation outcome. *Fertil. Steril.* **2012**, *97* (5), 1078−1084.e8.
(36) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717−O111.016717.
(37) Łukaszuk, K.; Kunicki, M.; Liss, J.; Bednarowska, A.; Jakiel, G. Probability of live birth in women with extremely low anti-Müllerian hormone concentrations. *Reprod. BioMed. Online* **2014**, *28* (1), 64−69.
(38) Łukaszuk, K.; Pukszta, S.; Wells, D.; Cybulska, C.; Liss, J.; Płóciennik, Ł.; Kuczyński, W.; Zabielska, J. Routine use of next-generation sequencing for preimplantation genetic diagnosis of blastomeres obtained from embryos on day 3 in fresh in vitro fertilization cycles. *Fertil. Steril.* **2015**, *103* (4), 1031−1036.
(39) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. Characterization of the Low Molecular Weight Human Serum Proteome* □ S. *Mol. Cell. Proteomics* **2003**, *2*, 1096−1103.
(40) Merrell, K.; Southwick, K.; Graves, S. W.; Esplin, M. S.; Lewis, N. E.; Thulin, C. D. Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *J. Biomol. Technol.* **2004**, *15* (4), 238−248.
(41) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr. Protoc. Mol. Biol.* **2009**.
(42) Thomas, P. D.; Campbell, M. J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* **2003**, *13* (9), 2129−2141.
(43) Mi, H.; Dong, Q.; Muruganujan, A.; Gaudet, P.; Lewis, S.; Thomas, P. D. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **2010**, *38* (suppl_1), D204−D210.
(44) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32* (3), 223−226.
(45) Vizcaíno, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **2016**, *44* (D1), D447−D456.
(46) Zhang, Y.; Guo, Z.; Zou, L.; Yang, Y.; Zhang, L.; Ji, N.; Shao, C.; Sun, W.; Wang, Y. A comprehensive map and functional annotation of the normal human cerebrospinal fluid proteome. *J. Proteomics* **2015**, *119*, 90−99.
(47) Tong, L.; Zhou, X. Y.; Jylha, A.; Aapola, U.; Liu, D. N.; Koh, S. K.; Tian, D.; Quah, J.; Uusitalo, H.; Beuerman, R. W.; et al. Quantitation of 47 human tear proteins using high resolution multiple reaction monitoring (HR-MRM) based-mass spectrometry. *J. Proteomics* **2015**, *115*, 36−48.
(48) Selevsek, N.; Chang, C.-Y.; Gillet, L. C.; Navarro, P.; Bernhardt, O. M.; Reiter, L.; Cheng, L.-Y.; Vitek, O.; Aebersold, R. Reproducible and Consistent Quantification of the *Saccharomyces cerevisiae* Proteome by SWATH-mass spectrometry. *Mol. Cell. Proteomics* **2015**, *14* (3), 739−749.
(49) Schulz-Knappe, P.; Hans-Dieter, Z.; Heine, G.; Jürgens, M.; Schrader, M. Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb. Chem. High Throughput Screening* **2012**, *4* (2), 207−217.
(50) Betz, B. B.; Jenks, S. J.; Cronshaw, A. D.; Lamont, D. J.; Cairns, C.; Manning, J. R.; Goddard, J.; Webb, D. J.; Mullins, J. J.; Hughes, J.; et al. Urinary peptidomics in a rodent model of diabetic nephropathy highlights epidermal growth factor as a biomarker for renal deterioration in patients with type 2 diabetes. *Kidney Int.* **2016**, *89* (5), 1125−1135.
(51) Liu, Y.; Hüttenhain, R.; Surinova, S.; Gillet, L. C.; Mouritsen, J.; Brunner, R.; Navarro, P.; Aebersold, R. Quantitative measurements of

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

N -linked glycoproteins in human plasma by SWATH-MS. *Proteomics* **2013**, *13* (8), 1247−1256.
(52) Dwivedi, R. C.; Navarrete, M.; Choi, N.; Spicer, V.; Rigatto, C.; Arora, R. C.; Krokhin, O.; Ho, J.; Wilkins, J. A. A proteomic evaluation of urinary changes associated with cardiopulmonary bypass. *Clin. Proteomics* **2016**, *13* (1), 17.
(53) Vaswani, K.; Ashman, K.; Reed, S.; Salomon, C.; Sarker, S.; Arraztoa, J. A.; Pérez-Sepúlveda, A.; Illanes, S. E.; Kvaskoff, D.; Mitchell, M. D.; et al. Applying SWATH Mass Spectrometry to Investigate Human Cervicovaginal Fluid During the Menstrual Cycle. *Biol. Reprod.* **2015**, *93* (2), 39.
(54) Hamerly, T.; Heinemann, J.; Tokmina-Lukaszewska, M.; Lusczek, E. R.; Mulier, K. E.; Beilman, G. J.; Bothner, B. Bovine serum albumin as a molecular sensor for the discrimination of complex metabolite samples. *Anal. Chim. Acta* **2014**, *818*, 61−66.

DOI: 10.1021/acs.jproteome.7b00366
J. Proteome Res. 2017, 16, 3053−3067

Journal of Proteomics 191 (2019) 131–142

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Human follicular fluid proteomic and peptidomic composition quantitative studies by SWATH-MS methodology. Applicability of high pH RP-HPLC fractionation

Aleksandra E. Lewandowska[a,*], Katarzyna Macur[a], Paulina Czaplewska[a], Joanna Liss[b], Krzysztof Łukaszuk[b,c,d], Stanisław Ołdziej[a,*]

[a] Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, Abrahama 58, Gdańsk 80-307, Poland
[b] INVICTA Fertility and Reproductive Center, Trzy Lipy 3, Gdańsk 80-172, Poland
[c] Department of Obstetrics and Gynecological Nursing, Faculty of Health Sciences, Medical University of Gdańsk, Dębinki 7, Gdańsk 80-211, Poland
[d] Department of Gynaecological Endocrinology, Medical University of Warsaw, Karowa 2, Warsaw 00-315, Poland

ARTICLE INFO

Keywords:
Human follicular fluid
High pH RP-HPLC fractionation
SWATH-MS
Proteome
Peptidome

ABSTRACT

Analysis of proteomic composition of human follicular fluid (hFF) has been previously proposed as a potential tool of oocyte quality evaluation. In order to develop an efficient method to investigate the hFF proteome and peptidome components, we applied and tested a few prefractionation schemes of hFF material consisting of ultrafiltration, optional immunodepletion, and high pH RP-HPLC separation by building spectral libraries and comparing their quantification capabilities of unfractionated samples. Low Molecular-Weight Fraction peptides (LMWF, < 10 kDa) and High Molecular-Weight Fraction proteins (HMWF, > 10 kDa) resulting from ultrafiltration were analyzed separately. We identified 302 proteins in HMWF and 161 proteins in LMWF in all qualitative experiments. All LMWF peptidomic libraries turned out to be of poor quantification quality, however they enabled measurement of higher numbers of peptides with increasing input of experiment data, in contrast to HMWF proteomic libraries. We were able to quantify a total of 108 HMWF proteins and 250 LMWF peptides (from 84 proteins) in all experiments. Employment of high RP-HPLC fractionation allowed for identification of a much broader set of proteins, however did not significantly improve the quantification capabilities of the applied method. Data are available via ProteomeXchange with identifier PXD008073.

Significance: In the search of biomarkers for assessment of oocyte quality in assisted reproductive technology, many studies are devoted to analysis of follicular fluid composition. Candidates for such biomarkers can be located in both the proteome and the recently investigated peptidome of hFF. Reliable qualitative and especially quantitative analysis of complex mixtures such as hFF, requires development of a fast and preferably inexpensive analytical procedure. The powerful SWATH-MS technique is well suited for quantitative label-free analysis of complex protein and peptide mixtures. However, for efficient usage it needs well designed and constructed MS-spectral libraries as well as a proper protocol for sample preparation. We investigated the influence of the size and quality of MS-spectral libraries (different spectral libraries are constructed using various sample prefractionation protocols) on SWATH experiments on hFF proteome and peptidome. In the case of peptidome investigation, increasing the size of spectral libraries led to quantification of more peptides in a single experiment. For the proteome, increasing the size of spectral libraries improved quantification only to a limited extend, and further extension of spectral libraries even worsened results. Nevertheless, using the best selected prefractionation schemes and spectral libraries we were able to quantify as many as 79 proteins of hFF proteome and 106 peptides (from 53 proteins) of hFF peptidome in single experiments. The spectral libraries and prefractionation protocols we developed allow for a large scale fast scan of hundreds of clinical hFF samples in the search for biomarkers for evaluation of oocyte quality.

* Corresponding authors.
E-mail addresses: aleksandra.lewandowska@biotech.ug.edu.pl (A.E. Lewandowska), stanislaw.oldziej@biotech.ug.edu.pl (S. Ołdziej).

https://doi.org/10.1016/j.jprot.2018.03.010
Received 30 October 2017; Received in revised form 25 January 2018; Accepted 7 March 2018
Available online 10 March 2018
1874-3919/

1. Introduction

A significant fraction of proteins present in human follicular fluid (hFF) affect regulation of the folliculogenesis process and thus oocyte growth and development in diverse individual ways [1–3]. Therefore, qualitative or quantitative analysis of hFF proteins has been proposed as a new direction of developing a method for oocyte quality evaluation [4–7]. In the long term, such non-invasive quality assessment could allow fertilization of only those oocytes which are most promising in terms of achieving a successful pregnancy. Proteomic composition of hFF has been investigated in numerous studies [4–14]. hFF is a selective filtrate of blood enriched by secretions from the few cells present in the follicle and its composition is closely related to blood serum. Considering this similarity, proteomic investigation of hFF is facing the same issues affecting blood serum studies: a great abundance of distinct proteins and substantial differences in their dynamic concentrations [15–17]. These properties cause considerable difficulties in simultaneous quantification or even identification of high and lower abundant proteins in one analysis. In order to, at least partly, overcome this problem various fractionation methods have been used in hFF proteome studies including: 1-D and 2-D electrophoresis (1-,2-DE) [6,8,9], isoelectric focusing (IEF) [8–10], liquid chromatography (LC) separations [7,9,11–13], ultrafiltration [5,14], or immunodepletion [5,8,9,11,12]. Increasingly more approaches to hFF analysis involve combination of two or more distinct fractionation methods, especially utilizing immunodepletion of most abundant blood serum proteins (e.g., albumin/IgG removal kits or removal of the 14 most abundant proteins by MARS-14 kit) [5,8,9,11,12]. Immunodepletion of the 14 most abundant proteins leads to 95% reduction of protein content in blood serum, and similarly in proximate in composition hFF, allowing the analysis of remaining proteins present at much lower concentrations. Other technique used more and more frequently in hFF proteome research as a stand-alone fractionation scheme [7,13] or in second dimension of more complicated fractionation procedures [11,12] is high pH RP-HPLC. This method constitutes an attractive alternative to other off-line orthogonal separations (e.g., strong cation exchange – SCX) offering simplicity, reduction in sample losses, and importantly higher proteome coverage [18].

In our previous study [14], we employed a relative quantification technique SWATH-MS (Sequential Windowed Acquisition of All Theoretical Fragment Ion Spectra) [19] allowing label-free quantification of virtually any number of proteins in one analysis. We divided the hFF material to Low Molecular-Weight Fraction (LMWF, < 10 kDa) and High Molecular-Weight Fraction (HMWF, > 10 kDa) by ultrafiltration and examined proteomes of both fractions. The number of quantifiable proteins/peptides and quality of their quantification in SWATH-MS technique relies heavily on the spectral library previously constructed from IDA (Information Dependent Acquisition) measurements conducted on samples of analyzed material [20]. Utilizing only such simple fractionation method as ultrafiltration in combination with immunodepletion in several HMWF samples, we were able to quantify 72 proteins in HMWF and 43 distinct peptides of 23 proteins in LMWF in a single experiment. Ultrafiltration is a simple fractionation method that leads to detection of a relatively small number of proteins (fewer than 200) in comparison with more sophisticated multistep procedures which allow identification of 700–800 proteins [11,12]. It is also important to mention that, to our knowledge as the only group working on hFF proteome, we utilize a setup of a microLC system coupled to a mass spectrometer, in contrast to a more popular setup with a nanoLC system. While nanoLC offers a higher threshold of sensitivity, allowing for more protein identifications in a single run, the advantage of microLC are shorter analysis times resulting in a higher throughput, necessary for potential large-scale clinical studies [21]. Therefore, to maximize the number of quantified proteins using our previously developed methodology for a potential clinical study, in the present study we decided to use a more complex fractionation scheme to create more comprehensive spectral libraries: ultrafiltration to LMWF and HMWF, optional immunodepletion in HMWF samples, and high pH RP-HPLC separations at various conditions. This procedure allowed us to identify a larger number of proteins and to build 18 spectral libraries resulting from those experiments (10 for HMWF and 8 for LMWF). Afterwards, we evaluated the potential of newly built spectral libraries in quantification of proteins/peptides in unfractionated pool hFF samples. The aim of this study was to examine the SWATH-MS technique's potential to quantify proteins identified after application of a multistep fractionation procedure to investigate relevance of fractionation during spectral library formation in SWATH-MS quantification of hFF proteome and peptidome.

2. Materials and methods

2.1. Research approval

Development of the analytical method described in this work is a part of the “Identification of biomarkers of early embryonic development and pregnancy” project that has been approved by the Independent Bioethics Commission at the Medical University of Gdansk (decision 62/2016). Each couple undergoing the IVF procedure has signed written informed consent regarding the treatment and all included procedures. The obtained written consents also include agreement for publication of treatment related data as long as patient anonymity is maintained.

2.2. Initial sample preparation

Results of our previous study [14] show that concentrations of some proteins may vary substantially between hFF samples taken from the same donor, regardless of their medical condition or age. Therefore, a pool sample of hFFs, composed of samples from two to five individual follicles taken from several randomly chosen patients undergoing in vitro fertilization treatment was obtained from the INVICTA Fertility and Reproductive Center in Gdansk. The general protocol of the patients' hormonal stimulation and hFF retrieval procedure was described in our previous work [14]. Obtained sample was free of visible blood contamination. The material was centrifuged at 1000 × *g* for 10 min and separated from cell pellets. Subsequently, resulting sample was divided into 2 ml aliquots stored in − 20 °C until further processing and analysis.

2.3. Sample processing for individual quantitative measurements

Each 400 µl of hFF sample was mixed with 100 µl of acetonitrile (ACN), vortexed briefly to break protein-protein and peptide-protein interactions, and subjected to ultrafiltration on a 10 kDa Amicon membrane (Merck-Millipore) at 14,000 g for 15 min followed by a reverse spin at 1000 g for 2 min. Resulting filtrates containing peptides are further referred to as Low Molecular-Weight Fraction (LMWF), and retentates containing proteins as High Molecular-Weight Fraction (HMWF). Five samples of LMWF resulting from separate ultrafiltration of 400 µl of the pool sample of hFF were individually evaporated to dryness in a SpeedVac. Subsequently, LMWF samples were dissolved in 50 µl of 0.5% trifluoroacetic acid (TFA) in water, filtrated through 0.22 µm cellulose acetate membrane at 16000 g for 2 min, and desalted on C18 Pierce SpinTips according to manufacturer's protocol with the following exceptions: whole procedure was repeated 3 times with increasing ACN concentration in the elution step (30, 50, and 80% ACN in water with 0.1% formic acid (FA). Resulting eluates were evaporated to dryness in a SpeedVac and reconstituted in 30 µl of 50% ACN 0.1% FA in water and subjected to LC-MS/MS measurements (SWATH-MS measurements and IDA experiments for basic spectral library construction). Obtained HMWF was diluted by 50 mM NH_4HCO_3 to protein concentration of 1 mg/ml. Five samples of diluted HMWF containing

18 μg of protein were further subjected to proteolytic digestion individually according to a protocol proposed by Gundry et al. [22]. First, samples were incubated with 10 mM of dithiotreitol (reducing agent) in 56 °C for 30 min. After cooling off, samples were incubated with 20 mM of iodoacetamide (alkylating agent) in darkness at room temperature for 30 min. Afterwards, trypsin in 1:50 ratio of enzyme to substrate was added to the samples, and the digestion reaction was carried out in 37 °C for 19 h. Digestion was stopped by addition of 5% TFA in 50% ACN/water by decreasing solution pH to 3. Samples were then evaporated to dryness in SpeedVac, reconstituted in 50 μl of 0.5% TFA in water, desalted on SpinTips, and prepared for the LC-MS/MS measurements as it was described above for the LMWF fraction.

2.4. Sample processing for spectral libraries construction

The material was divided into Low and High Molecular Weight Fraction as described in Sample Processing for Individual Quantitative Measurements paragraph. Multiple LMWF samples resulting from ultrafiltration were combined before HPLC separation to increase the amount of proteomic material resulting in samples of 0.8, 1.2, 1.6, 2.4 ml of ultrafiltrated hFF. Prior to fractionation, samples were reconstituted in HPLC separation buffer A and filtrated on 0.22 μm cellulose acetate filters at 16000 g for 2 min. Five samples of 100 μg of diluted HMWF were subjected to reduction, alkylation, and tryptic digestion as it was described in Sample Processing for Individual Quantitative Measurements paragraph. Samples were evaporated to dryness and reconstituted in a respective HPLC separation buffer A prior to HPLC fractionation.

2.4.1. Immunodepletion

A part of the pool material was subjected to fractionation prior to ultrafiltration to obtain immunodepleted material for HMWF measurements. Those samples were depleted of 14 most abundant human serum proteins (albumin, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) using a Multiple Affinity Removal Spin Cartridge Human 14 (MARS-14) kit (Agilent Technologies, Santa Clara, CA) according to manufacturer's protocol. Both, immunodepleted material and high abundant proteins eluted from the resin were subjected to trypsin digestion and desalted as described in Sample Processing for Individual Quantitative Measurements paragraph. Three samples of 100 μg of immunodepleted protein material were not desalted after digestion and instead reconstituted in a respective HPLC separation buffer A and further fractionated by HPLC.

2.4.2. Additional concentrated LMWF samples

Three LMWF samples resulting from ultrafiltration of 800 μl of hFF were evaporated to dryness, desalted and prepared for IDA experiments to build a more comprehensive basic spectral library without fractionation (according to procedures described in Sample Processing for Individual Quantitative Measurements paragraph).

2.4.3. High pH RP-HPLC fractionation

To increase the number of protein identifications in both analyzed fractions of hFF we further divided the material in a series of high pH RP-HPLC fractionation experiments. The details of each experiment (sample type, chromatographic gradient, buffers, and column used along with the resulting final spectral library name) are listed in Supplementary Table 1. All experiments were carried out on analytical Nexera XR HPLC System with the PDA detector (Shimadzu, Kyoto, Japan). The column temperature was kept at 40 °C and the flow rate was 1 ml/min during each separation. The combination of column and buffer system used was either Jupiter Proteo 90 Å column (4 μm, 250 × 2.6 mm, Phenomenex, Torrance, CA) and buffers A: 100 mM ammonium bicarbonate pH 8, and B: 100% ACN or Zorbax Extend-C18 column (4.6 × 150 mm, 5 μm particle size, Agilent Technologies, Santa Clara, CA) and buffers A: 0.1% NH_4OH in water, pH 10, and B: 0.1% NH_4OH in ACN, pH 10. Chromatographic gradients used for fractionations are described in detail in Supplementary Table 1. In each 30 min chromatographic run 16 fractions of 2 ml were manually collected; 15 fractions were collected one fraction per 2 min and one additional fraction collected after the end of the separation during the column washing. All collected fractions from all separations were evaporated to dryness in a SpeedVac and reconstituted in 12 μl of 50% ACN 0.1% FA in water for the LC-MS/MS measurements.

2.5. LC-MS/MS measurements

Chromatographic separation was performed on the Ekspert MicroLC 200 Plus System (Eksigent, Redwood City, CA) using the ChromXP C18CL column (3 μm, 120 Å, 150 × 0.3 mm). Samples were loaded onto the column using the CTC Pal Autosampler (CTC Analytics AG, Zwinger, Switzerland) with each injection of 5 μl of sample. Buffers A and B constituted of 0.1% FA in water and 0.1% FA in ACN, respectively. The chromatographic gradient was 10–90% B in 30 min for each run. Eluate from the column was analyzed in a positive ion mode on a TripleTOF 5600$^+$ mass spectrometer with DuoSpray Ion Source (SCIEX, Framingham, MA). The microLC-MS/MS system was controlled by the SCIEX Analyst TF 1.7.1 software.

2.5.1. Shotgun mass spectrometry experiments

The IDA experiments were conducted for all investigated samples. The TOF MS survey scan was performed in the m/z range of 100–2000 with the accumulation time of 40 ms, which resulted in a duty cycle of 1.11 s. Each analysis was carried out with rolling collision energy, and the precursor ions were excluded from reselection for 5 s after two occurrences.

2.5.2. SWATH mass spectrometry experiments

SWATH-MS analyses, according to a method developed by Ruedi Aebersold's group [19], were performed for the 5 respective repetitions of non-fractionated pool sample filtrates (LMWF) and digested retentates (HMWF). Experiments were performed in a looped product ion mode. A set of 25 transmission windows (25 Da wide) was constructed and covered the precursor mass range of 400–1000 m/z. The collision energy for each window was calculated for a + 2 to +5 charged ions centered upon the window with a spread of 2. The SWATH-MS1 survey scan was acquired in high sensitivity mode in the range of 100–2000 Da in the beginning of each cycle with the accumulation time of 50 ms, and it was followed by 40 ms accumulation time high sensitivity product ion scans, which resulted in the total cycle time of 1.11 s.

2.6. Data analysis

2.6.1. Database search

Protein database searches were conducted in ProteinPilot 4.5 Software (SCIEX) using the Paragon algorithm against the SwissProt *Homo sapiens* database (ver. 31.07.2017, 20,214 entries) with an automated false discovery rate. The search parameters for HMWF included: instrument TripleTOF 5600, alkylation of cysteines by iodoacetamide, trypsin enzyme digestion, ID focus on biological modifications, search effort “thorough ID”, and detected protein threshold [Conf] > 10%. The LMWF samples were processed accordingly with following changes in parameters: no modification of cysteines and no enzyme specified in digestion. Protein identifications in HMWF were considered to be genuine only at protein FDR < 0.01% to reduce the possibility of false discovery as much as possible, considering the fact that hFF proteome has already been studied extensively. We have set more relaxed conditions for peptide identifications in the less studied LMWF: (1) peptide confidence ≥99%, and (2) peptide FDR < 0.5%, to obtain more comprehensive results that could

be narrowed in the future research. Database searches of HMWF were performed for listed combinations: (1) five non-fractionated samples, (2) desalted immunodepleted and most abundant proteins fractions, (3) database search number 1, and database search number 2, (4) combined searches of 16 fractions resulted from a single RP-HPLC separation run described in Supplementary Table 1. for each individual fractionation (eight in total), (5) database search number 3 with database searches number 4 (eight in total), and (6) all conducted HMWF measurements. Database searches for LMWF were carried out in a similar manner, excluding samples resulting from, not performed in this fraction, immunodepletion and instead utilizing additional concentrated LMWF samples (see Materials and Methods section). Functional properties of proteins identified in database searches were assigned using PANTHER Gene Ontology database [23,24].

2.6.2. SWATH MS data processing

SWATH MS measurements were processed with libraries resulting from database searches number 1, 5, and 6, as it was described in the previous paragraph. Each resulting ProteinPilot.group file was loaded into MS/MS All with SWATH Acquisition MicroApp 2.01 in PeakView 2.2 (SCIEX) to automatically create a spectral library with the following parameters: maximum number of proteins equal to number of proteins at the FDR threshold set for a given fraction, modified peptides allowed, and shared peptides excluded. Each spectral library was processed with SWATH measurements of the pool hFF sample (three technical and five biological repetitions resulting in 15 runs). During data processing the maximum number of peptides per protein was set to 6. In HMWF only proteins with 2 valid peptides were quantified, while for LMWF proteins one peptide was sufficient for quantification as we focused on specific peptides. The number of transitions for each peptide was 6, and any peptide with fewer measurable transitions was excluded from the analysis. Starting parameters for peptides were [Conf] ≥ 99 and FDR < 1%. Extracted ion chromatogram (XIC) width was set to 75 ppm, and the offset XIC extraction window was set to 15 min. In HMWF experiments we performed retention time calibration based on 3–6 peptides equally distributed according to their elution time (with the exception of libraries LIB6 and LIB10 due to differences in elution times from different experiments). After calibration the XIC extraction window was narrowed to 5–12 min depending on the library. In LMWF we did not perform the retention time calibration and instead used a wider extraction window of 10 min from the beginning. In each library we manually selected peptides and transitions used for quantification according to their quality. Afterwards, we processed the data automatically and removed any remaining peptides that failed the evaluation in this step. Resulting data was exported to .xml files and normalized using total area sums (TAS) approach. We calculated coefficient of variation (CV%) values in each experiment (created library) for each protein in HMWF and each peptide in LMWF to assess the quality of our quantification experiments.

Data related to all qualitative and quantitative experiments were deposited to ProteomeXchange Consortium [25] via the PRIDE [26] partner repository with the data set identifier PXD008073.

3. Results and discussion

3.1. Qualitative measurements of High Molecular-Weight Fraction

We identified a total of 302 distinct proteins in all experiments on HMWF (see Supplementary Table 2). Numbers of proteins identified in each fractionation experiment are listed in Table 1. We found 85 proteins in unfractionated desalted samples. Using only immunodepletion based on MARS-14 commercial kit, we detected 19 proteins previously unidentified in the first experiment, however the total number of protein identifications in this experiment was lower (81) despite using fractionation (see Table 1 and Supplementary Table 2). This result is somewhat surprising, because we analyzed in a single database search both proteins unbound to MARS-14 resin and high abundant proteins which were retained and subsequently eluted from the resin (see Materials and Methods section). Such procedure should allow to analyze both high and low abundant proteins. However, some of the high abundant proteins (especially serum albumin) possess considerable protein binding capabilities, which may result in the deprival of lower abundant proteins from the fraction unbound to the MARS-14 resin. Therefore, such proteins are not present in sufficient quantities in either fraction resulting from immunodepletion, as their presence in the bound fraction is concealed by the vast amounts of high abundant proteins. Nonetheless, the quality of identifications improved with employment of fractionation. In the MARS-14 experiment there were only 3 single-peptide identifications compared to 17 single-peptide identifications in unfractionated samples (see Supplementary Table 2). The overall number of identifications in both experiments (104) is comparable with our previous study [14], where we identified 103 proteins of HMWF using the same sample preparation and fractionation scheme, as well as with other studies, where one-step fractionation procedures were utilized [5,6,10]. We reported 79 protein identifications overlapping with our previous study, which means that we found 25 new proteins and failed to detect 24 proteins identified previously. 11 of those 25 new proteins were identified only after immunodepletion. We consider those analyses of HMWF to be rather reproducible, and the difference in discovered protein sets can be explained mostly by two factors: (1) we used a different FDR cutoff in this study, and (2) the hFF fraction resulting from immunodepletion can be variable in composition, especially in proteins present at lower concentrations and/or proteins which are able bind to high abundant proteins, particularly albumins. It should be noted that in the case of research on the hFF proteome composition none of the so far published papers (even partially) touch upon the issue of the results reproducibility for identical or very similar experiment setups.

3.2. Effect of high pH RP-HPLC fractionation on qualitative measurements of HMWF

In total we identified nearly three times more proteins (296) in high pH RP-HPLC fractionation experiments than in the first two experiments without HPLC fractionation (104) in HMWF (see Fig. 1: left panel). Moreover, only 6 proteins reported in HMWF in this study have not been detected in any of the HPLC fractionation experiments and were only found in unfractionated HMWF samples. There were more proteins identified in all but one single fractionation experiments (HPLC1-HMW2) than in unfractionated HMWF samples. However, there were up to 34 single-peptide identifications in each fractionation experiment (aside from HPLC1-HMW2; see Supplementary Table 2), Nevertheless, the numbers of confident protein identifications proven by the detection of at least 2 tryptic peptides were higher for HPLC fractionation experiments in nearly all cases. Combination of immunodepletion with HPLC fractionation did not result in a substantially higher yield of identified proteins, as the numbers were roughly the same or even higher in experiments with only HPLC fractionation scheme. This can be explained by the fact that immunodepletion may cause not only elimination of targeted proteins from the sample but also proteins which can interact with those targeted proteins, thus reducing the sample's complexity, as it was mentioned in Qualitative Measurements of High Molecular-Weight Fraction paragraph. Because in our HPLC fractionation experiments we used only the immunodepleted hFF material as an analyzed sample, we could have lost some of the proteins present at lower concentration with the high abundant protein fraction. A few groups have previously used high pH RP-HPLC prefractionaction scheme in hFF proteome studies, including prior immunodepletion [11,12] or as a stand-alone technique [7,13]. Shen et al. reported 219 high-confidence proteins employing HPLC off-line HPLC fractionation in a pH of 10, and subsequently analyzing fractions in on-line low pH HPLC system coupled with MALDI TOF/TOF

MS detection [7]. Their results are comparable to the results of our experiments with the use of only HPLC at the pH value of 10 as a fractionation technique (HPLC2-HMW1, HPLC2-HMW2, HPLC2-HMW3, see Table 1). Zamah et al. [11] and Oh et al. [12] were able to report 742 and 1079 proteins, respectively, using immunodepletion followed by off-line HPLC fractionation and analysis by nano-LC coupled to an Orbitrap instrument. Considering that Oh et al. [12] reported isoforms of the same protein as distinct identifications, the number of unique protein identifications is similar to that presented in the work of Zamah et al [11]. Our list of reported protein identifications is still > 2 times shorter than in those studies. The reason for that might be differences in experimental setup, such as variety of fractionation gradients (composition of solvents, gradient steepness, as well as time of separation), use of different chromatographic columns, LC and mass spectrometers. We can speculate that the primary factor influencing the number of identifications is the use of different LC system. A microLC system such as utilized in this study allows for a shorter analysis time suited for clinical applications at the cost of sensitivity, resulting in a lower yield of identified proteins than with a widely used nanoLC system [21]. Nevertheless, in our study we still were able to found about 110 proteins not reported in the work of Zamah et al. [11], and about 90 proteins which were not reported in the work of Oh et al [12]. Identification of new proteins (not reported in previous studies) and large variation in the number of protein identification in different laboratories suggest that studies on proteome composition of hFF including standardization of procedures used in research are far from completion.

3.3. Qualitative measurements of Low Molecular-Weight Fraction

Our knowledge about hFF peptidome is very limited, and so far only two papers devoted to this subject have been published [14,27]. Chen and coworkers [27] managed to identify as many as 159 proteins in hFF peptidome. We identified a total of 161 proteins in experiments on LMWF (see Supplementary Table 3). Numbers of proteins detected in each fractionation experiment are listed in Table 1. In unfractionated samples of LMWF we found 28 proteins at established peptide identification criteria (confidence ≥99, FDR < 0.5%). Hoping to expand the initial unfractionated LMWF library, we analyzed three samples concentrated from 0.8 ml of hFF (concentration of peptides analyzed previously was increased 2-fold). However, among 17 proteins identified in that experiment we found only one additional protein. In our previous study [14], using a comparable sample processing scheme we were able to identify 91 proteins. In contrary to the results obtained for HMWF, we found a large discrepancy in the results obtained from experiments repeated in the same laboratory in nearly identical experimental setup. We suspect that such disproportions in our own results may be mainly due to the instability of hFF peptidome caused by the presence of proteases in stored material, which may diversify the LMWF by degrading more proteins or further degrading peptides to amino acids depending on the storage time and conditions. Moreover, due to interactions between peptides and proteins, ultrafiltration may not result in a complete division of these components by molecule size [28]. Therefore, the peptide composition of LMWF may differ in separate ultrafiltration experiments. In contrast to this part of the study, in our previous research we conducted a considerable number of analyses of LMWF samples prepared in a slightly different conditions (e.g., different desalting techniques) to obtain a proper spectral library, whereas at this stage we consider only 8 samples. Nonetheless, despite the difference in protein identification numbers, 14 out of 29 proteins identified in unfractionated samples were not found among 91 LMWF proteins reported previously. This finding further supports our hypothesis of the instability of hFF peptidome.

Table 1
Numbers of proteins identified in each experiment of this study. Total number states quantity of all proteins identified in each experiment, while unique number applies only to those protein identifications which were found in only one experiment (fractionation scheme). Uniqueness applies separately to proteins identified in High Molecular-Weight Fraction and Low Molecular-Weight Fraction.

Experiment	Fractionation scheme	Number of protein identifications	
		Total	Unique
High molecular-weight fraction			
Unfractionated HMW	None	85	6
MARS-14	Immunodepletion	81	0
HPLC1-HMW1	High pH RP-HPLC	127	13
HPLC1-HMW2	High pH RP-HPLC	74	0
HPLC1-MARS	Immunodepletion High pH RP-HPLC	128	17
HPLC2-HMW1	High pH RP-HPLC	204	23
HPLC2-HMW2	High pH RP-HPLC	192	20
HPLC2-HMW3	High pH RP-HPLC	141	3
HPLC2-MARS1	Immunodepletion High pH RP-HPLC	138	10
HPLC2-MARS2	Immunodepletion High pH RP-HPLC	97	6
Low molecular-weight fraction			
Unfractionated LMW	None	28	6
2× LMW 1–3	None	17	1
HPLC1-LMW1	High pH RP-HPLC	18	7
HPLC1-LMW2	High pH RP-HPLC	32	7
HPLC1-LMW3	High pH RP-HPLC	49	20
HPLC2-LMW1	High pH RP-HPLC	20	5
HPLC2-LMW2	High pH RP-HPLC	75	38
HPLC2-LMW3	High pH RP-HPLC	56	26

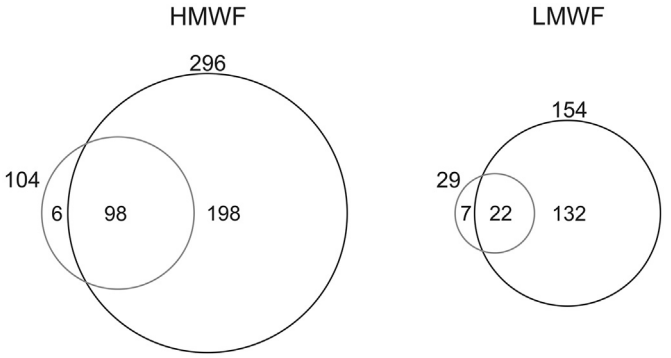


Fig. 1. Venn plots demonstrating distributions of distinct protein identifications reported in this study in experiments including or non-including HPLC fractionation. The left panel shows hFF proteins in HMWF identified in experiments involving high pH RP-HPLC fractionation (296 in total) or not involving this type of fractionation (104 in total). The right panel shows hFF proteins in LMWF identified in experiments involving high pH RP-HPLC fractionation (154 in total) or not involving this type of fractionation (29 in total).

3.4. Effect of high pH RP-HPLC fractionation on qualitative measurements of LMWF

We found 154 distinct proteins in HPLC fractionation experiments of LMWF of hFF. It is above five times more identifications than in unfractionated samples (see Fig. 1: left panel), and 63 more than in our previous study [14]. 7 proteins were identified only in unfractionated samples; this number seems small in comparison to all identified in fractionated samples proteins, yet it is about $^{1}/_{4}$ of all proteins found in unfractionated samples. This result can be explained by the very possible instability of hFF peptidome mentioned in earlier section Qualitative Measurements of Low Molecular-Weight Fraction. Most total and unique protein and peptide identifications were found in fractionation experiments on more concentrated LMWF fractions, resulting from ultrafiltration of 1.6 or 2.4 ml of hFF (see Table 1), as it was expected. Yet, unique identifications were found in all fractionation

experiments, what sustains the assumptions of peptidome instability, which may diversify samples in an unpredictable way. In all fractionation experiments we found a substantial percentage of single-peptide protein identifications, however in LMWF we consider them valid due to the nature of peptidome material.

3.5. Functional analysis of identified hFF proteins

The results of functional gene ontology analysis of proteins identified in both HMWF and LMWF are depicted in Supplementary information Figs. S1–S3. Leading molecular function classifications for both HMWF and LMWF are binding and catalytic activity, which is not surprising considering hFF's proximity to blood serum. A considerable percentage of especially LMWF proteins was also classified as having structural molecule activity and transporter activity. Major biological processes that proteins of both fractions are involved in are cellular and metabolic. As expected, most HMWF proteins that could be assigned to a cellular component are localized in extracellular region (81 proteins). However, many proteins originated also in cell part (48 proteins) or even organelle (18 proteins). It is unexpected, as processed hFF should not contain any cells; however, it is not a substantial part of all protein identifications. Presence of such proteins in HMWF can be explained by: (1) effect of sample processing method, which left cell debris in analyzed material, or (2) presence of the remains of natural cell death processes. In contrast, most of the LMWF proteins localize in cell parts, and then nearly equally often in macromolecular complexes, organelles, or extracellular region. However, this result is not as surprising due to the peptidomic nature of LMWF. Besides the two explanations for the presence of cell proteins in hFF, it also may be possible for certain peptides to be transported outside the cell. Moreover, it is more probable for a peptide to be present in hFF after a cell's breakdown (which could be found in the LMWF) than for a whole protein (which could be found in the HMWF). Most common protein classes for HMWF proteins are: enzyme modulator, defense/immunity protein, and hydrolase. In LMWF fraction most proteins were assigned to classes: enzyme modulator, nucleic acid binding, and cytoskeletal protein. Two latter classes were most probably assigned to proteins which localize inside cells. Most proteins in both fractions may be involved in blood coagulation pathway, which could be anticipated due to hFF's proximity to blood serum, however only up to 23 proteins (HMWF) were assigned to this category. The same number of LMWF proteins (10) can also take part in integrin signaling pathway.

3.6. Summary of qualitative analysis

We identified 400 distinct proteins in all experiments of this study. Most of those proteins were identified in the HMWF. More than $^{1}/_{3}$ of all proteins identified in LMWF were also found in HMWF (63; see Fig. 2), which is about $^{1}/_{5}$ of all proteins identified in HMWF. It is possible that not all proteins present in hFF undergo proteolysis at a measurable rate, what could cause their absence in the peptidomic pool of hFF. Moreover, we carried out more experiments on HMWF and used a more sophisticated fractionation technique. On the other hand, some proteins may be present at low concentrations in hFF, yet they may undergo proteolysis at a substantial rate, and therefore it could be easier to detect such proteins in the peptidomic fraction of hFF.

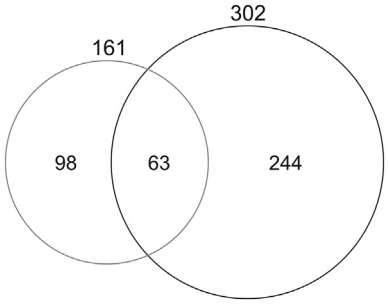


Fig. 2. Venn plot demonstrating whole hFF proteome identified in this study in HMWF (302 in total) or in LMWF (161 in total).

As expected, in this study we identified substantially more proteins than in our last study [14] without the use of fractionation. In total, in the present study we missed 59 protein identifications reported previously in both fractions. Most of those originate in LMWF, as in that fraction we detected only 33 identical protein identifications in this and previous study, meaning that we failed to identify 58 previously reported proteins and at the same time found as many as 128 new proteins. It further supports our hypothesis of instability of the peptidome fraction, described further in Qualitative Measurements of Low Molecular-Weight Fraction section. Despite the use of a more strict FDR filter (0.01%) for HMWF, only 18 of proteins reported in our previous research were not detected. Nonetheless, those results prove the efficiency of HPLC fractionation in the qualitative study of hFF proteome and peptidome.

We used two different chromatographic columns in conjunction with two buffer systems in HPLC fractionation with different pH values: 8 (marked HPLC1 in tables) and 10 (marked HPLC2 in tables). The exact composition of both buffer systems and specifications of chromatographic columns is described in Materials and Methods section and Supplementary Table 1. It seems that the second chromatographic setup with higher pH works more effectively in both HMWF and LMWF (see Table 1). Yet, considerable amounts of unique identifications were also found using the first buffer combination (pH 8), especially in the case of LMWF. Moreover, simple usage of divergent concentration gradients in our experiments also led to identification of unique proteins (e.g., compare HPLC2-HMW1 and HPLC2-HMW2 in Table 1). It demonstrates that, regardless of a number of studies on hFF proteomic composition, there is still a potential for new discoveries after performance of diverse fractionation schemes, particularly in usually neglected peptidome fraction.

3.7. SWATH measurements of High Molecular-Weight Fraction proteins

Few studies are devoted to quantitative measurements of hFF proteins [11,13,14,29] and among them only our previous work [14] used label-free SWATH quantification. Therefore, in the following part we will mainly compare obtained results with those presented in previous work [14]. In order to be able to correctly evaluate the influence of HPLC fractionation on the SWATH quantification of hFF proteins we built all our spectral libraries on the basis of analysis of samples not fractionated by HPLC: a primary library consisting of the search of only samples unfractionated by HPLC (LIB1), libraries containing data from one fractionated experiment and the primary library (LIB2-LIB9), and final library containing all HMWF data collected in this study (LIB10; see Table 2). In the primary library (LIB1) we were able to analyze 62 proteins by SWATH-MS (see Fig. 3). It is 10 proteins fewer than in our previous study [14], however in this case we used more strict processing parameters. We analyzed 10 new proteins, and failed to measure 20 proteins quantified before. Most of the proteins were measured by > 2 quantifiable peptides (above 60%, see Fig. 4: top panel). 17 proteins were quantified with CV% under 20%, and 37 proteins under 40% (see Supplementary Table 4). The quality of quantification of nearly 30% of all proteins analyzed in library LIB1 would be sufficient for a proper biomarker discovery study.

3.8. Effect of high pH RP-HPLC fractionation on the quality of HMWF spectral libraries

We were able to quantify 108 proteins across all created libraries. In single libraries we quantified from 39 (LIB6, LIB8) to 79 proteins (LIB3, LIB5; see Fig. 3). As expected, in most libraries we quantified more

proteins than in the primary library. We obtained the best results in terms of number of quantified total and unique proteins in libraries created from HPLC fractionation not preceded by immunodepletion (HPLC1-HMW2, HPLC2-HMW1). It was expected in the case of HPLC2-HMW1 experiment, in which we simply got the most protein identifications (204; see Table 1). However, this result is surprising in the case of HPLC1-HMW2 experiment, as we got the lowest number of protein identifications in that experiment (74), and we were able to measure more proteins quantitatively than we identified from the data experiment alone, including one protein which was not identified in any of the experiments (immunoglobulin kappa variable 2D-29; see Supplementary Table 4). It was probably possible to identify a new protein using the evidence from both unfractionated and HPLC fractionated experiments. An explanation for this phenomenon might be a good quality of HPLC1-HMW2 identifications, which could fit in the correct retention time window of the primary library data. Even though all libraries were based on the data from the primary library, in some of them we were able to quantify fewer proteins than in the original library, i.e., LIB6, LIB7, and LIB8. It is also unexpected as in those experiments (HPLC2-HMW2, HPLC2-HMW3, and HPLC2-MARS1) we were able to identify numbers of proteins second only to previously mentioned HPLC2-HMW1 experiment (192, 141, and 138, respectively; see Table 1). We can only assume that the raw number of protein identifications, even of good quality (FDR < 0.01%) is not a main factor affecting the quality of spectral libraries. Other factors that may also play a role in this process are the data quality, such as transitions intensity, narrowness of resulting peaks, integrity of retention times with the original library, and lack of overlapping transition retention times of different peptides. Further proof of this statement is the fact that the number of quantified proteins in the library consisting of all obtained HMWF data is only 71 (see Fig. 3), even though the number of identifications is obviously the highest in this search. Percentage of proteins identified using 5–6 peptides decreased from the level of the primary library in every secondary library, despite the availability of higher amount of measurable peptides (see Fig. 4: top panel). However, in almost all cases the number of proteins quantified by > 2 peptides was higher than in library LIB1, except for libraries LIB6 and LIB8, previously mentioned as libraries with the lowest numbers of quantified proteins (see Supplementary Table 4). Libraries LIB6, LIB9, and LIB2 are characterized by the highest percentage of 2-peptide protein measurements (above 40%), while libraries LIB4, LIB3, and LIB7 have the smallest percentage of such measurements (slightly above 30%; see Fig. 4: top panel). In terms of quantification quality, only libraries LIB3 and LIB9 have the same number of proteins quantified at CV% < 20% (17; see Supplementary Table 4), however the percentage of those proteins remains highest in the library LIB1, with following libraries LIB3 and LIB9 (see Fig. 4: bottom panel). Library LIB3 also has the smallest fraction of proteins quantified at CV > 60%. Surprisingly, the highest percentage of proteins quantified at CV > 60% was established in the case of library LIB10. Apparently, data on many available peptides and their transitions from different experiments may result in differing retention times affecting variation of quantification across samples.

3.9. SWATH measurements of Low Molecular-Weight Fraction peptides

As in the case of HMWF (see SWATH Measurements of High Molecular-Weight Fraction Proteins paragraph), in LMWF we also based all spectral libraries (LIB2-LIB8) on the primary library (LIB1, see Table 2). We quantified 38 peptides derived from 14 distinct proteins in LIB1 (see Fig. 5). In our previous study we were able to quantify 43 peptides from 23 proteins [14]. 4 peptides were quantified in both studies: SSKITHRIHWESASLLR (complement C3), KVPQVSTPTLVEVSR (serum albumin), DDPDAPLQPVTPLQLFEGRRN (complement C4-B), and AHYDLRHTFMGVVSLGSPSGEVSHPRKT (alpha-2-HS-glycoprotein). Peptides derived from 7 proteins quantified in library LIB1 were not quantified previously. Despite the fact that we qualified one peptide per protein for quantification in this fraction, in the primary library above 20% of proteins had 4–6 quantified peptides, and about 50% 2–3 quantified peptides (see Fig. 6: top panel). CV% values for LMWF peptides were generally substantially higher than in the case of HMWF proteins. Only one peptide was quantified with CV < 50% in library LIB1, and above 40% of peptides were quantified at CV% > 90% (see Fig. 6: bottom panel). In our opinion such high values of coefficients of variation are caused by the instability of LMWF described further in the Qualitative Measurements of Low Molecular-Weight Fraction paragraph. Even though the samples for SWATH measurements were prepared and analyzed in sequence at roughly the same time, the variation of the measurements might still be substantial. Therefore, we believe that quantitative measurements of LMWF of hFF cannot be a reliable source of biomarkers of oocyte quality without further studies on the distinct peptides origin in the material (source, nature of protein cleavage or breakdown processes, and means of transport to hFF).

Table 2
Construction of spectral libraries for SWATH analyses in both hFF fractions. Data from marked experiments were used in database search leading to preparation of a given library.

Experiment	LIB1	LIB2	LIB3	LIB4	LIB5	LIB6	LIB7	LIB8	LIB9	LIB10
High molecular-weight fraction										
Unfractionated HMW	X	X	X	X	X	X	X	X	X	X
MARS-14	X	X	X	X	X	X	X	X	X	X
HPLC1-HMW1		X								X
HPLC1-HMW2			X							X
HPLC1-MARS				X						X
HPLC2-HMW1					X					X
HPLC2-HMW2						X				X
HPLC2-HMW3							X			X
HPLC2-MARS1								X		X
HPLC2-MARS2									X	X
Low molecular-weight fraction										
Unfractionated LMW	X	X	X	X	X	X	X	X		
2× LMW 1–3	X	X	X	X	X	X	X	X		
HPLC1-LMW1		X						X		
HPLC1-LMW2			X					X		
HPLC1-LMW3				X				X		
HPLC2-LMW1					X			X		
HPLC2-LMW2						X		X		
HPLC2-LMW3							X	X		

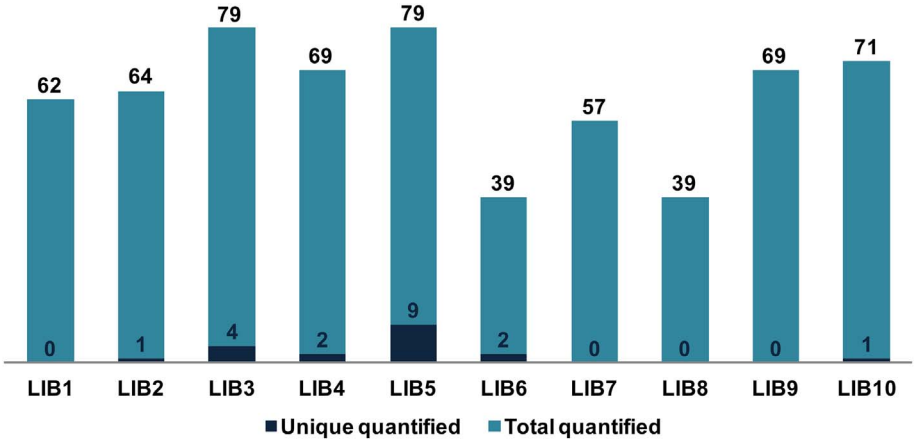


Fig. 3. Numbers of hFF proteins quantified by SWATH analysis conducted using each HMWF spectral library: total is a number of all quantified proteins, while unique is a number of proteins quantified only using the specific library.

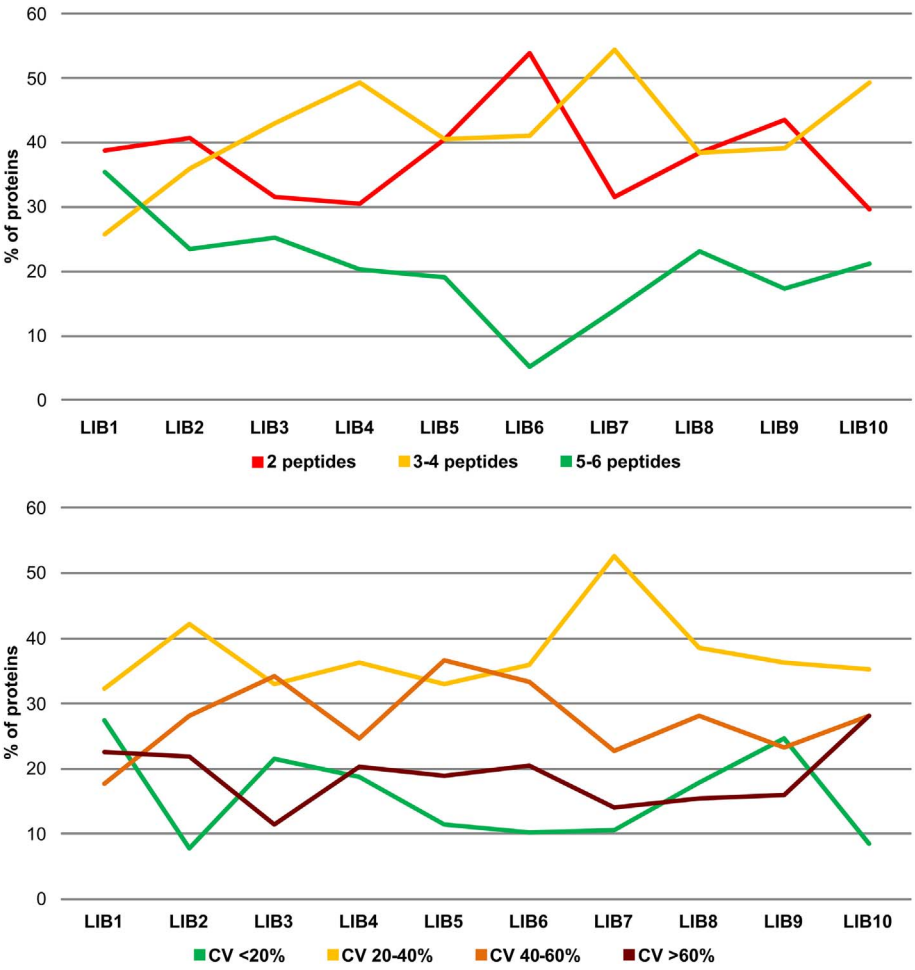


Fig. 4. Quality assessment of hFF protein quantification in HMWF spectral libraries. Top panel: percentage distribution of proteins quantified by different numbers of distinct peptides (2, 3–4, 5–6) in tested HMWF libraries. Bottom panel: Percentage distribution of proteins quantified with different CV% values (< 20%, 20–40%, 40–60%, > 60%) in tested HMWF libraries.

3.10. Effect of high pH RP-HPLC fractionation on the quality of LMWF spectral libraries

We quantified 250 peptides deriving from 84 distinct proteins across all LMWF libraries. In all libraries constructed with the use of HPLC experiments (LIB2-LIB8) we were able to quantify more peptides than in the primary library (see Fig. 5). The numbers of quantified peptides in a single library varied from 54 originating from 20 proteins (LIB5) to 106 originating from 53 proteins (LIB8). In contrast to HMWF, in the case of LMWF we obtained the best results in terms of the number of quantified peptides in the library consisting of all data collected from LMWF experiments. The reason for this difference between hFF fractions might be the fact that in the LMWF we focused on peptides, therefore we allowed single-peptide quantification of proteins, whereas in HMWF such measurements were excluded. We obtained the lowest number of quantified peptides from libraries built on fractionation of

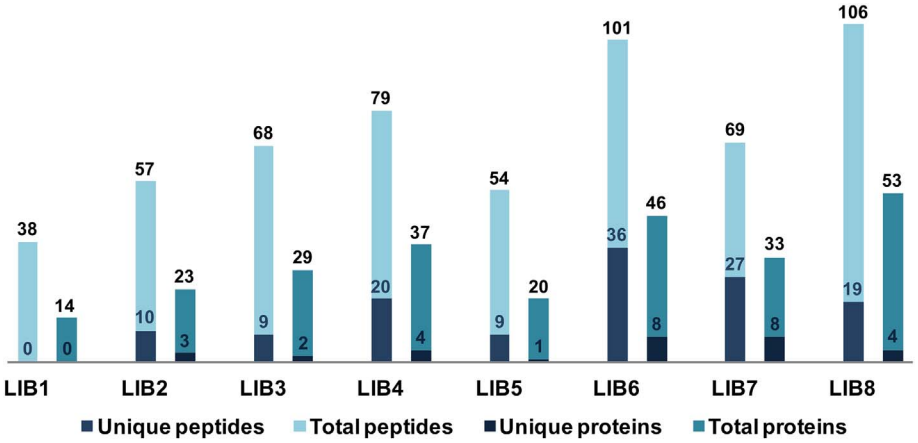


Fig. 5. Numbers of hFF peptides (bars to the left) and proteins (bars to the right) quantified by SWATH analysis conducted using each LMWF spectral library: total is a number of all quantified peptides or proteins, while unique is a number of peptides or proteins quantified only using the specific library.

material obtained from ultrafiltration of smaller volumes of hFF (LIB2–0.8 ml, LIB5–1.2 ml), and the numbers rose with the increasing volume of starting material used with the exception of unusually rich library LIB6, which was also characterized by the highest number of distinct protein identifications (see Table 1). In all secondary libraries we were able to quantify unique peptides, the most in libraries LIB6 and LIB7. We also quantified peptides originating from 6 proteins which were not identified in any of the experiments (see Supplementary Table 5). As we mentioned in the Effect of High pH RP-HPLC Fractionation on the Quality of HMWF Spectral Libraries paragraph, combination of data from the primary library and the fractionation experiments could have enabled gathering of enough spectral evidence for a protein identification, especially in the case of such variable material as LMWF of hFF. All secondary libraries have a higher percentage of proteins with one quantified peptide with the highest values in the case of libraries LIB7 and LIB8 (see Fig. 6: top panel). It is not

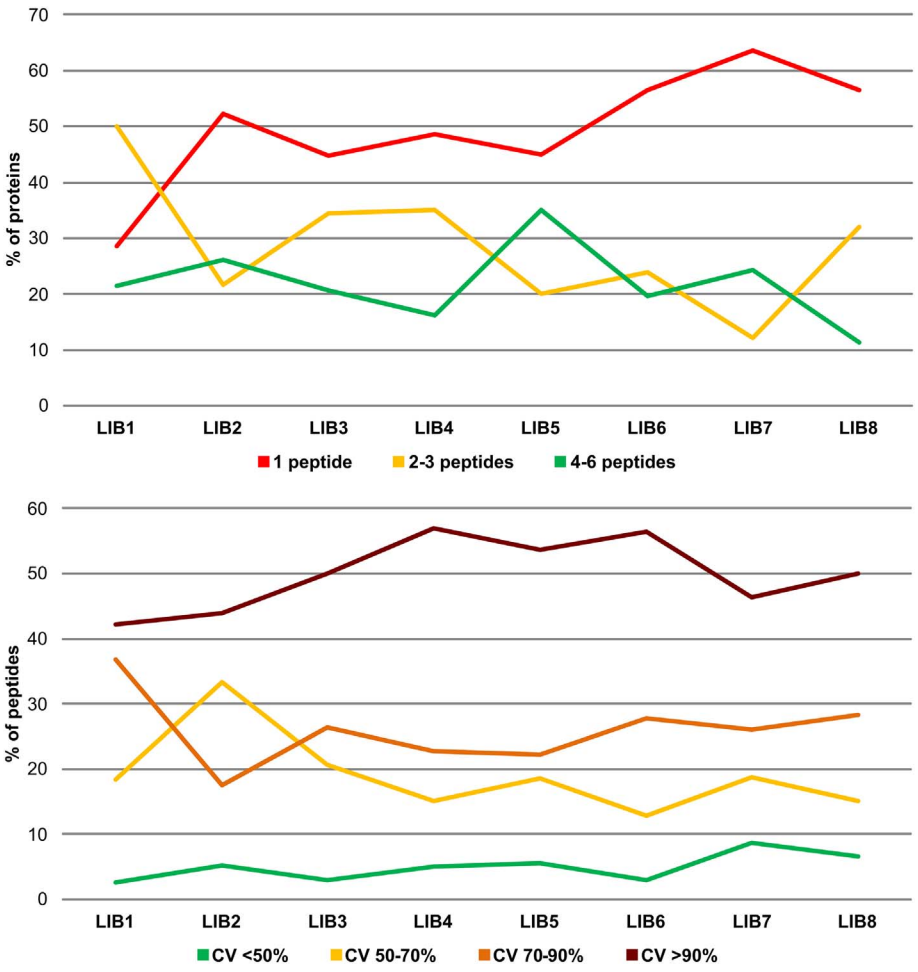


Fig. 6. Quality assessment of hFF peptide and protein quantification in LMWF spectral libraries. Top panel: percentage distribution of proteins represented by different numbers of distinct quantified peptides (1, 2–3, 4–6) in tested HMWF libraries. Bottom panel: Percentage distribution of peptides quantified with different CV% values (< 50%, 50–70%, 70–90%, > 90%) in tested LMWF libraries.

surprising, as it was easier to quantify additional proteins in secondary libraries (considering higher number of peptides quantified/protein originators) based on one peptide, especially in the case of one-peptide identifications, what was impossible in the case of HMWF. The percentage of proteins quantified by 5–6 proteins is higher in libraries built on samples of less volume (LIB2 – HPLC1-LMW1, LIB5 – HPLC2-LMW1). In both of those experiments, numbers of protein identifications were lower than in the unfractionated samples (see Table 1), hence many of the quantified proteins could overlap with proteins in the primary library sufficiently enough to generate more quantifiable peptides of a single protein. As we stated in the SWATH Measurements of Low Molecular-Weight Fraction Peptides paragraph, the CV% values for peptides quantified in LMWF did not reach a level lower than 20% in any case. In all libraries only < 10% of peptides were characterized by CV% < 50%, and above 50% of peptides reached CV% values above 90% (see Fig. 6: bottom panel). Most peptides were quantified at CV < 50% in libraries LIB7 and LIB8, however even in those libraries the numbers of such peptides are 6 and 7, respectively.

3.11. Comparison of SWATH quantification of HMWF proteins and LMWF peptides employing libraries built with and without the use of high pH RP-HPLC fractionation scheme

Fig. 7 presents, in a simplified manner, how the quantitative results from the primary library were affected by the addition of data from the HPLC fractionation experiments. Quantitative measurements of all proteins in HMWF and peptides in LMWF were compared to their respective measurements conducted with the use of one of the primary libraries.

In HMWF, the best, in terms of improved quantification, were libraries LIB3, LIB9, and LIB4. LIB3 and LIB5 were already mentioned as libraries with most quantified proteins (see Fig. 3). The highest numbers of proteins absent in library LIB1 were also quantified in libraries LIB3 and LIB5. Worse quantification was noted for the highest number of proteins in libraries LIB10, followed by LIB2 and LIB5. As we discussed in the Effect of High pH RP-HPLC Fractionation on the Quality of HMWF Spectral Libraries paragraph, a greater amount of data present in library LIB10 might have affected the quality of quantification in a negative way. Most proteins which were quantified in library LIB1 were absent in libraries LIB6 and LIB8, where the total numbers of quantified proteins were lowest (see Fig. 3). Taking all those factors into account, LIB3 and LIB9 can be indicated as libraries with the most improved overall quantification, and LIB6 and LIB8 as libraries with the most worsened quantification. General balance of factors in other libraries points to a quality similar to that of the primary library.

It is quite surprising that libraries LIB8 and LIB9 differ substantially in quantification quality, despite the fact that virtually the only difference between the experiments on which they were based (HPLC2-MARS1 and HPLC2-MARS2) was the chromatographic gradient applied during the HPLC fractionation (see Supplementary Table 1). The best libraries (LIB3, LIB9) were constructed on the basis of experiments utilizing different columns, chromatographic gradients and starting material in terms of prior immunodepletion (HPLC1-HMW2, and HPLC2-MARS2, respectively), and one similarity of those experiments is that they resulted in the lowest yield of protein identifications among HPLC fractionated data sets (74 and 97, respectively; see Table 1). It appears that this may be caused by differences in peptide transition retention times or abundance of peptides eluting in the same retention

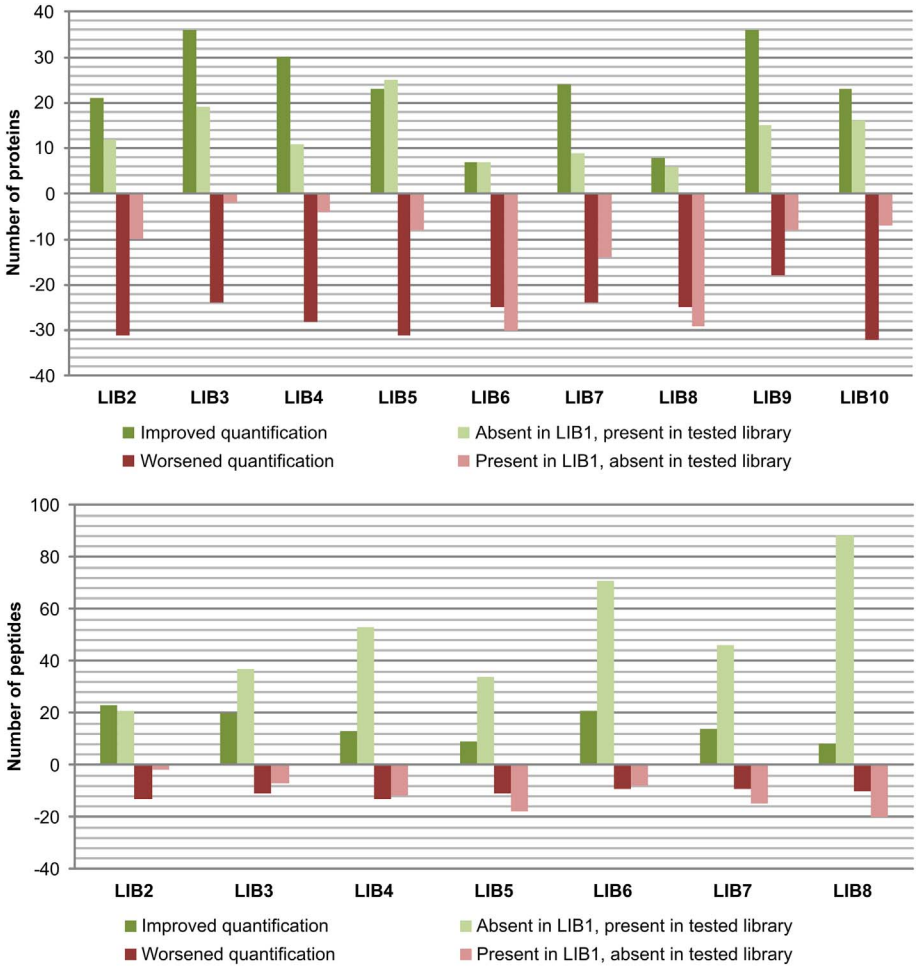


Fig. 7. Comparison of the SWATH quantification quality performed with the use of proteins (top panel, HMWF) and peptides (bottom panel, LMWF) spectral libraries built from high pH RP-HPLC fractionation experiments MS/MS spectra (top panel: libraries LIB2-LIB10, bottom panel: libraries LIB2-LIB8) to the primary library (LIB1) built without the application of this type of fractionation. The evaluation of quality was based on the differences of coefficient of variation among the libraries (lower CV% than in library LIB1 resulted in improved quantification, and higher CV% than in library LIB1 was treated as worsened quantification. Lighter bars represent either proteins/peptides newly quantified in relation to library LIB1 (in green) or proteins/peptides quantified in LIB1, but non-quantifiable using analyzed library (in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

time interval resulting in overlapping data. In case of less rich libraries (LIB3, LIB9), data processing was more efficient leading to improved outcomes. In summary, it seems that the involvement of data collected from the HPLC fractionation experiments usually affected the quality of quantification in a rather slightly negative manner, while generally keeping the number of quantified proteins at a comparable level (with a few already mentioned exceptions). Perhaps, to be able to fully utilize the potential of libraries most abundant in proteins, a distinct SWATH-MS method should be developed, e.g. employing variable window acquisition to avoid such extensive data overlapping. This approach was recently proposed in several studies [30–32].

There is a considerable difference between data obtained from the LMWF experiments versus the HMWF experiments. It is apparent at a first glance that the involvement of a substantial number of peptides absent in library LIB1 in each secondary library outweighs all other factors of the comparison pointing to the advantage of secondary libraries. Most such peptides were quantified in libraries LIB8 and LIB6, mentioned before due to the highest numbers of quantified peptides overall (see Fig. 5). Improved quantification was noted mostly for peptides quantified in libraries LIB2, LIB6, and LIB3. The highest numbers of peptides with worsened quantification characterize libraries LIB2 and LIB4, and the highest numbers of peptides quantified in library LIB1 were missing from libraries LIB8 and LIB5. In general, mostly due to the presence of newly quantified peptides, all secondary libraries seem to have improved quantification capabilities in comparison to the primary library (LIB1). The best libraries were built on database searches with the highest numbers of protein identifications: LIB6, and combined LIB8, whereas the worst results were obtained in the case of LIB5 and LIB2, both based on experiments with the lowest protein identification yield. It seems that in general, fractionation did not meaningfully influence an already bad quality of peptide quantification in LMWF, however it managed to considerably increase the number of quantifiable peptides. Still, to be able to correctly interpret quantitative peptidomic hFF data, more information on peptide origin as well as established sample preparation procedures for distinct peptide recovery are necessary.

4. Conclusions

Previous studies on hFF proteome showed that multi-step fractionation [8,9,11,12] procedure leads to identification of a large number (up to 800) of proteins in a single experiment. By employing prefractionation consisting of ultrafiltration, optional immunodepletion, and high pH RP-HPLC separation we were able to identify about 300 proteins (HMWF), as compared to our previous studies with less sophisticated fractionation protocol, where only about one hundred identifications were obtained [14]. Not only the number of identified proteins has increased, we also noticed that statistical significance and data quality rose substantially when high pH RP-HPLC separation was employed. Employing high pH RP-HPLC separation also raised the number of proteins identified as peptide fragments in peptidome (LMWF) of hFF. We were able to identify about 154 proteins in LMW fraction using a multistep fractionation protocol as compared to 91 identification where simpler fractionation protocol was employed [14]. Moreover, during our work we found much evidence that the peptidome content of hFF samples (LMWF) is very variable. Such observation could be connected with the presence of active proteolytic enzymes in the sample, which may create a very dynamic peptidome content leading to sample instability over time.

Employing high pH RP-HPLC prefractionation allowed us to extend knowledge about hFF proteome (identification of several proteins never reported before). We found out that expanding spectra libraries to some extent allowed to quantify more proteins, however there are some limitations. Building spectral libraries based on high confidence identifications of proteins does not indicate their quantitative (spectral libraries) quality. We observed that the quantifying capabilities of libraries rich in peptide and protein identifications were inferior to libraries with lowest numbers of identifications. Addition of low amounts of MS data provided by HPLC fractionation experiments with lower identifications numbers allowed us to build spectral libraries with best quantification potential in this study. Increasing MS data input above this limit resulted in data overlapping and worsened quantification quality. In our opinion, this issue could be resolved in the future by the development and optimization of specialized SWATH-MS acquisition method, e.g., recently proposed use of variable isolation windows [30–32].

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jprot.2018.03.010.

Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no 316094. The scientific work was financed from funds for science in the years 2013-2016 allocated for the internationally co-financed project. Computational resources used in this project were provided by the Informatics Center of the Metropolitan Academic Network (IC MAN-TASK) in Gdańsk.

References

[1] H.-C. Lee, S.-W. Lee, K.W. Lee, S.-W. Lee, K.-Y. Cha, K.H. Kim, S. Lee, Identification of new proteins in follicular fluid from mature human follicles by direct sample rehydration method of two-dimensional polyacrylamide gel electrophoresis, J. Korean Med. Sci. 20 (2005) 456, http://dx.doi.org/10.3346/jkms.2005.20.3.456.
[2] G.F. Erickson, S. Shimasaki, The physiology of folliculogenesis: the role of novel growth factors, Fertil. Steril. 76 (2001) 943–949 http://www.ncbi.nlm.nih.gov/pubmed/11704115 , Accessed date: 23 October 2017.
[3] C. Freitas, A.C. Neto, L. Matos, E. Silva, Â. Ribeiro, J.L. Silva-Carvalho, H. Almeida, Follicular fluid redox involvement for ovarian follicle growth, J. Ovarian Res. 10 (2017) 44, , http://dx.doi.org/10.1186/s13048-017-0342-3.
[4] A. Revelli, L. Delle Piane, S. Casano, E. Molinari, M. Massobrio, P. Rinaudo, Follicular fluid content and oocyte quality: from single biochemical markers to metabolomics, Reprod. Biol. Endocrinol. 7 (2009) 40, http://dx.doi.org/10.1186/1477-7827-7-40.
[5] M.M. Kushnir, T. Naessén, K. Wanggren, A.L. Rockwood, D.K. Crockett, J. Bergquist, Protein and steroid profiles in follicular fluid after ovarian hyperstimulation as potential biomarkers of IVF outcome, J. Proteome Res. 11 (2012) 5090–5100, http://dx.doi.org/10.1021/pr300535g.
[6] L. Bianchi, A. Gagliardi, G. Campanella, C. Landi, A. Capaldo, A. Carleo, A. Armini, V. De Leo, P. Piomboni, R. Focarelli, L. Bini, A methodological and functional proteomic approach of human follicular fluid en route for oocyte quality evaluation, J. Proteome 90 (2013) 61–76, http://dx.doi.org/10.1016/j.jprot.2013.02.025.
[7] X. Shen, X. Liu, P. Zhu, Y. Zhang, J. Wang, Y. Wang, W. Wang, J. Liu, N. Li, F. Liu, Proteomic analysis of human follicular fluid associated with successful in vitro fertilization, Reprod. Biol. Endocrinol. 15 (2017) 58, , http://dx.doi.org/10.1186/s12958-017-0277-y.
[8] J. Twigt, R.P. Steegers-Theunissen, K. Bezstarosti, J.A.A. Demmers, Proteomic analysis of the microenvironment of developing oocytes, Proteomics 12 (2012) 1463–1471, http://dx.doi.org/10.1002/pmic.201100240.
[9] A.S. Ambekar, R.S. Nirujogi, S.M. Srikanth, S. Chavan, D.S. Kelkar, I. Hinduja, K. Zaveri, T.S.K. Prasad, H.C. Harsha, A. Pandey, S. Mukherjee, Proteomic analysis of human follicular fluid: a new perspective towards understanding folliculogenesis, J. Proteome 87 (2013) 68–77, http://dx.doi.org/10.1016/j.jprot.2013.05.017.
[10] J. Hanrieder, A. Zuberovic, J. Bergquist, Surface modified capillary electrophoresis combined with in solution isoelectric focusing and MALDI-TOF/TOF MS: a gel-free multidimensional electrophoresis approach for proteomic profiling-exemplified on human follicular fluid, J. Chromatogr. A 1216 (2009) 3621–3628, http://dx.doi.org/10.1016/j.chroma.2008.12.026.
[11] A.M. Zamah, M.E. Hassis, M.E. Albertolle, K.E. Williams, Proteomic analysis of human follicular fluid from fertile women, Clin. Proteomics 12 (2015) 1–12, http://dx.doi.org/10.1186/s12014-015-9077-6.
[12] J.W. Oh, S.K. Kim, K.-C. Cho, M.-S. Kim, C.S. Suh, J.R. Lee, K.P. Kim, Proteomic analysis of human follicular fluid in poor ovarian responders during in vitro fertilization, Proteomics 17 (2017) 1600333, , http://dx.doi.org/10.1002/pmic.201600333.
[13] T. Regiani, F.B. Cordeiro, L. do V.T. da Costa, J. Salgueiro, K. Cardozo, V.M. Carvalho, K.J. Perkel, D.S. Zylbersztejn, A.P. Cedenho, E.G. Lo Turco, Follicular fluid alterations in endometriosis: label-free proteomics by MS(E) as a functional tool for endometriosis, Syst Biol Reprod Med 61 (2015) 263–276, http://dx.doi.org/10.3109/19396368.2015.1037025.
[14] A.E. Lewandowska, K. Macur, P. Czaplewska, J. Liss, K. Łukaszuk, S. Ołdziej, Qualitative and quantitative analysis of proteome and peptidome of human follicular fluid using multiple samples from single donor with LC-MS and SWATH

methodology, J. Proteome Res. 16 (2017) 3053–3067, http://dx.doi.org/10.1021/acs.jproteome.7b00366.
[15] S. Qin, A.S. Ferdinand, J.P. Richie, M.P. O'Leary, S.C. Mok, B.C.-S. Liu, Chromatofocusing fractionation and two-dimensional difference gel electrophoresis for low abundance serum proteins, Proteomics 5 (2005) 3183–3192, http://dx.doi.org/10.1002/pmic.200401137.
[16] J.F. Timms, E. Arslan-Low, M. Kabir, J. Worthington, S. Camuzeaux, J. Sinclair, J. Szaub, B. Afrough, V.N. Podust, E.-O. Fourkala, M. Cubizolles, F. Kronenberg, E.T. Fung, A. Gentry-Maharaj, U. Menon, I. Jacobs, Discovery of serum biomarkers of ovarian cancer using complementary proteomic profiling strategies, Proteomics Clin. Appl. 8 (2014) 982–993, http://dx.doi.org/10.1002/prca.201400063.
[17] M. Pernemalm, J. Lehtiö, Mass spectrometry-based plasma proteomics: state of the art and future outlook, Expert Rev. Proteomics 11 (2014) 431–448, http://dx.doi.org/10.1586/14789450.2014.901157.
[18] J.A. Dowell, D.C. Frost, J. Zhang, L. Li, Comparison of two-dimensional fractionation techniques for shotgun proteomics, Anal. Chem. 80 (2008) 6715–6723, http://dx.doi.org/10.1021/ac8007994.
[19] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol. Cell. Proteomics 11 (2012) (O111.016717-O111.016717), https://doi.org/10.1074/mcp.O111.016717.
[20] O.T. Schubert, L.C. Gillet, B.C. Collins, P. Navarro, G. Rosenberger, W.E. Wolski, H. Lam, D. Amodei, P. Mallick, B. MacLean, R. Aebersold, Building high-quality assay libraries for targeted analysis of SWATH MS data, Nat. Protoc. 10 (2015) 426–441, http://dx.doi.org/10.1038/nprot.2015.015.
[21] M. Zhang, B. An, Y. Qu, S. Shen, W. Fu, Y.-J. Chen, X. Wang, R. Young, J.M. Canty, J.P. Balthasar, K. Murphy, D. Bhattacharyya, J. Josephs, L. Ferrari, S. Zhou, S. Bansal, F. Vazvaei, J. Qu, Sensitive, high-throughput, and robust trapping-micro-LC-MS strategy for the quantification of biomarkers and antibody biotherapeutics, Anal. Chem. (2018), http://dx.doi.org/10.1021/acs.analchem.7b03949.
[22] R.L. Gundry, M.Y. White, C.I. Murray, L.A. Kane, Q. Fu, B.A. Stanley, J.E. Van Eyk, Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow, Curr. Protoc. Mol. Biol. (2009), http://dx.doi.org/10.1002/0471142727.mb1025s88.
[23] P.D. Thomas, M.J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, A. Narechania, PANTHER: a library of protein families and subfamilies indexed by function, Genome Res. 13 (2003) 2129–2141, http://dx.doi.org/10.1101/GR.772403.
[24] H. Mi, Q. Dong, A. Muruganujan, P. Gaudet, S. Lewis, P.D. Thomas, PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the gene ontology consortium, Nucleic Acids Res. 38 (2010) D204–10, http://dx.doi.org/10.1093/nar/gkp1019.
[25] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, 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-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. 32 (2014) 223–226, http://dx.doi.org/10.1038/nbt.2839.
[26] J.A. Vizcaíno, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools, Nucleic Acids Res. 44 (2016) D447–D456, http://dx.doi.org/10.1093/nar/gkv1145.
[27] F. Chen, C. Spiessens, T. D'Hooghe, K. Peeraer, S. Carpentier, Follicular fluid biomarkers for human in vitro fertilization outcome: proof of principle, Proteome Sci. 14 (2016) 17, http://dx.doi.org/10.1186/s12953-016-0106-9.
[28] R.S. Tirumalai, K.C. Chsan, D.A. Prieto, H.J. Issaq, T.P. Conrads, T.D. Veenstra, Characterization of the low molecular weight human serum proteome, Mol. Cell. Proteomics 2 (2003) 1096–1103, http://dx.doi.org/10.1074/mcp.M300031-MCP200.
[29] A. Iwase Bayasula, H. Kobayashi, M. Goto, T. Nakahara, T. Nakamura, M. Kondo, Y. Nagatomo, T. Kotani, F. Kikkawa, A proteomic analysis of human follicular fluid: comparison between fertilized oocytes and non-fertilized oocytes in the same patient, J. Assist. Reprod. Genet. 30 (2013) 1231–1238, http://dx.doi.org/10.1007/s10815-013-0004-3.
[30] Y. Zhang, A. Bilbao, T. Bruderer, J. Luban, C. Strambio-De-Castillia, F. Lisacek, G. Hopfgartner, E. Varesio, The use of variable Q1 isolation windows improves selectivity in LC–SWATH–MS acquisition, J. Proteome Res. 14 (2015) 4359–4371, http://dx.doi.org/10.1021/acs.jproteome.5b00543.
[31] S. Li, Q. Cao, W. Xiao, Y. Guo, Y. Yang, X. Duan, W. Shui, Optimization of acquisition and data-processing parameters for improved proteomic quantification by sequential window Acquisition of all Theoretical Fragment ion Mass Spectrometry, J. Proteome Res. 16 (2017) 738–747, http://dx.doi.org/10.1021/acs.jproteome.6b00767.
[32] B. Schilling, B.W. Gibson, C.L. Hunter, Generation of high-quality SWATH® acquisition data for label-free quantitative proteomics studies using TripleTOF® mass spectrometers, Methods Mol. Biol, 2017, pp. 223–233, , http://dx.doi.org/10.1007/978-1-4939-6747-6_16.

International Journal of
Molecular Sciences

MDPI

Article

Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid

Aleksandra E. Lewandowska [1,*], Anna Fel [1], Marcel Thiel [1], Paulina Czaplewska [1], Krzysztof Łukaszuk [2,3], Jacek R. Wiśniewski [4] and Stanisław Ołdziej [1,*]

[1] Intercollegiate Faculty of Biotechnology UG&MUG, University of Gdańsk, Abrahama 58, 80-307 Gdańsk, Poland; anna.fel@phdstud.ug.edu.pl (A.F.); marcel.thiel@ug.edu.pl (M.T.); paulina.czaplewska@ug.edu.pl (P.C.)
[2] INVICTA Fertility and Reproductive Center, Polna 64, 81-740 Sopot, Poland; luka@gumed.edu.pl
[3] Department of Obstetrics and Gynecological Nursing, Faculty of Health Sciences, Medical University of Gdańsk, Dębinki 7, 80-211 Gdańsk, Poland
[4] Department of Proteomics and Signal Transduction, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany; jwisniew@biochem.mpg.de
* Correspondence: aleksandra.lewandowska@ug.edu.pl (A.E.L.); stanislaw.oldziej@ug.edu.pl (S.O.)

Abstract: We present two separate label-free quantitative workflows based on different high-resolution mass spectrometers and LC setups, which are termed after the utilized instrument: Quad-Orbitrap (nano-LC) and Triple Quad-TOF (micro-LC) and their directed adaptation toward the analysis of human follicular fluid proteome. We identified about 1000 proteins in each distinct workflow using various sample preparation methods. With assistance of the Total Protein Approach, we were able to obtain absolute protein concentrations for each workflow. In a pilot study of twenty samples linked to diverse oocyte quality status from four donors, 455 and 215 proteins were quantified by the Quad-Orbitrap and Triple Quad-TOF workflows, respectively. The concentration values obtained from both workflows correlated to a significant degree. We found reasonable agreement of both workflows in protein fold changes between tested groups, resulting in unified lists of 20 and 22 proteins linked to oocyte maturity and blastocyst development, respectively. The Quad-Orbitrap workflow was best suited for an in-depth analysis without the need of extensive fractionation, especially of low abundant proteome, whereas the Triple Quad-TOF workflow allowed a more robust approach with a greater potential to increase in effectiveness with the growing number of analyzed samples after the initial effort of building a comprehensive spectral library.

Keywords: LC-MS/MS; Total Protein Approach; SWATH-MS; human follicular fluid; proteome; oocyte quality control; oocyte maturity; blastocyst development

Citation: Lewandowska, A.E.; Fel, A.; Thiel, M.; Czaplewska, P.; Łukaszuk, K.; Wiśniewski, J.R.; Ołdziej, S. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid. *Int. J. Mol. Sci.* **2021**, *22*, 7415. https://doi.org/10.3390/ijms22147415

Academic Editor: Laszlo Prokai

Received: 17 June 2021
Accepted: 8 July 2021
Published: 10 July 2021

1. Introduction

The accurate evaluation of the oocyte quality remains one of the main challenges in reproductive sciences. Although the proteomic analysis of a single oocyte cell has been recently accomplished [1], the oocyte cannot simultaneously be subjected to invasive analysis and maintain its conceptive potential. A consequence of those issues is a popular interest in the proximate environment of the developing oocyte, i.e., granulosa (GC) and cumulus cells or follicular fluid (FF) [2,3]. Human follicular fluid (hFF) from a single follicle is harvested along with the oocyte in the ovarian puncture procedure during in vitro fertilization (IVF) treatment. Therefore, it constitutes an easily obtainable material for the analysis of the processes occurring around the oocyte at the time of sample collection and has been appointed as a candidate for oocyte quality biomarker research [2,3]. Most of studies on hFF focused on utilizing diverse methodological strategies to characterize its proteome composition. Due to the proximity of hFF to human serum, its proteomic

analysis faces the same obstacles related to high diversity in dynamic protein concentrations. The use of distinct sample preparation workflows including fractionation schemes facilitated the identification of the most comprehensive sets of hFF proteins containing from few hundreds to 2461 proteins (analysis of human small antral follicle fluid—hSAF by Pla et al.) [4–10]. Quantitative hFF studies often concern analysis of pooled samples (coming from multiple follicles from one or more donors) [11] or single dominant follicles of each donor [6,12], yet it is essential that the oocyte quality studies (as methods and techniques used) must focus on multiple hFF samples related to individual oocytes from single donors [10,13,14]. Investigation of samples related to individual oocytes increases the number of measurements that should be performed and not all techniques or methods utilized so far in hFF proteome studies are well suited for large-scale clinical studies due to the cost of analysis, time and labor burden, or other factors. In order to achieve an in-depth analysis of hFF proteomic landscape and to propose an optimal methodology for the identification of proteins related to individual oocyte quality in a large-scale clinical study (involving many samples), we employed diversified sample preparation schemes to develop optimized quantitative protocols utilizing two independent instrument setups including the following high-resolution mass spectrometers: Q Exactive HF-X coupled with nano-LC and TripleTOF 5600+ coupled with micro-LC. The Total Protein Approach (TPA) [15] was applied for data registered on both instruments to achieve absolute protein concentrations in label-free workflows. The data-independent acquisition (DIA) method, Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS) [16], was employed to facilitate the quantification on TripleTOF 5600+ in a DIA-TPA workflow [17] (further referred to as the Triple Quad-TOF workflow), whereas the data-dependent acquisition (DDA) mode on Q Exactive HF-X was utilized in a standard DDA-TPA workflow [15] (further referred to as the Quad-Orbitrap workflow). The SWATH-MS method utilized on a less sensitive instrument setup allowed us to quantify lower abundant proteins in unfractionated samples by paying a one-time resource cost of extensive library construction. We illustrated the analytical capacity of our approaches in a small-scale pilot study of 20 hFF samples from four donors and derived lists of proteins possibly linked to oocyte maturity and resulting blastocyst development status. As a result, we demonstrate the compatibility of both distinct workflows in protein quantification and propose a relatively high-throughput methodological strategy of analysis of hFF proteomic composition that has potential applicability to a large clinical study planned in the future.

2. Results

2.1. Experimental Design

In this study, we have developed and optimized two distinct quantitative workflows for the analysis of the hFF material: (i) The Quad-Orbitrap workflow utilizing the Q Exactive HF-X instrument coupled to nanoL and facilitated by the DDA-TPA data analysis on the MaxQuant database search results and (ii) the Triple Quad-TOF workflow with MS measurements on the TripleTOF 5600+ coupled to microLC analyzed in the DIA-TPA method using ProteinPilot as a database search software. We have tested several analytical strategies of hFF proteomic investigation to optimize both workflows. The scheme of sample preparation involved stages of protein fractionation, proteolytic digestion, and peptide fractionation. Every experiment combined different techniques, but did not necessarily include all stages. The sample preparation scheme and instrument used in each experiment are listed in Table 1 (see Table S1 for more details).

Table 1. Sample preparation scheme including protein fractionation, method of digestion, peptide fractionation, and instrument used in each described experiment. Fraction analyzed designates the material, which was further prepared or analyzed as described, resulting from a given fractionation technique. More details on sample preparation of single samples are included in Table S1.

Workflow Abbreviation	Protein Fractionation	Fraction Analyzed	Digestion	Peptide Fractionation	Fraction Analyzed	Instrument
3TOF HFF T	None	-	FASP with trypsin	None	-	Triple TOF 5600+
3TOF HMW T	Ultrafiltration	Retentate (>10 kDa)	FASP with trypsin	None	-	Triple TOF 5600+
3TOF MF T	Immunodepletion	Low abundant proteins (MARS-14)	FASP with trypsin	None	-	Triple TOF 5600+
3TOF MR T	Immunodepletion	High abundant proteins (MARS-14)	FASP with trypsin	None	-	Triple TOF 5600+
3TOF HPLC1	None	-	FASP with trypsin	High pH RP-HPLC	60 separate fractions	Triple TOF 5600+
3TOF HPLC2	Immunodepletion	Low abundant proteins (MARS-14)	FASP with trypsin	High pH RP-HPLC	60 separate fractions	Triple TOF 5600+
3TOF HFF LTC	None		MED-FASP (LysC, trypsin, and chymotrypsin)	None	-	Triple TOF 5600+
3TOF HFF TC	None	-	MED-FASP (trypsin and chymotrypsin)	None	-	Triple TOF 5600+
3TOF HFF IS	None	-	In solution with trypsin	None	-	Triple TOF 5600+
3TOF LMW	Ultrafiltration	Filtrate (<10 kDa)	None	None	-	Triple TOF 5600+
QE HFF LTC	None	-	MED-FASP (LysC, trypsin, and chymotrypsin)	None	-	Q Exactive HF-X
QE HMW LTC	Ultrafiltration	Retentate (>10 kDa)	MED-FASP (LysC, trypsin, and chymotrypsin)	None	-	Q Exactive HF-X
QE MF LTC	Immunodepletion	Low abundant proteins (MARS-14)	MED-FASP (LysC, trypsin, and chymotrypsin)	None	-	Q Exactive HF-X
QE MR LTC	Immunodepletion	High abundant proteins (MARS-14)	MED-FASP (LysC, trypsin, and chymotrypsin)	None	-	Q Exactive HF-X
QE HFF IS	None	-	In solution with trypsin	None	-	Q Exactive HF-X
QE LMW	Ultrafiltration	Filtrate (<10 kDa)	None	None	-	Q Exactive HF-X

The Quad-Orbitrap workflow was aimed at the highest proteomic coverage with an inclusive quantitative analysis. Therefore, we analyzed the MED-FASP samples on the higher resolution-instrument in longer nanoLC gradients; a procedure which consumed the most time and resources in each single analysis. On the other hand, we designed the Triple Quad-TOF workflow to provide a higher throughput while maintaining a reasonable depth of proteomic analysis. For that reason, we performed shorter MS analysis of samples after only one digestion in the FASP procedure. However, using the advantage of the SWATH-MS methodology, we performed multiple fractionation experiments and utilized the results in the spectral library creation. In that manner, we went through the one-time resource cost to increase the quantification capability of the unfractionated clinical samples.

Due to instrument availability, more extensive investigation, including a number of fractionation experiments, was conducted on TripleTOF 5600+. Briefly, we have tested four methods of protein digestion: in solution digestion by trypsin; FASP by trypsin, MED-FASP with consecutive digestions by LysC, trypsin, and chymotrypsin; and MED-FASP with consecutive digestions by trypsin and chymotrypsin. Two of those, in solution digestion and MED-FASP with three enzymes, were employed in the analyses on both instruments and the other two were only used for samples analyzed on TripleTOF 5600+. We utilized three fractionation techniques. The first technique of protein fractionation applied in the study is ultrafiltration. We obtained two fractions after ultrafiltration on 10 kDa cutoff membrane Amicon filters: HMWF containing proteins as the retentate and LMWF containing endogenous peptides as the filtrate. HMWF was further digested, while LMWF was prepared for MS measurements without proteolytic digestion to identify peptides resulting from physiological protein breakdown in hFF. Next applied protein fractionation method was immunodepletion of high abundant proteins by the use of MARS Hu-14 cartridge. We analyzed both the depleted protein fraction enriched in low abundant proteins (MF) and the resulting high abundant proteins fraction (MR). Lastly, the experiments involving high pH RP-HPLC fractionation of peptides were conducted only on the TripleTOF 5600+ instrument to increase the proteomic coverage of the subsequently utilized SWATH-MS spectral library.

2.2. The Quad-Orbitrap Workflow

2.2.1. Workflow Optimization

We identified 942 proteins in all experiments conducted on the Q Exactive HF-X instrument using two methods of sample fractionation (immunodepletion and ultrafiltration, see Figure 1a–e) and two methods of sample digestion (MED-FASP and in solution digestion, see Figure 1a,e). The results for single experiments are listed in Table S2 and their summaries are shown in Figure 1a.

Protein digestion by MED-FASP resulted in a higher yield of identified and quantified proteins in comparison to in solution digestion: 565 and 438 proteins, respectively (1.49-fold and 1.65-fold more). Moreover, the quality of quantification was better in MED-FASP digested samples, which is especially evident in the numbers of proteins quantified by five and more peptides and their much lower CV values. Therefore, MED-FASP was employed as the method of choice in subsequent experiments.

An additional 36 identifications were achieved in HMWF after fractionation by ultrafiltration in comparison to unfractionated samples and 26 proteins were detected exclusively in HMWF. Moreover, a substantial part of LMWF proteins were found only in this fraction (44% unique identifications). Numbers of proteins quantified in HMW and unfractionated samples were very similar and so was the quality of quantification. However, the differences in absolute concentrations between those samples were apparent (see Figure 1b). With lowering concentration, the concentration deviation becomes higher, yet even medium abundant proteins display alterations in their concentration.

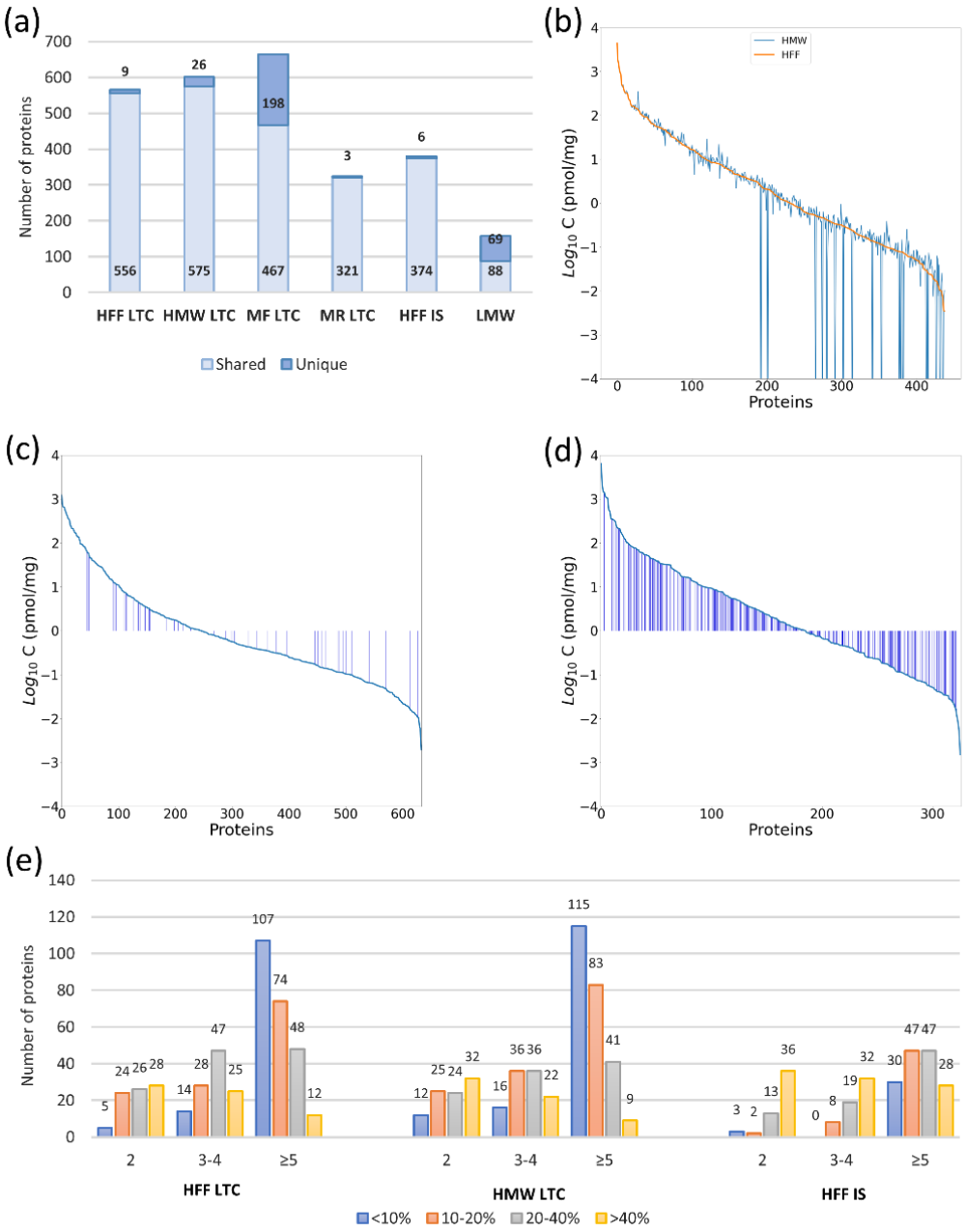


Figure 1. The effect of sample preparation on the identification and quantification of the result in the Quad-Orbitrap workflow. Sample names refer to abbreviations listed in Table 1 without the instrument designation (QE). (**a**) Numbers of proteins identified in all single experiments segregated into proteins found in multiple experiments (shared) and identified only in one experiment (unique). (**b**) Difference in protein concentration analyzed in the unfractionated (HFF) and fractionated by ultrafiltration (HMW) samples. Proteins are arranged in a decreasing mean concentration order in the unfractionated sample. Zero values in HMW experiment were implemented as 10^{-6} in order to present them in the chart. (**c**,**d**) Quantification of proteins after immunodepletion: low abundant protein fraction (**c**) and high abundant protein fraction (**d**). In each chart, proteins are arranged in a decreasing order of concentration. Blue bars represent proteins present in lower concentration than in the unfractionated sample. (**e**) Evaluation of protein quantification of samples: unfractionated sample digested by MED-FASP, unfractionated sample digested in solution, and sample fractionated by ultrafiltration digested by MED-FASP. Bars represent numbers of proteins quantified by a given number of peptides (grouped into columns) and at a given coefficient of variation (CV) value (showed in different colors).

The other applied fractionation technique was immunodepletion using the MARS-14 cartridge. A lot more proteins were captured by the cartridge than the anticipated 14 HAPs (high abundant proteins), resulting in 324 identifications. However, HAP fraction generally contained fewer identified proteins than the unfractionated sample. Conversely, the immunodepleted fraction was the richest in identifications among all experiments with 665 total and 198 uniquely identified proteins. Figure 1c,d depict the changes in concentrations of proteins quantified by the Quad-Orbitrap workflow in HAPs and immunodepleted fractions, respectively, in comparison to the unfractionated sample. Most proteins in the HAPs fraction are present at lower concentrations than in the unfractionated sample (proteins designated as blue bars in Figure 1c,d); however, general protein concentrations are mostly relatively high. On the other hand, a very small fraction of proteins in the immunodepleted fraction are present at lower concentrations than in the unfractionated sample, yet most proteins are present at relatively low concentrations and obviously more proteins in this fraction could be quantified.

2.2.2. Functional Enrichment Analysis on Interaction Networks Created in STRING Database

We divided the list of 438 proteins quantified by the Quad-Orbitrap workflow in an unfractionated hFF sample (see Table S3) ranked by their absolute concentration according to their abundance into three groups: (i) high abundant proteins (HAPs, 95% of total quantified proteins' mass; 72 most abundant proteins), (ii) medium abundant proteins (MAPs, 4.5% of total quantified proteins' mass; 102 proteins subsequent in concentration), and (iii) low abundant proteins (LAPs, 0.5% of total quantified proteins' mass; 264 low abundant proteins). The most abundant protein, which was serum albumin, was present at 4442.45 pmol/mg (30.82% of all protein content) and the least abundant protein in HAPs group was immunoglobulin lambda-like polypeptide 5, which was present at 51.76 pmol/mg (0.12%). Furthermore, top 20 proteins added to 75.70%, whereas top 10 proteins constituted as much as 64.93% of all protein content. One hundred and two lower abundant proteins were grouped as MAPs, which comprised approximately 4.5% of total quantified proteins' content. The first protein in this group was immunoglobulin lambda variable 1–47 (92.4 pmol/mg, 0.11%) and the last protein was immunoglobulin heavy variable 4–4 (9.33 pmol/mg, 0.01%). The least abundant proteins numbering 264 were classified as LAPs and comprised approximately 0.59% of the total quantified proteins' content. Moreover, we also included 309 proteins quantified by the Quad-Orbitrap workflow only after immunodepletion with the use of MARS-14 kit in our functional analysis to present a group of proteins masked by the presence of HAPs.

We created interaction networks for each protein group using the STRING database [18] and focused on the functional enrichment analysis for three categories: biological process (GO), molecular function (GO), and Reactome [19] pathways (Table S4). The reason for omitting other categories were multiple instances of repeating terms, the most notable were the following: extracellular region in cellular component (GO), signal or secreted in UniProt keywords, and coagulation cascade in KEGG [20] pathways (excluding LAPs group and group of proteins quantified only after immunodepletion with ECM-receptor interaction and pentose phosphate pathway as the top term, respectively). HAPs are mostly associated with platelet degranulation, protein activation cascade, enzyme (notably endopeptidase) inhibitor and regulator activity, and regulation of IGF transport and uptake by IGFBPs. However, the fact that not all protein identifiers were imported into the STRING database must be noted, as multiple immunoglobulin-related entries were discarded at that point. MAPs are also linked to the protein activation cascade; however, more focus is placed on complementing cascade regulation and the immune system. Interestingly, their molecular function is related to peptidase activity and its regulation. LAPs and proteins quantified only after immunodepletion were similar in terms of results, which included the following terms: regulated exocytosis, biological adhesion, protein and other molecule binding, and immune system.

As a next step, we also created interaction networks in the described manner for proteins identified in LMWF by the Quad-Orbitrap workflow: (i) all 157 identified proteins and (ii) 69 proteins identified exclusively in LMWF. In the case of all LMWF proteins (Figure S1a), many terms overlap with previous HMWF analysis: platelet degranulation, endopeptidase inhibitor activity, peptidase regulator activity, or hemostasis (Table S5). Moreover, even though most proteins were also assigned to extracellular region, we colored the resulting network according to different protein localization due to the formation of visible interaction clusters in the network. These clusters are mostly related to the nucleosome, ribonucleoprotein complex, and extracellular region and/or vesicle. Differences in assigned terms worth mentioning are the negative regulation of biological process, transport, and developmental biology. The network of proteins exclusive to LMWF consists mostly of proteins localized in nucleosome and ribonucleoprotein complex (Figure S1b). Therefore, most of the enriched functional terms in this network are related to translation and accompanying processes.

2.3. The Triple Quad-TOF Workflow

Initially, we tested the efficiency of the Triple Quad-TOF workflow in 1 h and 30 min LC gradients using sample preparation methods analogous to the Quad-Orbitrap workflow (see Section 2.2.1). In these experiments, we identified approximately 5 to 20% less proteins using shorter gradients, depending loosely on the complexity of the sample (e.g., 5% less in case of unfractionated sample digested in solution and 20% less in immunodepleted fraction digested by FASP). The 30 min LC gradient was selected for the study due to the relatively small decrease in identifications as compared to the substantial advantage of shorter analysis time. Using these short LC-gradients, we identified 259 proteins in total (results presented on Figure 2a–d).

In order to test if we could enhance these results, we conducted high pH RP-HPLC peptide fractionation. Thus, we obtained 1151 identifications in those experiments alone at the cost of additional work labor and instrument analysis time. Only 31 proteins identified in other experiments were not detected after this method of fractionation, which taken together resulted in a total of 1182 identifications. The results for single experiments are listed in Table S6 and their summaries are shown in Figure 2a.

MED-FASP was more effective than in solution digestion in terms of identification and quantification of proteins to an even higher extent than in the case of the Quad-Orbitrap workflow resulting in 154 and 122 proteins, respectively (1.66-fold and 3.49-fold more). Additionally, we tested the FASP digestion with a single enzyme (trypsin) to simplify the sample preparation protocol for the purpose of the SWATH-MS quantification (see Table S7). Here, we obtained a lower number of identifications than in the case of MED-FASP (25 proteins less); however, the quantification parameters have improved. We were able to quantify only seven more proteins, yet the total number of proteins quantified at CV < 10% rose from 46 to 70. We examined MED-FASP digestion with two enzymes only in terms of identified proteins and obtained a number between identifications for FASP digested samples and MED-FASP digested samples, which is as expected. Due to much-improved quantification results, we applied the FASP digestion with trypsin in preparation of other samples in the Triple Quad-TOF workflow. Moreover, we examined the quantification capabilities of this method also in the 1 h LC gradient and observed a 10% increase in quantified proteins using basic libraries without the high pH RP-HPLC measurements (126 proteins in 1 h as compared to 113 proteins in 30 min).

Fractionation by ultrafiltration resulted in a similar number of identified and quantified proteins in the HMW fraction as in the unfractionated sample (with 16 new identifications in HMW in respect to HFF). However, we observed a substantial decrease in lower CV values (<10%) for HMW samples. The analysis of LMW fraction was much less comprehensive in comparison to the Quad-Orbitrap workflow, yet we also managed to identify a few unique proteins. Immunodepletion also produced similar identification results, however, as much as 69 proteins identified in immunodepleted fraction constituted

identifications that were not found in the unfractionated sample. The HAPs fraction was characterized by the smallest number of identified proteins.

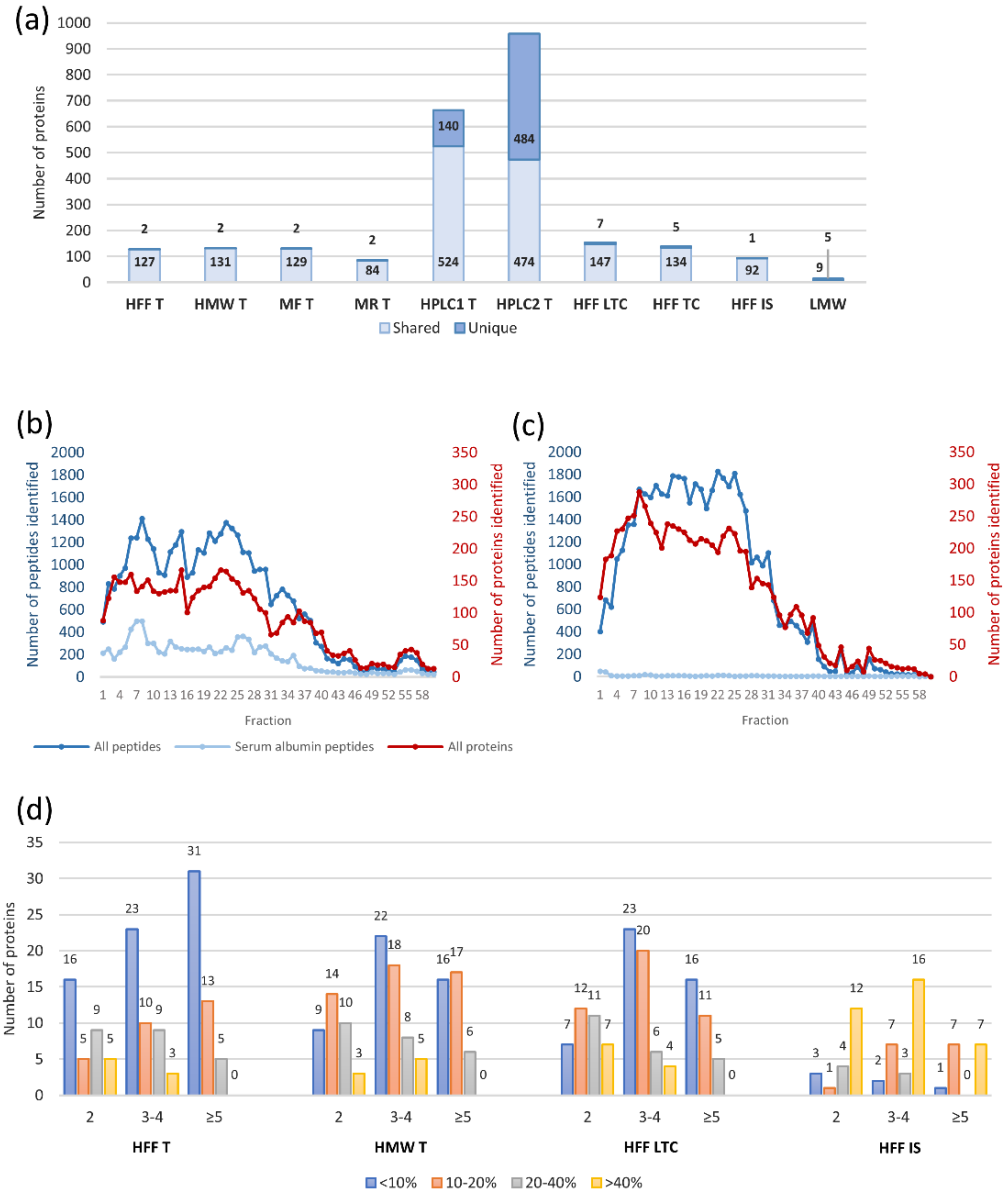


Figure 2. The effect of sample preparation on the identification and quantification result in the Triple Quad-TOF workflow. Sample names refer to abbreviations listed in Table 1 without the instrument designation (3TOF) (**a**) Numbers of proteins identified in all single experiments segregated into proteins found in multiple experiments (shared) and identified only in one experiment (unique). (**b**,**c**) Numbers of proteins (red), all peptides (dark blue), and serum albumin peptides (light blue) in each fraction in high pH RP-HPLC experiments: without prior protein fractionation—HPLC1; (**b**) and after immunodepletion of high abundant proteins—HPLC2 (**c**). (**d**) Evaluation of protein quantification of samples: unfractionated samples digested by MED-FASP, FASP, or in solution and sample fractionated by ultrafiltration digested by FASP. Bars represent numbers of proteins quantified by a given number of peptides (grouped into columns) and at a given coefficient of variation (CV) value (showed in different colors).

To obtain a more comprehensive view of hFF proteome with the use of TripleTOF 5600+, we employed the high pH RP-HPLC fractionation of peptides and additionally performed an experiment, where we coupled it with prior protein immunodepletion. The

use of this peptide fractionation method allowed us to achieve a similar or slightly higher number of protein identifications as with the use of Q Exactive HF-X when applied as a standalone technique (664 identifications, 140 unique) and to significantly exceed it in combination with immunodepletion (958 identifications, 484 unique). Nonetheless, this gain in identifications comes at a price of a significant instrument time increase as each experiment consisted of 60 separately analyzed fractions. Figure 2b,c demonstrates chromatogram-like charts of numbers of proteins, peptides, and peptides derived from albumin in each analyzed fraction. Both experiments show that proteins were mainly identified in the first 40 fractions and in especially high numbers in the first 25 fractions. Peptides derived from albumin were present in each fraction in the first experiment, possibly overshadowing other identifications, which could be detected after albumin depletion in the combined experiment.

2.4. Literature Comparison to Proteomic Studies of hFF and Related Biological Materials

In the conducted literature comparison, we focused on a few of the most comprehensive proteomic studies of human follicular fluid up to date: Zamah et al., 2015 (742 reported proteins) [6]; Bianchi et al., 2016 (617 reported proteins) [5]; Oh et al., 2017 (1079 reported proteins) [7]; Poulsen et al., 2019 (400 reported proteins) [4]; Zhang et al., 2019 (1153 reported proteins) [9]; and Pla et al., 2020 (2461 reported proteins) [10]. Furthermore, we compared the set of proteins identified in this study to the reported proteomes of biological materials in the proximate physiological environment, i.e., human plasma (Plasma Proteome Database—10,546 reported proteins) [21]; human oocyte (Virant-Klun et al., 2016—2154 reported proteins) [1]; and human granulosa cell (Bagnjuk et al., 2019—3642 reported proteins) [22]. The detailed comparison for each protein is presented in Table S8 and the general overview with cumulative data is presented in Table 2.

Table 2. Numbers of proteins identified in this study by Quad-Orbitrap (QE) or Triple Quad-TOF (3TOF) workflows, which were also reported in proteomic studies of hFF or proximate biological materials (plasma, oocyte, and granulosa cell).

	All Identified Proteins		All Proteins Identified in HMW Fraction		Proteins Identified only in HMW Fraction		All Proteins Identified in LMW Fraction		Proteins Identified only in LMW Fraction	
Resource	**QE (942)**	**3TOF (1182)**	**QE (873)**	**3TOF (1177)**	**QE (785)**	**3TOF (1168)**	**QE (157)**	**3TOF (14)**	**QE (69)**	**3TOF (5)**
Plasma Proteome Database [21]	773	975	723	975	644	966	129	9	50	0
Human oocyte [1] (oocyte specific)	226 (22)	301 (18)	211 (22)	301 (18)	183 (18)	294 (17)	43 (4)	7 (1)	15	0
Human granulosa cell [22]	436	599	390	599	346	591	90	8	46	0
HFF (Zamah et al., 2015) [6]	545	610	542	610	470	602	75	8	3	0
HFF (Bianchi et al., 2016) [5]	357	368	352	368	284	360	73	8	5	0
HFF (Oh et al., 2017) [7]	521	534	518	534	445	525	76	9	3	0
HFF (Poulsen et al., 2019) [4]	336	330	333	330	269	321	67	9	3	0
HFF (Zhang et al., 2019) [9]	567	647	540	647	463	640	104	7	27	0
HFF from hSAF (Pla et al., 2020) [10]	829	987	794	987	708	978	121	9	35	0
Unique	28	40	20	35	20	35	8	5	8	5

Identifications were divided into HMW and LMW fractions and identifications specific to each fraction were additionally separated. Numbers of total identifications in each instance are given in brackets. The proteomic study of human oocyte contained an inferred set of proteins specific to the oocyte; the comparison to this set is given in brackets. Numbers termed as unique denote proteins identified in this study and not reported in any listed resource.

As expected, most of the proteins identified in this study were also reported in human plasma: about 82% of all sets with the exception of LMWF-specific proteins. The next most overlapping proteome is the granulosa cell (about 45–50%) with a slight increase in overlap in LMWF. Lastly, about 22–27% of proteins were reported in human oocyte, 22 of which were described as oocyte-specific; most of those were identified in HMWF. In comparison to hFF proteomic studies mentioned above, from almost 28% (Poulsen et al. 2019 to Triple Quad-TOF-identified proteins) to about 88% (Pla et al. 2020 to Quad-Orbitrap-identified proteins) of identifications overlapped. The numbers of the same identifications relied mostly on the total number of proteins identified in the study. The overwhelming

majority of proteins reported in other hFF studies were found in the HMWF with a total of 42 proteins detected only in the LMWF. Of the proteins studied, 199 and 63 proteins identified in our study were not reported in the analyzed hFF publications and in any of the referred to sources, respectively.

2.5. Compatibility of the Quad-Orbitrap and Triple Quad-TOF Workflows

2.5.1. Quantification Capability in Pooled Material

We were able to quantify 129 and 438 proteins in an unfractionated hFF sample using the Triple Quad-TOF and Quad-Orbitrap workflows, respectively (Figure 3a, Tables S3 and S7). Five proteins were quantified exclusively by Triple Quad-TOF workflow. The majority of all proteins were quantified only with the use of the Quad-Orbitrap workflow—314 proteins. We obtained the Pearson correlation coefficient value of 0.86 for median concentrations of 124 proteins measured by both workflows and presented the scatterplot demonstrating their compatibility in Figure 3b.

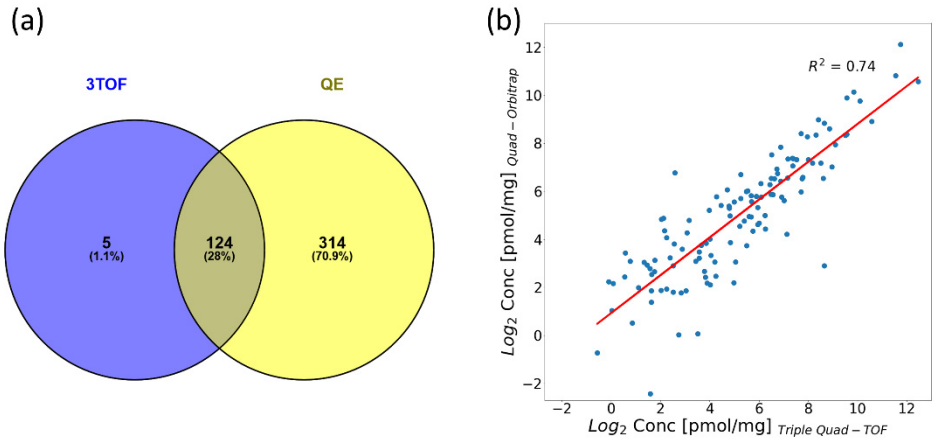


Figure 3. General comparison of compatibility of quantification capabilities of tested workflows. (**a**) Venn diagram illustrating numbers of quantified proteins by the Triple Quad-TOF (3TOF) and Quad-Orbitrap (QE) workflows. (**b**) Scatterplot of log2-transformed median concentration values from the Triple Quad-TOF and Quad-Orbitrap workflows.

2.5.2. Pilot Study on Clinical Samples

We set up a small-scale pilot study to further examine quantitative capabilities of the investigated methods and their effectiveness in elucidation of biological context in the clinical setup. Twenty hFF samples from individual follicles of four patients were included in the study (see Table S9). Selected samples differed in the maturity of retrieved oocyte and the resulting blastocyst status was obtained after embryo culture of fertilized mature oocytes. Thus, it allowed us to investigate differences in protein concentrations associated with oocyte quality along with differences stemmed from individual characteristics of patients.

Using both methodologies, we were able to quantify 471 proteins in total and 199 shared proteins by both methods (455 by the Quad-Orbitrap workflow and 215 by Triple Quad-TOF workflows, see Tables S10 and S11). We checked the consistency of obtained data for both methods by multiscatter plots (see Figures S2 and S3). P1F2 and P2F5 samples showed the lowest correlation with other samples, however, the P1F2 sample had such characteristics only in the case of the Triple Quad-TOF measurements which points to the possibility of error occurrence during the data acquisition of this sample. To compare the compatibility of the Quad-Orbitrap workflow and Triple Quad-TOF workflows in a general sense, we determined Pearson correlation values of concentrations: (i) between all clinical samples (Table S12) and (ii) in each sample resulting from both methods for each protein (Table S13). All samples presented inter-workflow correlation coefficient values higher than 0.75, excluding the P1F2 sample in which the correlation coefficient

value was lower than 0.7. In the whole set of samples, 24 proteins displayed correlation coefficient values higher than 0.7, while in the filtered set of 19 clinical samples (P1F2 sample was excluded), 38 proteins displayed correlation coefficient values higher than 0.7 (Figure S4). Due to this, we decided to eliminate the P1F2 sample from further analysis as a possible outlier. Moreover, the higher Pearson correlation value did not correspond to the overall protein abundance (see Table S13). For instance, serum albumin, which was present at the highest concentration in the analyzed material, demonstrated the value of 0.47, whereas C4b-binding protein alpha chain (less than 0.04% of mean pool protein content) displayed 0.9.

In order to discern proteome differences related to oocyte quality from individual characteristics of each patient, we employed a two-way ANOVA with one factor grouping of all individual samples from one patient (further referred to as the patient factor) and the second factor grouping according to the estimate of oocyte quality as shown in Table S9. The results of all conducted tests are listed in Table S14. As expected, proteome differences between patients exceeded characteristics associated with the assessed oocyte quality. At 5% FDR, almost 48% of all proteins measured by any method were associated with patient factor in comparison with the retrieved oocyte quality and about 45% in comparison with the blastocyst status, while at 1% FDR the proteins' concentration changes were statistically significant for about one third of all proteins. Out of 199 proteins analyzed by both methods, 139 proteins in total with 34 detected in each comparison and a total of 98 proteins and 20 detected in each comparison were linked to the patient factor at 5% and 1% FDR, respectively (see Figure S5). We did not further explore the biological significance of the established inter-patient differences due to the limited number of patients included in the study.

Instead, we focused on the proteomic signatures related to oocyte quality. Two separate ANOVA tests were conducted: The first testing the group of mature oocyte versus immature oocyte outcomes (14 and 5 samples, respectively) and the second testing the group of fertilized mature oocytes that reached blastocyst stage versus those that were arrested in development before reaching blastocyst stage (8 and 6 samples, respectively; see Table S9). The numbers of proteins relevant for oocyte maturity status for both methods at 5% FDR were 49 and 10 for the Quad-Orbitrap workflow and Triple Quad-TOF workflows, respectively. Similarly, the numbers of proteins relevant for blastocyst status were 45 and 7, respectively. Out of those proteins, all Triple Quad-TOF-appointed quality-related proteins were also quantified by the Quad-Orbitrap workflow, whereas 23 out of 49 oocyte maturity-related proteins significant from the Quad-Orbitrap workflow and 30 out of 45 blastocyst-related proteins were quantified by the Triple Quad-TOF workflow. In order to illustrate the general results of statistical analysis, we created interaction networks for each group of statistically significant proteins resulting from one of the four comparisons (see Figure S6). Most of the designated proteins displayed only minimal concentration differences between tested groups. Eight proteins related to oocyte maturity in the Quad-Orbitrap experiments exceeded 2-fold concentration changes and all those proteins were not quantified in the Triple Quad-TOF analysis. Conversely, there were three such proteins detected in the Triple Quad-TOF analysis, including the only protein appointed to be significant by both methods, which is the hepatocyte growth factor-like protein. However, Triple Quad-TOF analysis displayed a decrease in the concentration of this protein in the test group (0.39), whereas the Quad-Orbitrap workflow pointed to a slight increase (1.09). The number of proteins related to blastocyst status with a fold change higher than two was also eight in the case of the Quad-Orbitrap quantification and two of those quantified in Triple Quad-TOF experiments, while there was only one such protein significant from the Triple Quad-TOF workflow. Again, only one protein, which was carboxypeptidase B2, was designated as statistically significant by both methods and was present at decreased concentrations in the developed blastocyst group.

In order to evaluate the relevance of the obtained lists of statistically significant proteins and to further compare quantification concurrence of both methods, we calculated

ratios of median fold changes resulting from each comparison. Figure 4a, b presents charts showing differences for all 199 proteins quantified in both methods. This examination revealed the considerable agreement of quantification. For instance, less than one third of proteins were quantified with a fold change difference higher than 0.3, while more than one third of proteins displayed a fold change difference lower than 0.1 in both comparisons. Proteins determined as statistically significant by the Quad-Orbitrap workflow tended to cluster towards lesser differences in fold change ratio values in both comparisons. Proteins determined as statistically significant by the Triple Quad-TOF workflow displayed similar tendencies in the case of oocyte maturity comparison. On the other hand, most of the statistically significant proteins appointed by the Triple Quad-TOF workflow in the blastocyst status comparison have been measured with high discrepancies, except for a single protein, which was carboxypeptidase B2.

We filtered the lists of proteins determined for each comparison to those presenting 0.2-fold or lower-fold change ratio differences between the methods and presented them as potential markers related to oocyte quality (see Table S15). As a result, we obtained 20 and 22 proteins related to oocyte and blastocyst status, respectively. In Table S15, we included literature reports showing connection of listed protein appearance and/or concentration in hFF with oocyte quality [4,6,10–14,23–28]. We constructed interaction networks for proteins associated with oocyte maturity (Figure 4c) and blastocyst status (Figure 4d). Proteins with extreme changes of concentrations were filtered out entirely, either due to no quantification in another method or high inconsistencies. Proteins related to oocyte maturity with the highest fold changes were glutathione S-transferase A1 and prostatic acid phosphatase (0.35 and 1.37 mean median fold change, respectively). Both of those were unrelated to differences between patients. Proteins related to blastocyst status with the highest fold changes were vitamin D-binding protein (0.79) and plasma serine protease inhibitor (1.4). However, both of those proteins were also detected by the Triple Quad-TOF workflow to be significant for the patient factor. Concentration values in individual samples measured by both workflows for proteins with the highest fold changes are shown in Figure 5. Most of the proteins significantly related to oocyte maturity were present at slightly higher concentrations in the group of mature oocyte outcomes (12 out of 20 proteins). Moreover, these proteins were organized in interaction clusters associated mostly with complement cascade pathway, protease inhibitor activity, hemostasis, and blood coagulation. On the other hand, most of the proteins significantly related to blastocyst status were present at slightly lower concentrations in the group of developed blastocyst outcomes (18 out of 22 proteins). These proteins also assembled in interaction clusters associated with complement cascade pathway, protein activation, defense response, and platelet degranulation.

In order to obtain a closer look into differences between studied groups, we constructed hierarchically clustered heatmaps of Pearson correlations of proteins in groups of samples considered in Table S9 using results obtained by both quantitative methods (see Figures S7–S14). Both methods generated similar representations with clearly visible differences between studied groups. Analysis of the group of immature oocyte outcomes showed a few clusters of tightly positively correlated proteins, which were negatively correlated with other clusters. On the contrary, analysis of the group of mature oocyte outcomes displayed only a few highly correlated proteins, while most of the proteins remained uncorrelated in any manner. Most of the proteins determined to be significant to oocyte maturity status (listed in Table S15) were present in correlation clusters in the analysis of immature oocyte outcomes group and absent in the analysis of mature oocyte outcomes group. Representations of blastocyst development outcomes groups are more similar; however, both positive and negative correlations are stronger in the group of undeveloped blastocysts.

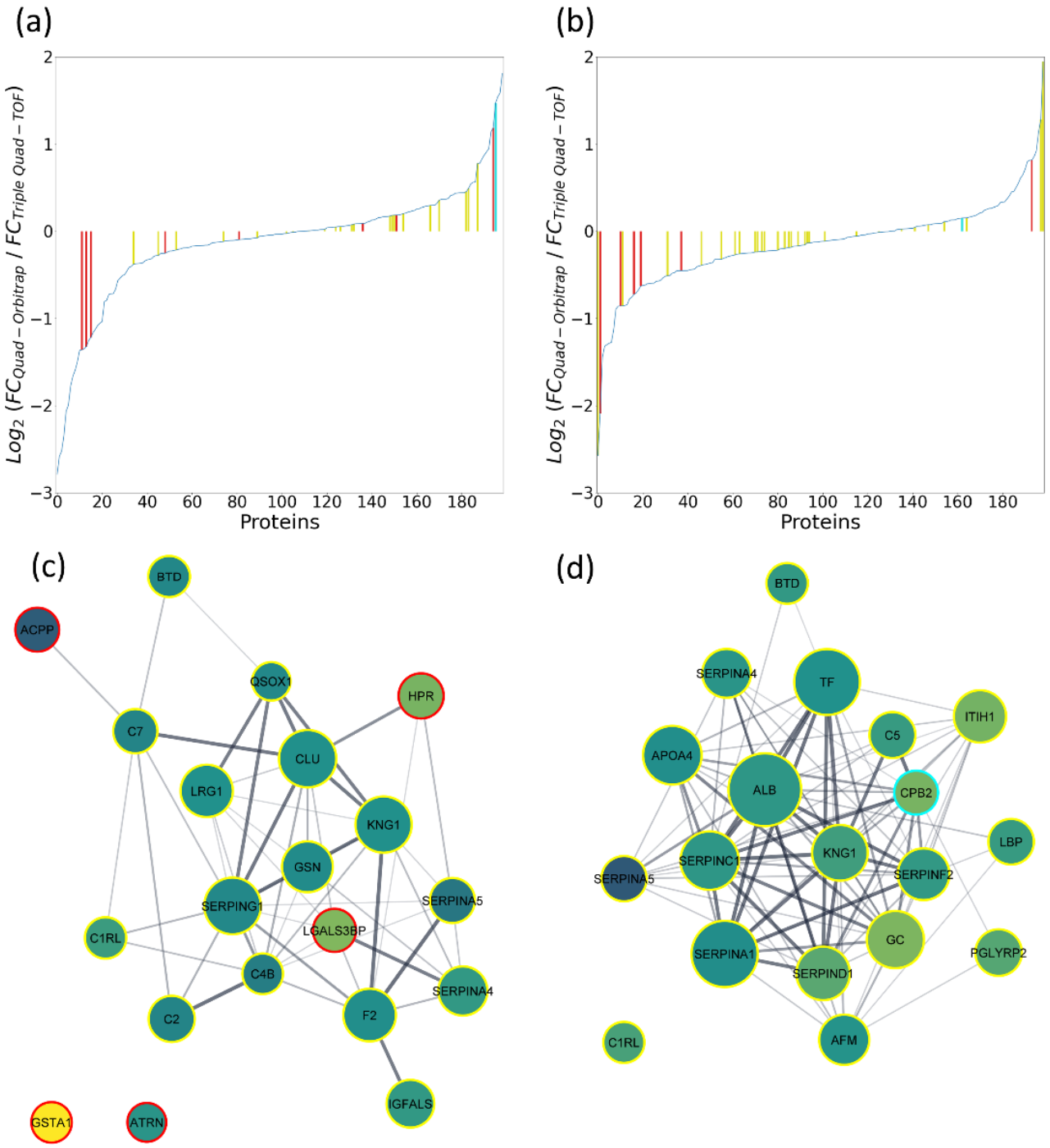


Figure 4. Proteins determined as statistically significant by two-factor ANOVA for the factor relating to (**a**,**c**) oocyte maturity or (**b**,**d**) blastocyst development status in both quantitative methods. (**a**,**b**) The ratio of fold change values of the Quad-Orbitrap workflow to Triple Quad-TOF workflow for each protein (log2). Statistically significant proteins are shown in colors (yellow—Quad-Orbitrap; red—Triple Quad-TOF; and both workflows—cyan). (**c**,**d**) Interaction networks for statistically significant proteins with less than 20%-fold change ratio difference between quantitative methods. Node edge color designates the quantitative method used to establish statistical significance (yellow—Quad-Orbitrap; red—Triple Quad-TOF; and both workflows—cyan). Fill color relates to the mean fold change of both methods, from 0.5 and below (yellow) through 1 (aquamarine) to 2 and above (purple). Node size represents mean log10 median abundance TPA concentrations of both methods in the test group (either mature oocyte or developed blastocyst).

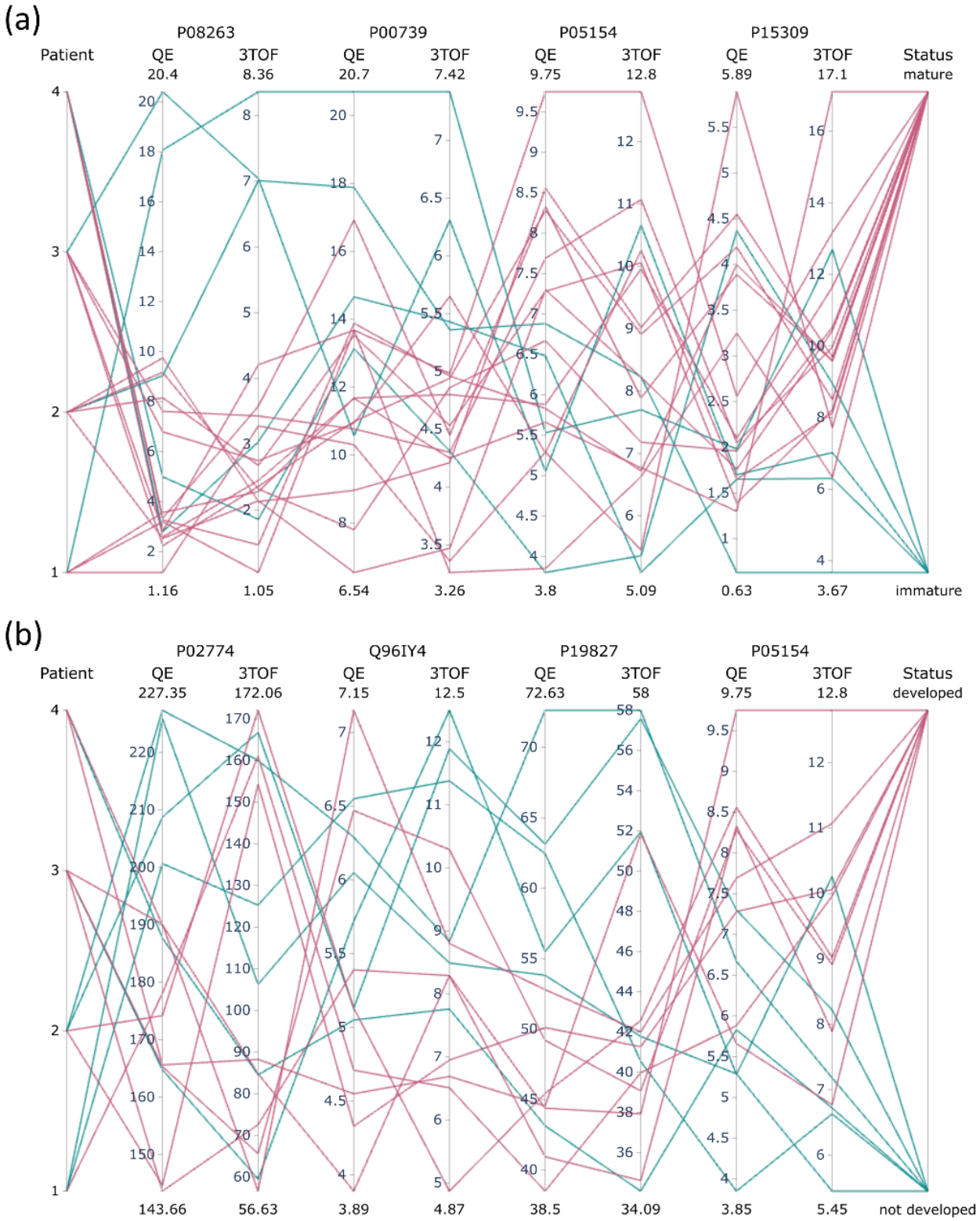


Figure 5. Median sample concentrations of proteins related to (**a**) oocyte maturity or (**b**) blastocyst development status with the highest fold changes measured by both workflows presented in parallel coordinates charts. Protein concentrations are scaled separately on each axis. Red and blue lines mark samples with the positive outcome (mature oocyte or developed blastocyst) and samples with the negative outcome (immature oocyte or not developed blastocyst), respectively.

3. Discussion

In this study, we attempted a directed development of two distinct quantitative proteomic workflows based on different high-resolution mass spectrometers: One aimed at the highest proteomic coverage (the Quad-Orbitrap workflow) and the other designed for the overall robustness of the analysis (the Triple Quad-TOF workflow). We further investigated their capabilities of hFF composition analysis and their potential to facilitate the search for oocyte quality biomarkers. Results obtained from both workflows were analyzed using TPA [15], allowing the unified comparison of the compatibility of the utilized workflows on absolute protein concentrations. The details concerning both workflows are summarized in Table 3.

Table 3. General description of the components and main outcomes of the two applied protein quantification workflows.

	Method 1 (Quad-Orbitrap Workflow)	**Method 2 (Triple Quad-TOF Workflow)**
Instrument	Q Exactive HF-X	TripleTOF 5600+
LC (flowrate)	nanoflow (0.3 µL/min)	microflow (5 µL/min)
Digestion method (enzymes)	MED-FASP (Lys-C, Trypsin, Chymotrypsin)	FASP (Trypsin)
Quantification method	DDA-TPA	DIA-TPA (SWATH-MS based)
Single LC-MS/MS analysis time (full in triplicate)	95 min (285 min)	30 min (90 min)
Number of proteins identified in the unfractionated sample using the described digestion method	565/380 (MED-FASP/IS)	129/154/93 (FASP/MED-FASP/IS)
Number of proteins identified in the LMW fraction	157	14
Number of proteins quantified in the pool sample	438	129
Number of proteins quantified in the pool sample with CV < 10%/< 20%	126/252	70/98
Number of proteins quantified in the clinical samples	455	215
Number of proteins linked to the patient factor (inter-patient differences) at 1%/5% FDR	68/101	64/96
Number of proteins linked to the oocyte status (inter-follicle differences) at 1%/5% FDR	11/49	1/10
Number of proteins linked to the blastocyst status (inter-follicle differences) at 1%/5% FDR	13/45	2/7

The Quad-Orbitrap workflow involved the measurements on Q Exactive HF-X. Such orbitrap-based spectrometers coupled with nanoflow LC are the instruments most widely used in discovery proteomics and instruments of that type have been applied in most of the hFF proteome studies presenting high numbers of protein identifications [4,6–8,13]. Additionally, we combined the sensitive MS setup advantage with the most exhaustive of the tested digestion methods—MED-FASP using three consecutive digestions to increase proteomic coverage [15]. The results obtained in this study confirm the anticipated advantage of the Quad-Orbitrap workflow in both the detection and quantification of high numbers of proteins without the need of extensive fractionation (see Figure 1, Table 3). This trend is especially evident in terms of low abundance proteome components [29], which can be assumed here based on three aspects of the analysis: (i) numbers of identified proteins, (ii) numbers of quantified proteins, and (iii) numbers of statistically significant differentiators of oocyte or blastocyst status. We have shown that fractionation has a high impact on absolute concentrations of single proteins, even in the case of simple ultrafiltration (see Figure 1b) and we recommend avoiding any fractionation procedures in quantitative experiments on clinical samples.

In the Triple Quad-TOF workflow, we used another high-resolution mass spectrometer TripleTOF 5600+, which could be coupled to a more robust microflow LC, reducing the time of a single analysis and overall workload of instrument maintenance [30]. We applied SWATH-MS as a DIA method [16], which allowed us to extend the spectral library with additional fractionation while keeping the preparation of clinical samples uncomplicated. Recent developments in SWATH-MS quantification on short microflow LC gradients point

to a possibility of further reduction in the analysis time [31]. Furthermore, we utilized FASP with a single trypsin digestion, which resulted in a better quantitative performance in combination with SWATH-MS than MED-FASP (see Figure 2d). Nevertheless, both FASP and MED-FASP protocols can be further shortened considerably by using reduced digestion times (1–2 h) [32], which can be cut down even more significantly with the use of the recently introduced ultrasonic-based FASP [33]. The extent of proteins identified with the Quad-Orbitrap workflow could be reached by the Triple Quad-TOF workflow only after the peptide RP-HPLC fractionation in high pH and exceeded with two-step fractionation (immunodepletion followed by HPLC). This comprehensive outcome came at a cost of additional 60 MS measurements; however, due to the principle of SWATH-MS quantification, proteins detected in the fractionation experiments could be included in the spectral library to improve quantification in unfractionated samples. The cost of additional measurements for fractionation experiments becomes less important with their application in large scale studies (on many clinical samples).

The complete separation of all stages of the proteomic analysis of clinical samples in both workflows ought to be kept in mind in the consideration of the compatibility of quantification. Due to the use of TPA, we were able to compare the compatibility of both workflows on absolute biological concentrations. In the pool unfractionated hFF sample, 309 more proteins could be quantified in the Quad-Orbitrap workflow, but only 154 more were quantified with CV less than 20% and even less—56 with CV less than 10%. This further confirms the suitability of the Quad-Orbitrap workflow for the analysis of a great number of low abundance proteome components, which is analyzed with slightly worse accuracy. The accuracy of the measurements could, however, be further improved by facilitating the concentration calculations by the implementation of the MaxLFQ algorithm [34] at the cost of lower proteomic coverage. The number of quantified proteins increases with the number of analyzed samples, but the difference is much more significant for the Triple Quad-TOF workflow. It may indicate that the Quad-Orbitrap workflow is near its limit of quantifiable proteins without fractionation, whereas the Triple Quad-TOF workflow may be capable of a moderate increase in the number of quantified proteins with larger number of samples analyzed. Proteins numbering 124 were quantified by both workflows in a pool sample. We achieved a high degree of correspondence between the workflows, which can be observed on the scatterplot in Figure 3b, and by the high Pearson correlation coefficient value of 0.86. In the set of clinical samples, the correlations between single samples analyzed by different workflows for 199 shared proteins were lower than in the case of the pool sample, yet what is noteworthy is that all samples subjected to subsequent statistical analysis (excluding the P1F2 sample) displayed Pearson correlation coefficients higher than 0.75 (see Table S12). However, in case of the analysis of single proteins in clinical samples, we found rather considerable discrepancies between both workflows as only about 20% of proteins displayed Pearson correlation higher than 0.7 in our comparison (see Table S13). The agreement of quantification was not related to the general protein concentration in hFF. On the other hand, the group quantifications of mature/immature oocyte or developed/not developed blastocyst comparisons were remarkably consistent as demonstrated by low fold change ratio differences (see Figure 4a,b). It must be noted that the overall group fold changes were rather minimal, which might have assisted the unity of both workflows in that aspect. Both workflows have shown significant similarity also in the hierarchically clustered heatmaps of Pearson correlations of proteins, which we constructed for each development status-linked group of samples (see Figures S7–S14). There, the differences between mature and immature oocyte groups and developed and not developed blastocyst groups are clearly visible due to noticeably higher numbers of positively and negatively correlated clusters in the immature oocyte/not developed blastocyst group.

Results obtained in this study point to much higher influences of inter-patient differences than the inter-follicle differences based on the status of oocyte or blastocyst on the hFF composition. Similar numbers of proteins linked to inter-patient differences have

been established by both workflows; however, significantly more proteins linked to the oocyte quality features have been determined by the Quad-Orbitrap workflow. Again, it is most likely related to lower concentrations of these proteins, as for instance more than half of the proteins determined to be associated with oocyte maturity by the Quad-Orbitrap workflow have not been quantified by the Triple Quad-TOF workflow. Although only one protein per maturity comparison was appointed to be significant from both workflows (hepatocyte growth factor-like protein and carboxypeptidase B2 for oocyte and blastocyst status, respectively, see Table S14), most of the remaining significant differentiators quantified by both methods were measured with good consistency, allowing us to obtain unified lists of proteins possibly linked to oocyte quality (see Table S15). It is crucial to acknowledge the difference between the maturity comparisons carried out in this study. Identification of proteins linked to oocyte maturity is more straightforward since it is not dependent on further events, while formation of the blastocyst relies heavily on the success of the fertilization process, sperm quality, and subsequent embryo culture; factors which are separated from hFF. Thus, proteins associated with the oocyte maturity may play a more direct role in the oocyte development or their concentrations may be an explicit consequence of physiological circumstances accompanying its course. The inference of proteins related to blastocyst status, however, may be too distant or even coincidental, yet its success would be a genuine fulfillment of the oocyte quality-related investigation. List of proteins appointed by both workflows and the literature reports of their possible connection to oocyte quality are summed up in Table S15.

4. Materials and Methods

4.1. Research Approval

Conducted experiments described here are part of the "Identification of biomarkers of early embryonic development and pregnancy" project that has been approved by the Independent Bioethics Commission at the Medical University of Gdansk (decision 62/2016). Each couple undergoing IVF treatment has signed a written informed consent regarding the treatment and all included procedures. The obtained written consent also include an agreement for the publication of treatment-related data if patient anonymity is maintained.

4.2. Collection of Samples

A pool sample of hFF from multiple patients along with 20 samples from single follicles of four patients was obtained from the INVICTA Fertility and Reproductive Center in Sopot. Patients who took part in this study underwent the IVF procedure due to male factor infertility and received hormonal stimulation as part of the procedure. The hormonal stimulation protocol and sample retrieval procedures were outlined in detail in our previous publication [14]. All samples were free of visible blood contamination and were stored at −20 °C until analysis. Retrieved mature oocytes were fertilized by intracytoplasmic sperm injection (ICSI). Embryos were cultured as previously described [35] and evaluated according to the 2011 Istanbul consensus criteria [36].

4.3. Experimental Design

Specific details on the samples prepared for all the experiments described in this work are listed in Table S1, including the specification of digestion, fractionation, spiking of retention time calibration standard, mode of acquisition, instrument used, number of samples and repetitions, and the general purpose in the experimental task. Briefly, the experiments involved comparative qualitative and quantitative studies, spectral library preparation for SWATH-MS quantification, and a pilot study on clinical samples. The pilot study included five samples from single follicles per each of the four patients. Out of twenty total collected samples: five were linked to the immature oocyte and fifteen to mature oocytes out of which, after fertilization, nine developed relative to the mature blastocysts and six arrested before the compactation stage (see Table S9). Data processing for both tested workflows was performed separately for clinical samples and each pool experiment.

4.4. Sample Preparation

Pool samples of hFF were centrifuged at 1000× *g* for 10 min to separate cell remains from the fluid and pellets were discarded. Next, samples were either subjected to protein fractionation (Protein fractionation subsection) or left unfractionated. Protein concentrations were measured by spectrophotometer measurements of absorbance at 280 nm. The material was digested either by FASP, MED-FASP (Multi-Enzyme Digestion FASP) or in-solution digestion (see Protein digestion subsection). The resulting proteolytic peptides were fractionated by RP-HPLC in high pH (Peptide fractionation by high pH RP-HPLC subsection) or desalted in STAGE (STop And Go Extraction) Tips procedure [37] on in-house prepared tips filled with C18 solid phase (3M™ Empore™, St. Paul, MN, USA). Briefly, 10 µg of peptides was added on the tip equilibrated beforehand by 1% acetic acid in water. After washing, peptides were eluted by a buffer containing 60% acetonitrile (ACN)/1% acetic acid in water and evaporated in a SpeedVac to obtain volumes ready for MS measurements (5 µL for Q Exactive HF-X or 10 µL for Triple TOF 5600+). The iRT (indexed retention time) Kit (Biognosys, Zurich, Switzerland) was spiked into samples used for SWATH-MS spectral library preparation or SWATH-MS quantification in a 1:10 standard to sample volume ratio in order to perform the retention time calibration.

Clinical samples were prepared as described above for pooled samples; however, no protein or peptide fractionations were applied. Proteins were digested either by FASP for the Triple Quad-TOF workflow or MED-FASP for the Quad-Orbitrap workflow.

4.4.1. Protein Fractionation

Samples were depleted of high abundant serum proteins (HAP) with the use of Multiple Affinity Removal Spin Cartridge (MARS Hu-14, Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. Briefly, the material was diluted with the supplied buffer, filtered through a spin filter of 0.22 µm cutoff membrane for 2 min at 16,000× g, and applied to the cartridge according to the manual. HAP-depleted fractions and HAP-enriched fractions resulting from multiple procedures were combined individually, concentrated on 5000 kDa MW 4 mL Spin Concentrators by subsequent centrifugations at 4000× *g* 10 °C for 30 min, and finally the buffer was exchanged into 50 mM NH_4HCO_3 on concentrators. Both fractions were subjected to digestion.

Samples were divided into low molecular weight fraction (LMWF) and high molecular weight fraction (HMWF) by ultrafiltration on Amicon filters with 10 kDa cutoff membrane (Merck-Millipore, Burlington, MA, USA) for 15 min at 14,000 g with the addition of 20% ACN as described before [14,38]. LMWF was subsequently prepared for MS analysis in the STAGE Tips procedure. HMWF was subjected to proteolytic digestion.

4.4.2. Protein Digestion

Protein material was digested by trypsin (1:50 enzyme to protein weight ratio) in a standard FASP procedure [39] on Microcon with 30 kDa cutoff membrane (Merck-Millipore, Burlington, MA, USA). The MED-FASP procedure involved three consecutive digestions by LysC (1:50), trypsin (1:100), and chymotrypsin (1:100); and the modified MED-FASP [39] consisted of two digestions: trypsin (1:50) and chymotrypsin (1:100) (all enzymes from Promega Corporation, Madison, WI, USA). First, hFF was lysed with the use of buffer containing 1% sodium dodecyl sulfate (SDS), 50 mM dithiothreitol (DTT) in 100 mM Tris-HCl pH 8 for 10 min in 95 °C (all reagents from Sigma-Aldrich, St. Louis, MO, USA). The amount of 100 µg of protein was administered to each filter. Briefly, the filters were washed with the buffer containing 8 M urea in 100 mM Tris-HCl pH 8.5 multiple times by centrifugation at 10,000× *g* in 20 min. Proteins were alkylated with the use of 55 mM iodoacetamide (IAA, Sigma-Aldrich, St. Louis, MO, USA) at room temperature in the dark for 20 min. Finally, traces of IAA and urea were washed with 100 mM Tris-HCl pH 8.5 and enzyme was added to the filters for the overnight digestion at 37 °C. The resulting peptides were eluted with 100 mM Tris-HCl pH 8.5. In the case of MED-FASP, filters were placed in new tubes and digestion and elution steps were repeated with other enzymes. Digestion

with chymotrypsin was performed for 3 h in the buffer containing 10 mM $CaCl_2$ in 100 mM Tris-HCl pH 7.8. Eluted peptides were desalted in the STAGE Tips procedure [37].

In-solution protein digestion was executed according to previously used protocol [14,40] suggested by Gundry et al. [41]. Briefly, proteins diluted in 50 mM NH_4HCO_3 solution were subjected to reduction by 10 mM DTT for 30 min at 56 °C and subsequent alkylation by 20 mM IAA for 30 min at room temperature. Trypsin in 1:50 enzyme to protein weight ratio was added to samples for overnight incubation at 37 °C. Digestion was stopped by 50% ACN/5% trifluoroacetic acid solution in water and samples were desalted in the STAGE Tips procedure [37].

4.4.3. Peptide Fractionation by High pH RP-HPLC

Non-desalted samples of peptide material were fractionated by RP-HPLC separations in high pH in a similar manner as described before [40] on the analytical Prominence HPLC System with the UV-VIS detector (Shimadzu, Kyoto, Japan). Applied buffer system consisted of Buffer A: 0.1% NH_4OH in water, pH 10; B: 0.1% NH_4OH in ACN, pH 10. The amount of 1 mg of peptides was separated on the Zorbax Extend-C18 column (4.6 × 150 mm, 5 µm particle size, Agilent Technologies, Santa Clara, CA, USA) into 60 2 mL fractions in 120 min gradient (0–40% Buffer B in 100 min followed by 40–100% Buffer B in 20 min). All collected fractions were evaporated to dryness in a SpeedVac, dissolved in 60% ACN/1% acetic acid in water and again evaporated to a volume of 10 µL, with the exception of 2 first fractions from each separation which were desalted in the STAGE Tips procedure [37] due to salt accumulation in those fractions.

4.5. LC-MS/MS Measurements and Quantitative Data Processing

4.5.1. Triple Quad-TOF Workflow

The LC-MS/MS measurements for the Triple Quad-TOF workflow were acquired on the TripleTOF 5600+ hybrid mass spectrometer with DuoSpray Ion Source (AB SCIEX, Framingham, MA, USA) coupled with the Eksigent microLC (Ekspert MicroLC 200 Plus System, Eksigent, Redwood City, CA, USA) in a similar manner as described before [14,40]. Samples were loaded onto the LC column using the CTC Pal Autosampler (CTC Analytics AG, Zwinger, Switzerland), with a 5 µL injection. The Buffers A and B constituted of 0.1% (v/v) formic acid in water and ACN, respectively. LC separations were carried out on the ChromXP C18CL column (3 µm, 120 Å, 150 × 0.3 mm; Eksigent, Redwood City, CA, USA) using a gradient of 8–40% Buffer B in 30 min with a flowrate of 5 µL/min. All measurements were conducted in a positive ion mode. The system was controlled by the Analyst TF 1.7.1 software (AB SCIEX, Framingham, MA, USA). The data-dependent acquisition (DDA) analyses consisted of a 250 ms TOF survey scan in the m/z range of 400–1000 Da followed by a 100 ms Product Ion scan in the m/z range of 100–1500 Da, which resulted in a 2.3 s cycle time. Top 20 candidate ions with charge state of 2 to 5 were selected for collision-induced dissociation (CID) fragmentation with a rolling collision energy. Former target ions were excluded after 2 occurrences for 5 s. SWATH-MS [16] analyses were performed in a looped product ion mode. A set of 25 variable-width windows were constructed by equalized ion frequency distribution with the use of SWATHTuner [42] to cover the m/z range of 400–1000 Da. The collision energy for each window was calculated for +2 to +5 charged ions centered upon the window with a spread of 5. The SWATH-MS1 survey scan was acquired in high sensitivity mode in the range of 400–1000 Da in the beginning of each cycle with the accumulation time of 50 ms and it was followed by 40 ms accumulation time high sensitivity product ion scans, which resulted in the total cycle time of 1.1 s. The database search for the construction of spectral library (see Experimental Design section) was performed in ProteinPilot 4.5 Software (AB SCIEX, Framingham, MA, USA) using the Paragon algorithm against the SwissProt *Homo sapiens* database (ver. 26.07.2019; 20,428 entries) merged with iRT standard sequence and the following parameters: TripleTOF 5600 instrument; alkylation of cysteines by iodoacetamide; trypsin enzyme digestion, ID focus on biological modifications; search effort "thorough ID"; and

detected protein threshold [Conf] > 10%. The resulting group file was loaded into MS/MS All with SWATH Acquisition MicroApp 2.01 in PeakView 2.2 (AB SCIEX, Framingham, MA, USA) to automatically create a spectral library with the set parameters: modified peptides allowed and shared peptides excluded. The library was processed with SWATH-MS measurements of either the pool samples or the clinical samples (see Experimental Design section). The retention time calibration was performed manually with the use of iRT kit peptides. The maximum number of peptides per protein was 6 and extracted ion chromatogram (XIC) parameters were set to 10 min extraction window width and 75 ppm XIC width. Absolute concentration values were derived from the SWATH-MS intensities using the Total Protein Approach [15].

4.5.2. Quad-Orbitrap Workflow

The LC-MS/MS measurements for the Quad-Orbitrap workflow were acquired on the QExactive HF-X mass spectrometer (ThermoFisher Scientific, Palo Alto, CA, USA) coupled with nanoLC in a similar manner as described before [43]. LC separations were conducted on a 50 cm column packed with C18 material with 75 µm inner diameter in an ACN gradient of 5–30% in 95 min at the flowrate of 0.3 µL/min. The mass spectrometer operated in DDA mode with survey scans acquired at a resolution of 60,000. Top 15 ions with charge $\geq +2$ were selected from the survey scan in the range of 300–1650 m/z with an isolation window of 1.4 m/z and fragmented by high-energy collision dissociation (HCD) with normalized collision energies of 25. The dynamic exclusion time was 30 s. The maximum ion injection times for the survey scan and the MS/MS scans were 20 and 28 ms, respectively. The ion target value for MS1 and MS2 scan modes was set to 3×10^6 and 10^5, respectively. The raw spectra of either all pool samples or all clinical samples (see Experimental Design section) were processed in the MaxQuant software [44], version 1.6.2.6a against the SwissProt *Homo sapiens* database (see Triple Quad-TOF workflow subsection) with the following parameters: carbamidomethylation as a fixed modification, trypsin enzyme digestion, and 0.01 false discovery rate (FDR). Quantification of protein concentrations was performed using the Total Protein Approach [15] for proteins with a minimum of 2 unique peptides and 70% valid values across the measurements.

4.6. Data Analysis

Database search of DDA runs was conducted with ProteinPilot and MaxQuant software for Triple Quad-TOF and Quad-Orbitrap workflows, respectively. The specific search settings in all database search engines were dependent on the experiment (see Table S1) in each software with respect to the following aspects: proteolytic digestion (LysC, trypsin, trypsin + chymotrypsin, or no digestion) and fixed modifications (carbamidomethylation of cysteines or none). All results were filtered for 0.01 FDR at the protein level. Functional analysis of proteins and interaction network construction was performed using STRING database [18], version 11. Network visualization was conducted in Cytoscape 3.8.2 [45]. Venn plots were constructed using the online Venny 2.1 tool [46]. Bar plots were created in Microsoft Excel. Additional calculations and plot visualizations, including Pearson correlation calculation and hierarchically clustered heatmaps, were performed using Python 3.8 including libraries: numpy 1.18.5, pandas 1.1.1, matplotlib 3.3.1, and seaborn 0.11.0. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [47] via the PRIDE [48] partner repository with two dataset identifiers PXD024223 and PXD024347 for projects divided into data acquired with the Triple Quad-TOF workflow and Quad-Orbitrap workflow, respectively.

4.7. Statistical Analysis

Statistical analysis was performed using Perseus 1.6.14.0 [49]. Replicates were averaged by the median value. Data were log2-transformed, any missing values were imputed from normal distribution, inspected using scatterplots, and normalized by z-score. Two separate two-way ANOVA analyses were performed with one factor grouping samples from a

single patient and another factor grouping samples associated with the mature/immature oocyte or developed/not developed blastocyst. Post hoc Tukey's HSD tests were performed for each factor in both analyses on 0.05 and 0.01 FDR level.

5. Conclusions

We have developed and presented here results of tests of two completely separate quantitative proteomic workflows involving the use of different MS instruments and analyzed their compatibility in absolute quantification of hFF proteins with the use of TPA. We observed remarkably high correlation between analyses of single samples (see Figure 3b, Table S12) and moderate correlation between analyses of single proteins (see Figure S4, Table S13) obtained by applying the Quad-Orbitrap and Triple Quad-TOF workflows. Moreover, the final biological information was widely consistent. We have obtained combined lists of proteins possibly associated with oocyte quality: 20 proteins linked to oocyte maturity and 22 proteins linked to blastocyst development status (see Table S15). Demonstrated here are proteins linked to oocyte or blastocyst status that may pose as primary targets for the candidates of the oocyte quality which should be verified on a larger set of samples.

Both tested workflows are interchangeable in general terms of the obtained results; however, each workflow has its own advantages and limitations. The Quad-Orbitrap workflow (see Table 3) was indisputably best suited for in-depth proteomic analysis of protein targets present in a wide range of physiological concentrations, especially low abundant proteome; however, high sensitivity comes at a price of longer LC-MS/MS measurement time and the cost of digestion protocol. The use of the less effective Triple Quad-TOF workflow in terms of numbers of identified and quantified proteins (see Table 3) comes with a significant advantage of more than three times shorter LC-MS/MS analysis time as well as shorter and less expensive sample preparation protocol (one enzyme used in digestion instead of three) in comparison to the Quad-Orbitrap workflow. It was already mentioned that both workflows have a potential for further analysis time reduction as well as the increase in their detection and quantification capacity with the number of analyzed samples, which is more evident in the case of the Triple Quad-TOF workflow (see Discussion). Since the results obtained by both workflows are well correlated, a choice of the workflow for a given task should be steered by its limitations and advantages as well as the access to a given type of equipment.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms22147415/s1, Supporting Material 1, Figure S1: Interaction networks of proteins identified in LMW fraction generated in STRING, Figures S2 and S3: Multiscatter plots displaying correlations of all clinical samples in quantitative workflows, Figure S4: Pearson correlation of concentration changes measured by both quantitative methods for each protein in the full set of samples and in the set excluding the P1F2 sample, Figure S5: Proteins determined as statistically significant by two-factor ANOVA for the factor relating to inter-patient differences in each comparison (patient-oocyte status and patient-blastocyst status) for both quantitative methods, Figure S6: Interaction networks of all proteins determined as statistically significant by two-factor ANOVA for the factor relating to oocyte maturity or blastocyst status for the Quad-Orbitrap or Triple Quad-TOF workflow at 5% FDR (PDF); Supporting Material 2, Table S1: Detailed description of the purpose and analysis workflow of all samples described in this study, Table S2: List of all proteins identified by the Quad-Orbitrap workflow in pool experiments described in detail in Table S1, Table S3: The results of quantitative analysis of unfractionated pool samples by the Quad-Orbitrap workflow along with TPA absolute concentration and coefficient of variation calculation, Tables S4 and S5: Functional enrichment analysis of proteins identified in HMWF and LMWF, Table S6: List of all proteins identified by the Triple Quad-TOF workflow in pool experiments described in detail in Table S1, Table S7: The results of quantitative analysis of unfractionated pool samples by the Triple Quad-TOF workflow along with TPA absolute concentration and coefficient of variation calculation, Table S8: List of all proteins identified in the study with the literature comparison to proteomic studies of hFF and related biological materials, Table S9: Assignment of clinical samples to analyzed

groups, Tables S10 and S11: The results of quantitative analysis of clinical samples by both workflows, Table S12: Pearson correlation calculated for each sample comparison in both workflows, Table S13: Pearson correlation for each protein in both workflows for all clinical samples and excluding the P1F2 sample, Table S14: The results of all conducted two-way ANOVA tests for both workflows, Table S15: List of proteins determined to be statistically significant in relation to oocyte or blastocyst status by any applied quantification method, with less than 20% difference in median fold change ratio established by both methods (XLSX); Supporting Material 3, Figures S7–S14: Hierarchically clustered heatmaps of Pearson correlations of proteins in groups of samples considered in Table S10 using results obtained by both quantitative workflows (PDF).

Author Contributions: Conceptualization, A.E.L. and S.O.; methodology, A.E.L., J.R.W. and S.O.; formal analysis, A.E.L. and M.T.; investigation, A.E.L. and A.F.; resources, P.C., K.Ł. and J.R.W.; writing—original draft preparation, A.E.L. and S.O.; writing—review and editing, A.E.L., A.F., M.T., P.C., K.Ł., J.R.W. and S.O.; visualization, A.E.L. and M.T.; supervision, P.C., K.Ł., J.R.W. and S.O.; funding acquisition, K.Ł. and J.R.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the OPUS 14 financial grant (UMO-2017/27/B/NZ5/02393) and Max-Planck Society for the Advancement of Science. In part, this work is a result of a research stay funded by the Erasmus+ programme. The APC was funded by University of Gdańsk.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Independent Bioethics Commission at the Medical University of Gdansk (decision 62/2016) as a part of the approved project "Identification of biomarkers of early embryonic development and pregnancy".

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [47] via the PRIDE [48] partner repository with two dataset identifiers PXD024223 and PXD024347.

Acknowledgments: The authors are grateful to Matthias Mann for support during the research stay. We thank Anna Knight for administrative assistance and helpful discussions and Katharina Zettl for technical assistance. Computational resources used in this project were provided by the Informatics Center of the Metropolitan Academic Network (IC MAN-TASK) in Gdańsk.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Virant-Klun, I.; Leicht, S.; Hughes, C.; Krijgsveld, J. Identification of Maturation-Specific Proteins by Single-Cell Proteomics of Human Oocytes. *Mol. Cell. Proteomics* **2016**, *15*, 2616–2627. [CrossRef]
2. Revelli, A.; Piane, L.D.; Casano, S.; Molinari, E.; Massobrio, M.; Rinaudo, P. Follicular Fluid Content and Oocyte Quality: From Single Biochemical Markers to Metabolomics. *Reprod. Biol. Endocrinol.* **2009**, *7*, 40. [CrossRef]
3. Hernández-Vargas, P.; Muñoz, M.; Domínguez, F. Identifying Biomarkers for Predicting Successful Embryo Implantation: Applying Single to Multi-OMICs to Improve Reproductive Outcomes. *Hum. Reprod. Update* **2020**, *26*, 264–301. [CrossRef] [PubMed]
4. La Cour Poulsen, L.; Pla, I.; Sanchez, A.; Grøndahl, M.L.; Marko-Varga, G.; Yding Andersen, C.; Englund, A.L.M.; Malm, J. Progressive Changes in Human Follicular Fluid Composition over the Course of Ovulation: Quantitative Proteomic Analyses. *Mol. Cell. Endocrinol.* **2019**, *495*, 110522. [CrossRef]
5. Bianchi, L.; Gagliardi, A.; Landi, C.; Focarelli, R.; Leo, V.D.; Luddi, A.; Bini, L.; Piomboni, P. Protein Pathways Working in Human Follicular Fluid: The Future for Tailored IVF? *Expert Rev. Mol. Med.* **2016**, *18*. [CrossRef] [PubMed]
6. Zamah, A.M.; Hassis, M.E.; Albertolle, M.E.; Williams, K.E. Proteomic Analysis of Human Follicular Fluid from Fertile Women. *Clin. Proteomics* **2015**, *12*, 5. [CrossRef] [PubMed]
7. Oh, J.W.; Kim, S.K.; Cho, K.-C.; Kim, M.-S.; Suh, C.S.; Lee, J.R.; Kim, K.P. Proteomic Analysis of Human Follicular Fluid in Poor Ovarian Responders during in Vitro Fertilization. *Proteomics* **2017**, *17*, 1600333. [CrossRef] [PubMed]
8. Lehmann, R.; Schmidt, A.; Pastuschek, J.; Müller, M.M.; Fritzsche, A.; Dieterle, S.; Greb, R.R.; Markert, U.R.; Slevogt, H. Comparison of Sample Preparation Techniques and Data Analysis for the LC-MS/MS-Based Identification of Proteins in Human Follicular Fluid. *Am. J. Reprod. Immunol.* **2018**, *80*, e12994. [CrossRef] [PubMed]

9. Zhang, X.; Xu, X.; Li, P.; Zhou, F.; Kong, L.; Qiu, J.; Yuan, Z.; Tan, J. TMT Based Proteomic Analysis of Human Follicular Fluid from Overweight/Obese and Normal-Weight Patients With Polycystic Ovary Syndrome. *Front. Endocrinol.* **2019**, *10*, 821. [CrossRef] [PubMed]
10. Pla, I.; Sanchez, A.; Pors, S.E.; Pawlowski, K.; Appelqvist, R.; Sahlin, K.B.; Poulsen, L.L.C.; Marko-Varga, G.; Andersen, C.Y.; Malm, J. Proteome of Fluid from Human Ovarian Small Antral Follicles Reveals Insights in Folliculogenesis and Oocyte Maturation. *Hum. Reprod.* **2021**, *36*, 756–770. [CrossRef]
11. Kushnir, M.M.; Naessén, T.; Wanggren, K.; Rockwood, A.L.; Crockett, D.K.; Bergquist, J. Protein and Steroid Profiles in Follicular Fluid after Ovarian Hyperstimulation as Potential Biomarkers of IVF Outcome. *J. Proteome Res.* **2012**, *11*, 5090–5100. [CrossRef] [PubMed]
12. Yoo, S.W.; Bolbot, T.; Koulova, A.; Sneeringer, R.; Humm, K.; Dagon, Y.; Usheva, A. Complement Factors Are Secreted in Human Follicular Fluid by Granulosa Cells and Are Possible Oocyte Maturation Factors. *J. Obstet. Gynaecol. Res.* **2013**, *39*, 522–527. [CrossRef] [PubMed]
13. Bayasula; Iwase, A.; Kobayashi, H.; Goto, M.; Nakahara, T.; Nakamura, T.; Kondo, M.; Nagatomo, Y.; Kotani, T.; Kikkawa, F. A Proteomic Analysis of Human Follicular Fluid: Comparison between Fertilized Oocytes and Non-Fertilized Oocytes in the Same Patient. *J. Assist. Reprod. Genet.* **2013**, *30*, 1231–1238. [CrossRef] [PubMed]
14. Lewandowska, A.E.; Macur, K.; Czaplewska, P.; Liss, J.; Łukaszuk, K.; Ołdziej, S. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology. *J. Proteome Res.* **2017**, *16*, 3053–3067. [CrossRef] [PubMed]
15. Wiśniewski, J.R.; Rakus, D. Multi-Enzyme Digestion FASP and the 'Total Protein Approach'-Based Absolute Quantification of the Escherichia Coli Proteome. *J. Proteomics* **2014**, *109*, 322–331. [CrossRef] [PubMed]
16. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717. [CrossRef] [PubMed]
17. He, B.; Shi, J.; Wang, X.; Jiang, H.; Zhu, H.-J. Label-Free Absolute Protein Quantification with Data-Independent Acquisition. *J. Proteomics* **2019**, *200*, 51–59. [CrossRef] [PubMed]
18. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef] [PubMed]
19. Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2020**, *48*, D498–D503. [CrossRef]
20. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
21. Nanjappa, V.; Thomas, J.K.; Marimuthu, A.; Muthusamy, B.; Radhakrishnan, A.; Sharma, R.; Ahmad Khan, A.; Balakrishnan, L.; Sahasrabuddhe, N.A.; Kumar, S.; et al. Plasma Proteome Database as a Resource for Proteomics Research: 2014 Update. *Nucleic Acids Res.* **2014**, *42*, D959–D965. [CrossRef]
22. Bagnjuk, K.; Mayerhofer, A. Human Luteinized Granulosa Cells—A Cellular Model for the Human Corpus Luteum. *Front. Endocrinol.* **2019**, *10*, 452. [CrossRef]
23. Zakerkish, F.; Brännström, M.; Carlsohn, E.; Sihlbom, C.; van der Post, S.; Thoroddsen, A. Proteomic Analysis of Follicular Fluid during Human Ovulation. *Acta Obstet. Gynecol. Scand.* **2020**, *99*, 917–924. [CrossRef]
24. Bianchi, L.; Gagliardi, A.; Campanella, G.; Landi, C.; Capaldo, A.; Carleo, A.; Armini, A.; De Leo, V.; Piomboni, P.; Focarelli, R.; et al. A Methodological and Functional Proteomic Approach of Human Follicular Fluid En Route for Oocyte Quality Evaluation. *J. Proteomics* **2013**, *90*, 61–76. [CrossRef] [PubMed]
25. Carbone, M.C.; Tatone, C.; Monache, S.D.; Marci, R.; Caserta, D.; Colonna, R.; Amicarelli, F. Antioxidant Enzymatic Defences in Human Follicular Fluid: Characterization and Age-dependent Changes. *Mol. Hum. Reprod.* **2003**, *9*, 639–643. [CrossRef] [PubMed]
26. Kleinman, D.; Insler, V.; Leiberman, J.R.; Glezerman, M.; Albotiano, S.; Potashnik, G.; Meizner, I. Acid Phosphatase Levels in Follicular Fluids Following Induction of Ovulation in in Vitro Fertilization Patients. *J. In Vitro Fertil. Embryo Transf.* **1987**, *4*, 181–184. [CrossRef]
27. Chen, F.; Spiessens, C.; D'Hooghe, T.; Peeraer, K.; Carpentier, S. Follicular Fluid Biomarkers for Human in Vitro Fertilization Outcome: Proof of Principle. *Proteome Sci.* **2016**, *14*, 17. [CrossRef] [PubMed]
28. Estes, S.J.; Ye, B.; Qiu, W.; Cramer, D.; Hornstein, M.D.; Missmer, S.A. A Proteomic Analysis of IVF Follicular Fluid in Women ≤32 Years Old. *Fertil. Steril.* **2009**, *92*, 1569–1578. [CrossRef] [PubMed]
29. Jones, K.A.; Kim, P.D.; Patel, B.B.; Kelsen, S.G.; Braverman, A.; Swinton, D.J.; Gafken, P.R.; Jones, L.A.; Lane, W.S.; Neveu, J.M.; et al. Immunodepletion Plasma Proteomics by TripleTOF 5600 and Orbitrap Elite/LTQ-Orbitrap Velos/Q Exactive Mass Spectrometers. *J. Proteome Res.* **2013**, *12*, 4351–4365. [CrossRef] [PubMed]
30. Bian, Y.; Zheng, R.; Bayer, F.P.; Wong, C.; Chang, Y.-C.; Meng, C.; Zolg, D.P.; Reinecke, M.; Zecha, J.; Wiechmann, S.; et al. Robust, Reproducible and Quantitative Analysis of Thousands of Proteomes by Micro-Flow LC–MS/MS. *Nat. Commun.* **2020**, *11*, 157. [CrossRef] [PubMed]

31. Sun, R.; Hunter, C.; Chen, C.; Ge, W.; Morrice, N.; Liang, S.; Zhu, T.; Yuan, C.; Ruan, G.; Zhang, Q.; et al. Accelerated Protein Biomarker Discovery from FFPE Tissue Samples Using Single-Shot, Short Gradient Microflow SWATH MS. *J. Proteome Res.* **2020**, *19*, 2732–2741. [CrossRef] [PubMed]
32. Wiśniewski, J.R. Filter-Aided Sample Preparation for Proteome Analysis. In *Microbial Proteomics: Methods and Protocols*; Becher, D., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2018; pp. 3–10. ISBN 9781493986958.
33. Carvalho, L.B.; Capelo-Martínez, J.-L.; Lodeiro, C.; Wiśniewski, J.R.; Santos, H.M. Ultrasonic-Based Filter Aided Sample Preparation as the General Method to Sample Preparation in Proteomics. *Anal. Chem.* **2020**, *92*, 9164–9171. [CrossRef] [PubMed]
34. Cox, J.; Hein, M.Y.; Luber, C.A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **2014**, *13*, 2513–2526. [CrossRef] [PubMed]
35. Skowrońska, P.; Kunicki, M.; Pastuszek, E.; Konieczna, L.; Bączek, T.; Łukaszuk, K. Follicular Fat-Soluble Vitamins as Markers of Oocyte Competency. *Syst. Biol. Reprod. Med.* **2020**, *66*, 112–121. [CrossRef] [PubMed]
36. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology; Balaban, B.; Brison, D.R.; Calderón, G.; Catt, J.; Conaghan, J.; Cowan, L.; Ebner, T.; Gardner, D.; Hardarson, T.; et al. The Istanbul consensus workshop on embryo assessment: Proceedings of an expert meeting. *Hum. Reprod.* **2011**, *26*, 1270–1283. [CrossRef]
37. Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips. *Nat. Protoc.* **2007**, *2*, 1896–1906. [CrossRef] [PubMed]
38. Tirumalai, R.S.; Chan, K.C.; Prieto, D.A.; Issaq, H.J.; Conrads, T.P.; Veenstra, T.D. Characterization of the Low Molecular Weight Human Serum Proteome. *Mol. Cell. Proteomics* **2003**, *2*, 1096–1103. [CrossRef]
39. Wiśniewski, J.R. Quantitative Evaluation of Filter Aided Sample Preparation (FASP) and Multienzyme Digestion FASP Protocols. *Anal. Chem.* **2016**, *88*, 5438–5443. [CrossRef]
40. Lewandowska, A.E.; Macur, K.; Czaplewska, P.; Liss, J.; Łukaszuk, K.; Ołdziej, S. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High PH RP-HPLC Fractionation. *J. Proteomics* **2019**, *191*, 131–142. [CrossRef] [PubMed]
41. Gundry, R.L.; White, M.Y.; Murray, C.I.; Kane, L.A.; Fu, Q.; Stanley, B.A.; Eyk, J.E.V. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Curr. Protoc. Mol. Biol.* **2010**, *90*, 10–25. [CrossRef]
42. Zhang, Y.; Bilbao, A.; Bruderer, T.; Luban, J.; Strambio-De-Castillia, C.; Lisacek, F.; Hopfgartner, G.; Varesio, E. The Use of Variable Q1 Isolation Windows Improves Selectivity in LC–SWATH–MS Acquisition. *J. Proteome Res.* **2015**, *14*, 4359–4371. [CrossRef]
43. Fel, A.; Lewandowska, A.E.; Petrides, P.E.; Wiśniewski, J.R. Comparison of Proteome Composition of Serum Enriched in Extracellular Vesicles Isolated from Polycythemia Vera Patients and Healthy Controls. *Proteomes* **2019**, *7*, 20. [CrossRef]
44. Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [CrossRef] [PubMed]
45. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
46. Oliveros, J.C. (2007–2015). Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. Available online: https://Bioinfogp.Cnb.Csic.Es/Tools/Venny/Index.Html (accessed on 14 June 2021).
47. Deutsch, E.W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.; Campbell, D.S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; et al. The ProteomeXchange Consortium in 2017: Supporting the Cultural Change in Proteomics Public Data Deposition. *Nucleic Acids Res.* **2017**, *45*, D1100–D1106. [CrossRef]
48. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [CrossRef] [PubMed]
49. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13*, 731–740. [CrossRef] [PubMed]

Attachment 2

This attachment contains the statements of contribution of all the authors of the publications, which the thesis was based on (see List of publications, page 7).

Gdańsk, 20.12.2021

Aleksandra E. Bogucka, MSc.
Laboratory of Biopolymers Structure
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology.** ***Journal of Proteome Research*****. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

  - investigation (i.e., affinity chromatography and ultrafiltration fractionation, in-solution protein digestion, peptide purification, execution of mass spectrometry measurements), formal analysis (i.e., protein identification, spectral library construction, SWATH-MS data processing, statistical, functional, and comparative literature analysis), visualization, and writing – original draft, review and editing of the manuscript.

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation.** ***Journal of Proteomics*****. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - conceptualization, methodology (i.e., high pH RP-HPLC fractionation experiments), investigation (i.e., affinity chromatography, ultrafiltration, and high pH RP-HPLC fractionation, in-solution protein digestion, peptide purification, execution of mass spectrometry measurements), formal analysis (i.e., protein identification, spectral libraries construction, SWATH-MS data processing, quality comparison of constructed spectral libraries), visualization, and writing – original draft, review and editing of the manuscript.

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid.** ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - conceptualization, methodology (i.e., Triple Quad-TOF workflow optimization), investigation (i.e., affinity chromatography, ultrafiltration, and high pH RP-HPLC fractionation, in-solution, FASP, and MED-FASP protein digestion, STAGE Tips peptide purification, execution of mass spectrometry

measurements on the TripleTOF 5600+ instrument), formal analysis (i.e., protein identification, spectral library construction, SWATH-MS data processing, absolute protein concentration calculation by TPA, statistical, functional, and comparative literature analysis), visualization (i.e., charts and figures created using MS Excel, Venny, STRING, and Cytoscape), and writing – original draft, review and editing of the manuscript.

Gdańsk, 20.12.2021

Prof. Paulina Czaplewska, Ph.D.
Laboratory of Mass Spectrometry
Core Facility Laboratories
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology.** ***Journal of Proteome Research*****. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation.** ***Journal of Proteomics*****. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - conceptualization, resources (i.e., access to necessary materials and equipment, including the TripleTOF 5600+ mass spectrometer), funding acquisition, project administration, and writing – review and editing of the manuscript.

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid.** ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - resources (i.e., access to necessary materials and equipment, including the TripleTOF 5600+ mass spectrometer), project administration, supervision, and writing – review and editing of the manuscript.

Międzyuczelniany Wydział
Biotechnologii UG-GUMed
Zespół Laboratoriów Specjalistycznych
Paulina Czaplewska
dr hab. Paulina Czaplewska, prof. UG

Gdańsk, 20.12.2021

Anna Fel-Tukalska, MSc.
Laboratory of Biopolymers Structure
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publication my contribution involved:

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid.** ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - investigation (i.e., proteolytic digestion of human follicular fluid samples by FASP and MED-FASP methods, peptide purification in STAGE Tips procedure, execution of all mass spectrometry measurements on the Q Exactive HF-X instrument) and writing – review and editing of the manuscript.

Anna M. Tukalska

Gdańsk, 21.12.2021

Joanna Liss, Ph.D.
Department of Medical Biology and Genetics
Faculty of Biology
University of Gdańsk

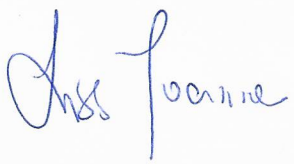
STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology.** ***Journal of Proteome Research*****. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation.** ***Journal of Proteomics*****. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - resources (i.e., collection and preparation of the human follicular fluid clinical material for further analysis) and writing – review and editing of the manuscript.



Gdańsk, 21.12.2021

Prof. Krzysztof Łukaszuk, M.D., Ph.D.
Division of Obstetric and Gynaecological Nursing
Institute of Nursing and Midwifery
Faculty of Health Sciences with the Institute of Maritime and Tropical Medicine
Medical University of Gdańsk

STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology. *Journal of Proteome Research*. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation. *Journal of Proteomics*. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - conceptualization, resources (providing access to the human follicular fluid clinical material and acquisition of the ethical approval required for the project execution), supervision (including guidance on the results analysis in a clinical setting), and writing – review and editing of the manuscript.

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid. *International Journal of Molecular Sciences*. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - resources (providing access to the human follicular fluid clinical material and acquisition of the ethical approval required for the project execution), supervision (including guidance on the results analysis in a clinical setting), funding acquisition (OPUS 14 financial grant: UMO-2017/27/B/NZ5/02393), and writing – review and editing of the manuscript.

Gdańsk, 20.12.2021

Katarzyna Macur, Ph.D.
Laboratory of Mass Spectrometry
Core Facility Laboratories
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology.** ***Journal of Proteome Research*****. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation.** ***Journal of Proteomics*****. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - methodology (i.e., all LC-MS/MS methods development on the TripleTOF 5600+ mass spectrometer coupled with microLC, sample preparation and data analysis workflow development), investigation (partial execution of mass spectrometry measurements on the TripleTOF 5600+ instrument), and writing – review and editing of the manuscript.

Katarzyna Macur

Gdańsk, 20.12.2021

Prof. Stanisław Ołdziej, Ph.D.
Laboratory of Biopolymers Structure
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publications my contribution involved:

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Qualitative and Quantitative Analysis of Proteome and Peptidome of Human Follicular Fluid Using Multiple Samples from Single Donor with LC–MS and SWATH Methodology.** ***Journal of Proteome Research*****. 2017; 16 (8):3053-3067. https://doi.org/10.1021/acs.jproteome.7b00366**

- **Lewandowska AE*, Macur K, Czaplewska P, Liss J, Łukaszuk K, Ołdziej S*. Human Follicular Fluid Proteomic and Peptidomic Composition Quantitative Studies by SWATH-MS Methodology. Applicability of High pH RP-HPLC Fractionation.** ***Journal of Proteomics*****. 2019; 191:131-142. https://doi.org/10.1016/j.jprot.2018.03.010**

  - conceptualization, methodology (development of the general workflow applied in the studies and data analysis strategies), resources (reagents, instrumentation, and computing resources), supervision (main supervision of the project), funding acquisition, and writing – original draft, review and editing of the manuscript.

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid.** ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - conceptualization, methodology (i.e., the Triple Quad-TOF workflow development and choice of data analysis strategy), supervision (main supervision of the project), funding acquisition (partial financing of conducted experiments and publication fee), and writing – original draft, review and editing of the manuscript.

Stanisław Ołdziej

Gdańsk, 20.12.2021

Marcel Thiel, MSc.
Laboratory of Biopolymers Structure
Intercollegiate Faculty of Biotechnology UG-MUG
University of Gdańsk

STATEMENT

I declare that in the following publication my contribution involved:

- **Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid.** ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

  - formal analysis (i.e., planning and carrying out of data analysis procedures using Python 3.8 programming language), visualization (i.e., preparation of charts and tables using Python 3.8), and writing – review and editing of the manuscript.

Max-Planck-Institut of Biochemistry
Proteomics and Signal Transduction

MPI für Biochemie • Am Klopferspitz 18 • D-82152 Martinsried (near Munich)

Prof. Dr. Jacek R. Wiśniewski

Phone: +49-(0)89 / 8578-2205
jwisniew@biochem.mpg.de

December 22, 2021

STATEMENT

I declare that in the publication

Lewandowska AE*, Fel A, Thiel M, Czaplewska P, Łukaszuk K, Wiśniewski JR, Ołdziej S*. Compatibility of Distinct Label-Free Proteomic Workflows in Absolute Quantification of Proteins Linked to the Oocyte Quality in Human Follicular Fluid. ***International Journal of Molecular Sciences*****. 2021; 22(14):7415. https://doi.org/10.3390/ijms22147415**

my contribution involved conceptualization, methodology (i.e., the Quad-Orbitrap workflow and absolute protein quantification by TPA development), resources (all means required for the experiments conducted by the Quad-Orbitrap workflow, including reagents, instrumentation, and computing resources), supervision, funding acquisition (by Max-Planck Society for the Advancement of Science), and writing – review and editing of the manuscript.

Am Klopferspitz 18
D-82152 Martinsried bei München

Tel.: +49-(0)89 / 8578-1
Fax: +49-(0)89 / 8578-3777

mpib@biochem.mpg.de
www.biochem.mpg.de