

# Chapter 2

## Proteomics

Ricardo P. Bertolla

### 2.1 Introduction

Proteomics, the comprehensive identification and quantification of proteins present in a fluid or a cell population (James 1997), has become an important and established tool in understanding the mechanics of biological systems (Bantscheff et al. 2012), as proteins themselves are postgenomic effectors of cellular and molecular events (Zhang et al. 2013). Moreover, proteins and peptides may act as messengers, or interact upstream (controlling gene expression) or downstream in the Omics cascade, in order to regulate cell function or the extracellular environment (Zhang et al. 2013). In the case of complex samples, in which there is extensive crosstalk between different cell types and the environment—and the reproductive system is a great example of such a relationship—understanding the role these proteins may play, as well as potential regulators of their functions, is of great significance (Del Giudice et al. 2013; Amann 1989; Regassa et al. 2011).

Proteomics experiments may be of an exploratory or confirmatory nature. In exploratory (untargeted) proteomics, extensive mapping of digested peptides from a proteome is performed, in order to identify and quantify as many proteins as possible. This approach, coined shotgun proteomics (Claassen 2012), may be performed using chemical labeling of proteins or peptides in order to differentiate groups or may be label-free. A similar approach enriching postgenomic modifications may also be performed in order to observe post-translational protein modifications such as sumoylation, phosphorylation, oxidation, glycosylation, acetylation, and

---

R.P. Bertolla (✉)  
São Paulo Federal University, São Paulo, Brazil  
e-mail: rbertolla@yahoo.com

ubiquitination, among others (Olsen and Mann 2013). In confirmatory (targeted) proteomics, a number of techniques may be employed, ranging from conventional immunological labeling of differentially expressed proteins (i.e., Western blotting and enzyme-linked immunosorbent assay [ELISA]) (Bianchi et al. 2013; Ambekar et al. 2015; McReynolds et al. 2014) to mass spectrometry (MS)-based techniques (Lange et al. 2008).

Recent advances in MS techniques as well as in orthogonal protein and peptide separation technologies have dramatically increased throughput, sensitivity, dynamic range, and reproducibility of proteomics studies (Bantscheff et al. 2012). Coupled with advances in downstream data analysis, proteomic studies have impacted studies in the clinical sciences, with the promise of a personalized medicine tailored to an individual's disease (as opposed to the current generalized approach) (Cline et al. 2007; Hood and Flores 2012). This is especially true of the so-called bottom-up proteomics studies, in which proteins are broken down into peptides, and this complex mixture of peptides is separated by ultra-high pressure liquid chromatography (UPLC) and analyzed through tandem mass spectrometry (MS/MS) for sequencing of peptides against an *in silico* database constructed from sequenced genomes (Zhang et al. 2013).

However, it remains a major challenge of proteomics studies to translate findings complex in design, statistical analysis, and technical settings to answer clinical/biological questions. Bearing this in mind, this chapter is set up to discuss untargeted shotgun proteomics studies, including the usual steps in a proteomics experiment, ranging from sample preparation to separation techniques, MS equipment, and finally to the univariate and multivariate data analyses steps, as well as to functional enrichment tests usually performed for shotgun proteomics studies. Finally, we will also discuss results validation through a number of different techniques.

## 2.2 Methods Used in Proteomics

MS-based proteomics has profoundly impacted the field of proteomics in general, because of increased dynamic range, possibility to integrate with chromatographic separation equipment, increased sensitivity, and high quality identification of proteins (Bantscheff et al. 2012; Aebersold 2003; Aebersold and Mann 2003; Mallick and Kuster 2010). Mass spectrometers are able to detect the masses of ions [or the mass-to-charge ratios ( $m/z$ )] in a gas phase. In order to do so, mass spectrometers are generally subdivided into (i) an ionization source, (ii) mass analyzer(s), and (iii) a detector, of which the first two will be further discussed [for a review, see (Scherl 2015)].

Ionization of proteins and peptides was made possible with the development of the so-called soft ionization techniques; ionization techniques which lead to little or no fragmentation of the generated ion within the ion source, of which are of special note the matrix-assisted laser desorption/ionization (MALDI) (Tanaka et al. 1988) and electrospray ionization (ESI) (Fenn et al. 1989) techniques, both of which have been applied to proteomics studies. For shotgun proteomics studies, ESI allows for direct integration with liquid chromatography equipment coupled to tandem mass spectrometers (LC-MS/MS), which adds a quantitative aspect to the studies (Whitehouse et al. 1985).

Mass analyzers select, separate, or analyze ions based on their  $m/z$  values. Quadrupole (Q) mass analyzers, for example, select ions by varying an electric potential using radio frequency and direct current, focusing on ions of interest and diverting ions that are to be removed. Time-of-flight (TOF) analyzers base ion separation on the time it takes for an ion (of a specific mass and charge state) to fly through a vacuum tube. Ion traps are able to trap ions in a chamber and, by varying the applied electric field, allow for expulsion of single (or few) ions at a time. Finally, orbitrap (OT) mass analyzers detect  $m/z$  values by analyzing the cycle resonance of ions around a central electrode (Scherl 2015).

Current high-end mass spectrometers for shotgun proteomics studies utilize hybrid mass spectrometers, in which the first analyzer is a selector (such as a quadrupole or an ion trap), followed by a second analyzer which is converted into a collision chamber, followed by a high resolution mass analyzer, such as a TOF or OT. The collision chamber is used to generate fragments (fragment ions) of selected ions in the first analyzer (parent ions), and to determine the masses of these fragments, in an approach termed tandem mass spectrometry (MS/MS) (Scherl 2015). This is discussed in greater depth in the “*Identification of proteins using mass spectrometry*” section. A similar setup, utilizing a scanning quadrupole mass analyzer in the first step for separation of one or few parent ions and targeting specific fragments in the second analyzer may be used for specific protein detection/quantification, in selective reaction monitoring (SRM), multiple reaction monitoring (MRM), or parallel reaction monitoring (PRM) studies (Lange et al. 2008; Gallien et al. 2015). This will be explored further in the “*Validation of mass spectrometry data*” section.

In terms of setting up a general workflow for shotgun proteomics studies, the following steps are usually followed: (i) sample acquisition, storage, and preparation; (ii) protein and peptide labeling and separation; (iii) tandem mass spectrometry; (iv) database searching for protein identification/quantification; and (v) data analysis (Bantscheff et al. 2012). Enrichment of post-translational modifications may be performed, and a number of labeling approaches may be employed in order to increase sample normalization.

## 2.3 Sample Preparation

Upon sample collection, much care should be taken with regards to inhibiting proteases. This is especially important for analysis of post-translational modifications (PTMs), as many are subject to alterations due to protease activity (Olsen and Mann 2013). However, in the case of studies that are performed on ejaculated semen in humans, a special comment should be made with regards to semen coagulation/liquefaction. Coagulation occurs mainly due to semenogelin-1 and -2 polymerization. Prostatic Kallikrein-3 will then cleave these proteins at specific sites (at the N-side of Tyrosine residues), ultimately leading to semen liquefaction (Mitra et al. 2010). Thus, semen samples from humans have been subjected to protease activity (albeit specific and for a relatively short amount of time) even before the sample is available for separation of sperm/seminal plasma.

Protease inhibitors may be added individually to specific classes (such as metalloproteinase inhibitors), or commercially available protease inhibitor cocktails may be used (Clifton et al. 2011). Samples are then processed by separating cells (and cellular debris) from the extracellular fluid (seminal plasma, blood plasma, follicular fluid, etc.). This is usually achieved by an initial centrifugation at a lower force (not more than 600  $\times$ G) in order to avoid cell lysis and contamination of the supernatant fluid with cellular proteins. The supernatant is then centrifuged at maximum speed for a longer period of time (around 1 h) at cold temperatures (4 °C) to remove any remaining cellular debris. Next, the cell pellet can be washed a couple of times to remove any contaminating extracellular fluid, bearing in mind that lower centrifuge force is recommended to avoid cell lysis and loss of soluble proteins. The sample may then be kept frozen (at  $-70$  °C or in liquid nitrogen) until sample preparation.

Sample preparation is different for cells compared to that for fluids. Cells must be lysed and proteins extracted. There are a number of different protocols for cell protein extraction, such as protocols for the extraction of membrane proteins, nuclear proteins, and whole cell lysate, among others (Tanca et al. 2013; Ahmed 2009; Weston et al. 2013; Intasqui et al. 2013a; Wisniewski et al. 2009). For fluids, most of the time protein extraction is not necessary, instead precipitation of proteins in order to remove contaminating lipids or sugars may be performed (Lo Turco et al. 2010; Camargo et al. 2013; Intasqui et al. 2013b).

The next step in the proteomics workflow is to consider options to decrease sample complexity. Current dynamic ranges of modern mass spectrometers utilized for shotgun proteomics are at or just above 4 orders of magnitude (according to their manufacturers), while protein dynamic range may achieve up to 9, or even 12 orders of magnitude (Anderson and Anderson 2002; Corthals et al. 2000). This means that highly concentrated proteins decrease the technical capability of detecting proteins at lower concentration. In seminal plasma, we have found that around 10 proteins represent 80 % of the proteome (*unpublished data*). In blood plasma (and its transudates, such as follicular fluid), albumin itself may represent 50 % of the proteome (Corthals et al. 2000). Thus, in order to detect (and quantify)

proteins on the lower end, protein depletion and sample fractionating are two options that ensure a more comprehensive coverage of the proteome. Protein depletion is achieved by using columns which will bind to the desired protein (that is to be removed). There are several commercially available columns to remove albumin, immunoglobulins, and a number of other proteins (Zhang et al. 2013).

Protein fractionating may be performed in one-dimensional sodium dodecyl sulfate polyacrylamide (1D SDS-PAGE) gels, where they are usually separated by molecular mass, and strips of the gels containing a portion of the proteome are subjected to downstream analysis (Kim et al. 2003). Another option is pre-fractionating according to protein isoelectric point—the pH at which their net charge is zero—(isoelectric focusing) in solid (Pernemalm and Lehtio 2013) or in liquid phase (Zuo and Speicher 2002). In either case, sample complexity is decreased, which leads to the possibility of observing proteins present in smaller amounts. Another orthogonal approach for protein fractionation is two-dimensional SDS-PAGE (2D SDS-PAGE), in which proteins are separated initially according to their isoelectric point in strip gels, and these are then submitted to gel electrophoresis for separation according to their molecular mass (Lopez 2007; Westermeier 2014).

2D SDS-PAGE experiments often target downstream MS protein identification to the differentially expressed gel spots (proteins). However, 2D SDS-PAGE is limited by the fact that (i) variability is quite high between different gels, and (ii) dynamic range is lower than that of LC-MS/MS platforms. Inter-gel variability has been dealt with by tagging proteins by fluorescent markers and then pooling samples from different groups to run in a same gel (Two-dimensional fluorescence difference gel electrophoresis—2D-DIGE), and dynamic range issues have been dealt with by depleting highly enriched proteins (Lopez 2007; Westermeier 2014). Also, an important advantage of 2D SDS-PAGE is that the initial mass of the protein spots is known before MS identification. Therefore, it is possible to know if the observed protein was intact or modified by cleavage, or by formation of doublets, for example. In seminal plasma, this may allow for visualization of specific semenogelin digests that present different biological roles (Robert and Gagnon 1999; Tomar et al. 2013; Mitra et al. 2010).

## 2.4 Identification of Proteins Using Mass Spectrometry

Because whole protein mass spectrometry is still limited, in terms of generating protein fragments for identification, protein samples need to be digested into peptides for identification and quantification (Bruce et al. 2013; Cox and Mann 2011). Trypsin is usually the enzyme of choice, because it produces peptides that fall into an optimal mass range (Vandermarliere et al. 2013). Trypsin cleaves proteins at the C-terminus of lysine and arginine residues, unless these are followed by a proline at the C-terminal end. Cleavage is highly specific and current protocols are quite efficient, leading to few missed cleavage sites (Vandermarliere et al. 2013). If a

label-free approach is being employed, immediately following digestion, peptides are injected into the LC-MS/MS system (Bantscheff et al. 2012; Wisniewski et al. 2009; Cox and Mann 2011). If a labeling approach is chosen, peptides may be subjected to a number of different labels, such as dimethyl (multiplex peptide stable isotope dimethyl labeling) or isobaric tag for relative and absolute quantification (iTRAQ) (Bantscheff et al. 2012; Zhang et al. 2013; Wisniewski et al. 2009; Bruce et al. 2013; Cox and Mann 2011). This will be further discussed in the “*Quantitative proteomics*” section.

Proteins are inferred in LC-MS/MS experiments based on the mass of the peptides (MS parent ion, generated in the first mass analyzer of a hybrid mass spectrometer) and on the masses of detected fragment ions (MS/MS), usually reporting to specific amino acid sequences (detected in the second mass analyzer of a hybrid equipment) (Bantscheff et al. 2012; Aebersold and Mann 2003; Scherl 2015). In this data-dependent approach, an initial scan is performed in order to indicate MS peaks for fragmentation, followed by a step in which each peak of interest is filtered in the first mass analyzer, fragmented in a collision induced dissociation (CID) chamber, and its fragment ions detected in the second mass analyzer (Bantscheff et al. 2012; Scherl 2015; Bruce et al. 2013). The generated MS and MS/MS spectra are compared against in silico-generated mass values of putative proteins (based on sequenced genomes), leading to the identification of shared peptides and unequivocal peptides, the latter of which report to a single protein (Bantscheff et al. 2012; Cox and Mann 2011; UniProt 2015). A similar approach is used in data-independent experiments, in which the first mass analyzer is switched off, and the CID chamber alternates between high and low collision energies, generating parent and fragment ions with no need to filter out unwanted MS peaks (thus with loss of information) (Law and Lim 2013).

## 2.5 Quantitative Proteomics

Quantification of proteins in a sample, either in absolute terms or relative to another sample, is of importance if differential expression is to be studied. In order to achieve quantification, a few different approaches may be used, of which are of special note the label-free quantification (Washburn et al. 2001; Liu et al. 2004; Bantscheff et al. 2007), labeled quantification and absolute quantification methods.

In label-free experiments, quantitative information is extracted by (i) quantifying the total mass spectrometry signal for each peptide of a given protein (in other words, quantitation directly from constructed ion chromatograms generated during LC-MS/MS) or by (ii) determining the number of fragment spectra which report to peptides of a given protein (spectral counting) (Bantscheff et al. 2007, 2012; Washburn et al. 2001; Liu et al. 2004). Label-free shotgun proteomics has improved dramatically with the development of software packages specialized in aligning chromatograms from different runs and designed to normalize experiment-wise in terms of the amount of protein injected (Bantscheff et al. 2012; Zhang et al. 2013;

Cox and Mann 2008). Moreover, developments in chromatographic separation have greatly increased peptide resolution (in retention time), decreasing co-elution, and allowing for quantification of a large number of identified peptides (Ow et al. 2011; Delmotte et al. 2007).

In order to decrease intra-sample variability, labeled quantification experiments have been successfully devised. This may be possible through metabolic labeling or through chemical labeling. Stable isotope labeling with amino acids in cell culture (SILAC) is one of the best examples of metabolic labeling (Hoedt et al. 2014; Chahrour et al. 2015; Ong et al. 2002). In this type of experiment, labeled amino acids are fed to cells in *in vitro* culture conditions and, after a few rounds of division, it is assumed that translated proteins have incorporated these amino acids which, according to the label, will report to either one or the other condition. Thus, after protein extraction and digestion, samples may be mixed and analyzed in a single LC-MS/MS run, which in turn decreases inter-assay variation (Hoedt et al. 2014; Chahrour et al. 2015; Ong et al. 2002).

For studies on biological fluids or tissues not submitted to culture, chemical labeling of amino acids with isobaric markers, such as in iTRAQ, has allowed for simultaneous LC-MS/MS runs of up to eight different conditions (Chahrour et al. 2015; Choe et al. 2007; Ross et al. 2004). A similar, less costly, approach is the labeling of peptides using dimethyl labeling, which allows up to three-plex analysis in a simultaneous LC-MS/MS run (Boersema et al. 2009). A number of other labels are available, such as Tandem Mass Tags (TMT), Isotope-coded affinity tag (iCAT), among many more (Chahrour et al. 2015).

Finally, absolute quantification of proteins in shotgun proteomics studies have usually been performed by spiking into the sample previously digested peptides from one or more proteins of a different species, at a known concentration (Bruce et al. 2013). While this is not an error-free absolute quantification *per se* (absolute quantification in targeted proteomics requires spiking-in of peptides specific to the target protein), this allows for an internal standard normalization and a stoichiometric calculation for the studied proteome (Bruce et al. 2013).

## 2.6 Differentially Expressed Proteins

Regardless of the method of protein quantification, statistical analysis of the generated proteome should be performed under careful considerations. Initially, our group has utilized both a univariate and a multivariate approach to complex samples (Intasqui et al. 2013a, b, 2015, 2016; Camargo et al. 2013; da Silva et al. 2013; Lo Turco et al. 2013; Antoniassi et al. 2016; Del Giudice et al. 2016). In a univariate approach, statistical differences may be performed using standard parametric tests (Student's T-test, One-way ANOVA, etc.), provided that the data are robust (enough biological replicates, standard statistical assumptions for normality and homoscedasticity, which refers to homogeneity of variance). If not, the use of non-parametric tests (Mann-Whitney, Kruskal-Wallis, etc.) is a valid alternative.

Moreover, setting a minimal fold-change value may be desired, in order to filter out noise differences. Some studies report minimal fold-change values as low as 2 (or 0.5), although other studies have demonstrated that quantitative data utilizing spectral counting has been able to detect fold-change differences as low as 1.4. Other quality control assumptions may be used, such as quantification results derived from at least two different peptides. A complete list of the Minimum Information about a Proteomics Experiment (MIAPE) Mass Spectrometry Quantification (MIAPE-Quant) is under constant development and update by the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI) (Martinez-Bartolome et al. 2013).

Finally, in a multivariate data analysis approach, extraction of components using a Partial Least Squares Discriminant Analysis (PLS-DA) has been performed in the study of the seminal plasma proteome. These components are then subject to a binary logistic calculation which ultimately calculates odds ratio for group detection for each protein, as well as builds a multiplex data model to calculate group distribution according to protein quantification levels. This is particularly useful to suggest diagnostic models, although downstream absolute quantification studies will be necessary (Intasqui et al. 2015, 2016).

## **2.7 Proteomic Identification of Post-translational Modifications (PTMs)**

One of the main mechanisms of biological variability is the wide array of PTMs present in proteins (Zhang et al. 2013; Olsen and Mann 2013; Clamp et al. 2007). Proteomic identification of PTMs—unbiased identification of multiple modifications in multiple proteins of a single cell type or fluid—presents additional challenges, because most PTMs are present in relatively low amounts, which leads to the need to enrich these modifications (Olsen and Mann 2013). Other than an enrichment step, a step to avoid degradation or artificial creation of PTMs, most of the proteomics workflow for the study of PTMs remains similar to a conventional proteomics workflow (Olsen and Mann 2013).

MS-based proteomics is the standard approach to identify PTMs in a proteomics workflow. However, enrichment of PTMs remains an important challenge, because, while for some modifications, such as phosphorylation and glycosylation, a relatively high efficiency is possible, other modifications usually require a larger amount of protein and protocol development (Olsen and Mann 2013). However, the promise of identifying multiple PTMs as associated to a specific biological condition is quite important, as it confers a functional aspect to a shotgun proteomics study. Indeed, differential protein expression does not shed light on all aspects of biological variability, while PTMs add an important functional modulation information to the proteome (Olsen and Mann 2013).



## 2.8 Interpretation of Protein Profiles Using Bioinformatics

Current proteomics studies are able to identify and quantify a high number of proteins—up to 5000 or 10,000 proteins have been identified from proteomes (Nagaraj et al. 2011; Beck et al. 2011). In seminal plasma, at least 2600 proteins have been identified with high confidence MS studies, over at least 4 orders of magnitude (da Silva et al. 2016). While this leads to a comprehensive biological view, it adds the need to filter out results which do not explain the biological question under study. In silico-derived protein–protein interaction (PPI) networks, functional enrichment of gene ontology terms and biological pathways in differentially expressed clusters, and multivariate statistical analysis are important approaches which may aid in filtering out a true signal from proteomic noise (Cline et al. 2007; Intasqui et al. 2013a, b; Camargo et al. 2013; Bindea et al. 2009).

In silico-derived protein–protein interaction networks are constructed based on large PPI databases, such as IntAct (Orchard et al. 2014), BioGRID (Stark et al. 2006), HPRD (Keshava Prasad et al. 2009), and STRING (Szklarczyk et al. 2015), among many others. Several software suites, such as MetaCore™ (Thomson Reuters, New York, NY, USA), Ingenuity Pathway Analysis® (IPA, QUIAGEN, Redwood City, CA, USA), and the open source platform Cytoscape (Cline et al. 2007), have made it possible to overlay identified proteins in a study to these interactomes, in order to suggest PPI subnetworks in any given study. Our suggested workflow is to (i) upload the list of identified proteins, (ii) identify PPIs between identified proteins, (iii) determine clusters specific/overexpressed in each group, and (iv) submit these clusters to functional enrichment studies. We have previously performed this workflow for the study of follicular fluid (Lo Turco et al. 2010), in seminal plasma (Intasqui et al. 2013b; Camargo et al. 2013; da Silva et al. 2013), and in sperm (Intasqui et al. 2013a) under different conditions.

For functional enrichment studies, functional annotations should be considered. The Gene Ontology (GO) Consortium has constructed a functional annotation database for a number of different species. Proteins, as gene products, inherit these functional annotations, which for the GO Consortium is subdivided into three major groups: Cellular Component, Biological Process, and Molecular Function. Each protein may receive a number of GO annotation terms, and these are shared by multiple proteins (Ashburner et al. 2000). If a cluster presents a higher frequency of GO terms than the whole GO database (or than a reference database, or when compared to the other group), these terms are said to be functionally enriched (Bindea et al. 2009; Maere et al. 2005). Many other ontology databases may be used, such as PantherDB (Mi et al. 2013), and highly curated databases exclusive to software packages, such as in MetaCore™ and Ingenuity Pathway Analysis®. Also, proteins may be annotated with pathway information, from such databases as Reactome (Croft et al. 2014), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000), and PantherDB (Mi et al. 2013), among others. These pathways may indicate upstream and downstream interactions between the studied proteins and other molecules, such as lipids, sugars, and metabolites. Moreover,

metabolic activity of the cell, and its assumed pathways brings information on response to the environment as well as on cell properties, such as senescence, aging, cell cycle activity, protein turnover (Merelli et al. 2014). This is particularly interesting when integrating metabolomics/lipidomics/steroidomics studies into proteomics studies. The Metscape app within Cytoscape, for example, utilizes pathway databases to generate a protein/metabolite/drug network with experimental data (Gao et al. 2010). Integration of these pathways may demonstrate not only the differentially expressed protein(s), but also its effects (such as metabolites) and effectors (such as steroid hormones) (Rouillard et al. 2015).

## 2.9 Validation of Mass Spectrometry Data

Validation of proteomics data usually falls into two major categories: validation through immuno-detection (Brown et al. 2013; Rifai et al. 2006; Surinova et al. 2011; Kohler and Seitz 2012; Kingsmore 2006) and validation using targeted mass spectrometry techniques (Lange et al. 2008; Gallien et al. 2015). Upstream validation (mRNA expression levels) and downstream validation (measurement of metabolites involved in the enriched pathways) are also important to consider as the former demonstrate the protein production cascade involved in the differential proteome, and the latter demonstrates that the assumed proteomic pathway is indeed active (Cline et al. 2007; Hood and Flores 2012).

Immunodetection may be performed on extracted proteins from the same sample or from a confirmatory prospective cohort, or may be performed directly on the studied cells, using fluorescent microscopy or flow cytometry. Of the many immune-detection methods available for extracted proteins, conventional ELISA are of special note (Brown et al. 2013; Rifai et al. 2006; Surinova et al. 2011; Kohler and Seitz 2012). In Western blotting, proteins are separated in 1D SDS-PAGE gels, transferred onto membranes which bind the proteins, and detected by the use of antibodies against the protein of interest. These intensities are usually normalized to constitutive proteins in the samples, in order to achieve lower intra- and inter-assay variation. ELISA, on the other hand, detects the protein(s) of interest by direct binding in microplates pretreated with antibodies against the protein of interest, signaling the presence, and intensity of the target protein (Scherl 2015).

Multiplexing immune-detection techniques are also currently available, utilizing the Luminex<sup>®</sup> system, which is an interesting development to consider for proteomics results validation, as they allow for the simultaneous intensity-based quantification of up to 50 different proteins (Purohit et al. 2015). However, while the available kits cover important pathways (such as oxidative stress), customization of detection kits for specific proteins is still prohibitive in terms of cost.

For detection of proteins directly on cells, two important techniques are immunohistochemistry (for tissue sections) or immunocytochemistry (for cells in suspension). Using fluorescently labeled antibodies allows for detection of

populations marked or unmarked for the target protein (Erikson et al. 2007a, b). This is an interesting approach because cell integrity tests, such as membrane integrity, DNA fragmentation, or reactive oxygen species activity, may often be performed on the same sample, adding cellular biology information to proteomics results.

Mass spectrometry studies utilizing selected reaction monitoring (SRM), multiple reaction monitoring (MRM), and parallel reaction monitoring (PRM) are very sensitive (able to sense up to attomolar ( $10^{-18}$  mol/L) concentrations), specific, and allow for absolute protein quantification (Lange et al. 2008; Gallien et al. 2015). A specific workflow needs to be designed for each protein of interest, however, and in practice, this can lead to a long discovery pipeline (Bantscheff et al. 2012). In this type of experiment, a specific peptide from the target protein is filtered in a first analyzer of a mass spectrometer, and a (or a few) specific fragment(s) is detected on the second analyzer, leading to very high confidence quantification of that specific protein. In this experiment, calibration curves utilizing different quantities of the purified (or synthesized) target peptide (of the target protein) allow for determination of absolute quantities in each sample (Lange et al. 2008; Gallien et al. 2015).

Finally, it should be mentioned that current shotgun proteomics techniques utilizing high-end mass spectrometers produce results which, in themselves, are high confidence. While quantitative information from these experiments is subject to a number of variation factors, often coefficients of variation (CVs) from these studies fall between 20 % and 30 % (Piehowski et al. 2013). However, it is important to document specific conditions in which identification and quantification calculations were made, in order to guarantee that results are indeed high confidence (Martinez-Bartolome et al. 2013). In lacking a true validation, many studies report peptide mass errors, minimum number of peptides identified/protein, b- and y-sequencing results and false discovery rates, and this information allows for determination of MS conditions which, if stringent, would render reproducible results (Martinez-Bartolome et al. 2013, 2014).

## 2.10 Final Remarks

Proteomics is a promising and exciting area of research for integration of systems biology, which is especially applicable to complex samples and for integration into studies focused on the Omics cascade, ranging from mutations and polymorphisms through gene expression, protein expression, and down to metabolite generation and control of epigenetic factors. These systems biology studies aim at allowing for a more personalized medicine, and stem from the concept of phenotypic singularities due to genotypic potential within a given environment. The number of potential environmental agents is immeasurable, and individual response is potentially as widespread. Therefore, focusing on observing specific molecular fingerprints which associate to a biological condition may help to characterize disease. Generation of single biomarkers for diagnosis is one of the possible results,

but most likely these studies will determine panels of differentially expressed proteins that associate to a phenotype, in that different nodes (proteins and other biomolecules) will weigh the biological network away or closer to a healthy state. In the case of infertility, which is often not a binary disease, the intensity of deviation from a healthy state of each individual in a couple will likely allow for understanding the causal factors of infertility and, potentially, therapeutic targets for intervention.

Moreover, while maintaining a Cartesian characteristic (in that it is hypothesis-driven), shotgun proteomics generates multiple hypotheses for downstream studies. On the other hand, much care should be taken, as is usually the case with technology-driven research platforms, in order to adhere to standards of quality that are under constant improvement. In the case of shotgun proteomics studies, the large amount of generated data is difficult to manually curate, which renders quality control of data acquisition fundamental, in order to increase repeatability. Data analysis and interpretation is also essential in order to transform lists of proteins into intra-assay variability is still quite high, due to technical limitations in peptide separation and identification. Targeted proteomics is, therefore, crucial for translation of results into actionable targets. It may well be that these results will in the future achieve the true nature of translational medicine in which the patient will benefit from novel diagnostic tests and potential druggable targets.

Proteomics in Human Reproduction

Biomarkers for Millennials

Agarwal, A.; Samanta, L.; Bertolla, R.P.;

Durairajanayagam, D.; Intasqui, P.

2016, XV, 124 p. 9 illus., Softcover

ISBN: 978-3-319-48416-7