

Chapter 21

Secretome Proteomics, a Novel Tool for Biomarkers Discovery and for Guiding Biomodulatory Therapy Approaches

Verena Paulitschke, Rainer Kunstfeld, and Christopher Gerner

Abstract Secretome analysis represents a novel technology for biomarker discovery based on proteome profiling of proteins secreted by both primary tumor cells and tumor associated cells. Tumor cells are able to establish a permissive and supportive environment for survival and cell growth and to facilitate invasion and metastasis by modulating the stromal host compartment. The onset of these characteristic events seems to precede tumor progression. Due to the leaky nature of newly formed blood vessels and the increased hydrostatic pressure within tumors, secreted proteins are most plausibly shed into the blood. Thus, proteins specifically secreted by these cells may serve as early disease biomarkers. Biomarker candidates identified by secretome proteomics combined with the application of appropriate bioinformatic tools can then be validated in human plasma/sera. Besides biomarker discovery secretome analysis will also shed light on mechanisms of tumor progression offering novel targets for therapeutic intervention. The tumor-stroma cell cooperativity is reversible and may thus be directly accessible to therapeutic intervention. In conclusion, secretome proteomics offers new insights into the pathophysiology of tumor progression, and allows the identification of novel biomarkers and of new drug targets.

Keywords Tumor-associated stroma cells • Secretome profiling • Biomodulatory therapy • Mass spectrometry • Primary human cells • Bioinformatics

Abbreviations

AFP	Alpha-fetoprotein
CAF	Cancer associated fibroblast
CEA	Carcinoembryonic antigen

V. Paulitschke (✉) and R. Kunstfeld
Department for Dermatology, Medical University Vienna,
Währinger Guertel 18–20, 1090, Vienna
e-mail: verena.paulitschke@meduniwien.ac.at

C. Gerner
Med. Universität Wien, Universitätsklinik für Innere Medizin 1, Institut für Krebsforschung,
Borschkegasse 8a, A – 1090, Wien, Austria

CPL/MUW – database	Database of the Clinical Proteomics Laboratories at the Medical University of Vienna
DIGE	Differential in-gel electrophoresis
Gpm	Global proteome machine organisation
ICAT	Isotope coded affinity tag
LMW	Low-molecular-weight
MIAPE	Minimum information about a proteomics experiment
PRIDE	PRoteomics IDentifications database
PSA	Prostate specific antigen
ROC	Receiver operating characteristic
SILAC	Stable isotope labeling by amino acids in cell culture
SOP	Standard operating procedure
SVM	Support vector machines
TIF	Tissue interstitial fluid

21.1 Introduction

21.1.1 *The Proteome*

The proteome, first defined by Williams in 1996 [1], is the protein complement of genomic functionality and is defined as the set of proteins which are present in a cell, tissue or organism. The proteome is highly dynamic and may respond to almost any kind of environmental stimuli, most obviously it varies according to cell type and functional state of cells. The proteome in a body fluid, cell, tissue, or organism represents only a subset of all possible gene products at a certain point of time and cannot be directly predicted from gene expression. Proteins may exist in multiple varieties due to posttranslational modifications which affect protein structure, localization, function and turnover. These specific changes may reflect immediate and characteristic changes in response to disease processes. Especially the low-molecular-weight (LMW) range proteome is believed to be very useful for analysis of disease progression and response to treatment [2].

21.1.2 *Clinical Proteomics*

The goal of clinical proteomics is to obtain the most comprehensive insight into pathophysiological conditions derived from protein expression profiles as they occur in vivo. Proteins play a fundamental role in controlling multiple functions within a cell's organization. They serve as building materials, enzymes and biological transport machines, as well as sensors processing and transferring information. Cells consist of thousands of proteins executing diverse operations, not only highly

coordinated, but also dependent upon each other. Cells may newly produce specific proteins when they encounter challenges for specific functions. When cells encounter unusual situations, they try to adjust to it by expressing proteins which may help to deal with the new situation. Such proteins, specifically synthesized on demand, may indicate characteristic disease states and may thus serve as diagnostic markers. Detection of such aberrations in protein expression in diseased tissues may lead to a better understanding of the cellular pathology and thereby support the development of new therapeutic strategies. Therefore, proteins have attracted attention to biomarker discovery: One of the central applications of proteomics has become the classic protein biomarker discovery and the uncovering of functional tumor-associated systems stages, e.g. inflammation, neoangiogenesis, proliferation behaviour and others.

Clinical proteomics focuses on the analytical and clinical implementation and validation of novel biomarkers and aims to gain a better understanding of disease processes which may support the implementation of novel treatment options. Therefore it is critically dependent on high-throughput analysis platforms which have to provide reproducible and reliable protein patterns, bioinformatics tools for data comprehension and interpretation. Furthermore it has to refer to a well-defined patient cohort including all necessary anamnestic and physiologic parameters for instance age, sex, hormonal status and treatment. Sample collection and biobank organization have to be SOP-driven. The samples should be rapidly analyzed since transportation and storage may lead to artifacts like selective damage or aggregation of specific cell subpopulations or shedding of cell surface markers. To collect comprehensive information about sample technical analyses such as genomics, metabolomics, lipidomics, glycomics, transcriptomics, flow cytometry with definition of specific cell populations may be combined [2].

As a matter of fact, despite of intensive efforts in proteomics in the recent years, few novel disease biomarkers have been discovered. Since 1998 the rate of introducing newly approved protein targets has been declining to an average of one per year in the USA [3,4]. Therefore, novel analysis models and procedures have to be defined for biomarker discovery, which are highlighted in this review.

21.1.3 Metastasis and Tumor Microenvironment

Especially in oncology novel biomarkers are urgently needed. Due to metastasis cancer is a major cause of mortality worldwide with ten million new cases and more than six million deaths per year [5]. Early detection of incipient remodeling processes indicating metastatic progression and the development of appropriate therapeutic approaches may substantially improve patient survival.

The tumor microenvironment consists of a multi-faceted spectrum of highly specialized cell types, e.g. mesenchymal cells, myelomonocytic cells, endothelial cells and immune cells. The metastatic process is decisively driven by stromal processes, particularly facilitated by neoangiogenesis, lymphangiogenesis and accompanying

inflammatory processes. Growth factors secreted by the stromal cells may serve as survival factors for cancer cells [6]. The tumor microenvironment, through the process of aberrant cell growth, cellular invasion and altered immune system function, contributes a unique sum of proteins secreted, with cytokine and chemokine or enzymatic activity (for example, matrix metalloproteinases) [7,8]. This generates an unbalanced or altered stoichiometry of agonists and antagonists within the tumor profile compared to the ‘normal’ milieu and can provide characteristic fingerprints applicable as specific and sensitive biomarkers for various purposes [9].

21.2 Biomarker

21.2.1 Definition

A biomarker is objectively measurable indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.

Different types of biomarker can be evaluated: prognostic, which characterize the course of disease, predictive to monitor the response to treatment, diagnostic which demonstrate the evidence of disease and pharmacodynamic for the purpose to show efficacy of treatment.

A surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint, a characteristic or variable that reflects how a patient feels, functions, or whether he is going to survive.

A surrogate endpoint is expected to predict clinical benefit such as decreased pain, quality of life, DFS (disease free survival), OS (overall survival) and cure.

Cancer biomarkers have to enhance the potential to screen, diagnose, prognosticate, localize and stage tumors, or predict and monitor the therapeutic responses to various cancers. Therefore cancer biomarkers have to be correlated with the clinical situation and can be classified into four broad categories related to tumor burden, cancer risk, tumor-host interaction and function.

21.2.2 Biomarker in Cancer

Metastatic cancer presents a substantial clinical challenge since there is a lack of adequate approaches to properly define disease subgroups for rational treatment design and selection. In addition the majority of cancers are initially diagnosed in advanced stages. Some important markers commonly employed in clinical diagnosis include CEA (carcinoembryonic antigen), PSA (prostate specific antigen), AFP (alpha-fetoprotein), CA 125, CA 15–3, and CA 19–9. Current diagnostic methods are limited in their ability to diagnose early disease and accurately predict individual

risk of disease progression and outcome. None of these markers is known to have high specificity and sensitivity or to exhibit prognostic value for neoplasms [10]. This may be attributed to the high heterogeneity in cancer patients with a lot of varying parameters such as tumor size, location, histology, depth, stage, grade, ulceration, age, sex etc. The emerging pattern of molecular complexity in tumors mirrors the clinical diversity of the disease. This highlights that cancer is not a single disease but a heterogeneous group of disorders that arise from complex molecular changes [11]. Thus, there is a growing consensus that marker panels, which are more sensitive and specific than any individual marker, will increase the accuracy of early-stage cancer detection.

21.2.3 Stages of Biomarker Development

The **discovery phase** represent an ‘unbiased’ experimental setup, here high-throughput methods are of outstanding relevance. The next phase, ‘**qualification**,’ serves for the confirmation that the differential expression of candidate proteins observed in the discovery phase can be verified using alternative, targeted methods. In addition the differential expression of candidate biomarkers has to be verified human plasma/serum samples. During the discovery and qualification phase the consistency of association between marker and disease and the marker sensitivity and specificity has to be demonstrated. In the ‘**verification**’ phase the analysis has to be extended to a larger number of human plasma samples, incorporating a broader range of cases and controls. Here the environmental, genetic, biological and stochastic variation in the population has to be considered. In the verification phase the sensitivity of biomarker candidates is affirmed and specificity has to be assessed [3].

21.2.4 Proteomic Technology in Biomarker Discovery

Important sources for biomarkers should be represented by proteins in the blood. The exact number of proteins in blood is not known. Efforts by different laboratories of the Plasma Proteome Project led to the identification 889 proteins identified with a confidence level of at least 95%. It is estimated that the plasma proteome may contain up to 10,000 proteins [12]. Proteome analysis is a promising tool for the discovery of novel and innovative cancer biomarkers [13]. Over the past decade, serum and plasma proteomics aimed to identify potential cancer biomarkers [14]. Since these markers are present in low amounts in blood samples, the direct isolation requires a labor-intensive process involving the depletion of abundant proteins and extensive protein fractionation.

This classical approach comparing the plasma protein profiles of the healthy donor to the patient largely failed during the discovery phase. An inherent problem

of blood proteomics is the **complexity** of the protein composition, comprising an enormous diversity of proteins and protein isoforms, the dynamic range of plasma and other biofluids and the tremendous extend of human and disease variation. In addition the anticipated **low relative abundance** of many disease-specific biomarkers represents a pitfall: the concentration range in human plasma covers ten orders of magnitude, which means that certain biomarkers may be ten billion fold less abundant than serum albumin. Due to these pitfalls of blood proteomics it has been proposed to rather analyze diseased tissue or biological fluids close to diseased sites (for example tissue interstitial fluid (TIF)). Here the relevant proteins are expected to occur at higher concentrations which facilitates biomarker discovery.

Alternatively, the secretome of cancer cells [15] and tumor associated cells can be analyzed and verified subsequently in human blood by ELISA analyses. Following completion of the Human Genome Project, scientists postulated that important cancer biomarkers will be secreted proteins, as about 20–25% of all cell proteins are secreted [16]. Actually some classical cancer biomarkers (e.g., CEA, Her2-neu) are cell-membrane bound, with their extracellular domains eventually shed into the circulation [14].

21.3 Secretome as Reservoir for Biomarker Discovery

21.3.1 Definition

The secretome is defined as the set of secreted proteins [17,18]. The term “secretome” was first referred by Tjalsma et al. [17] to secreted proteins of *Bacillus subtilis* in a genome-based global survey. The secretome is composed of proteins that are actively secreted, shed from the cell surface and intracellular proteins, which are accidentally released into the supernatant. Cell lysis resulting from necrosis releases relatively large amounts of protein when compared to secretion. The secretome harbors proteins released by a cell, tissue or organism through various mechanisms including classical and nonclassical secretion as well as secretion via exosomes [19]. Secretion may occur either constitutively (continuously) or be regulated and triggered on demand resulting from different functional cell states.

21.3.2 The Cancer Secretome

The cancer secretome, the totality of proteins released by cancer cells, has been attracting wide attention as it is a potential reservoir of cancer biomarkers. Secreted proteins may determine, control and coordinate many biological processes such as growth, cell division and differentiation, invasion, metastasis, angiogenesis and lymphangiogenesis via an endocrine, paracrine or autocrine way. In addition it is known that the tumor microenvironment contributes to tumor development and

progression via communicative processes, mediated by cytokines, chemokines, hormones and specifically secured communication structures (e.g. gap junctions) [8]. Therefore also secreted proteins shed by tumor associated cells need to be considered [9]. Protein secretion exerts autocrine and paracrine biological functions rather than maintenance of basic metabolism. Therefore, specifically secreted proteins may much better be related to the exertion of biological functions compared to cytoplasmic proteins. These proteins eventually end up in the bloodstream, and thereby may have a potential as non-invasive biomarkers [9]. Their biological key roles make them good targets and sources for therapeutical and drug-based intervention as well as tools for diagnosis and prognosis. Thus, great interest is currently focused on the characterization of secreted proteins in order to identify novel biomarkers. The leaky nature of newly formed blood vessels and the increased hydrostatic pressure within tumors increase the chance to find secreted proteins in the blood stream [9]. A pathological situation thus tends to push molecules from the tumor interstitium into the circulation. Therefore it seems to be plausible that proteins produced by the microenvironment will be shed into the blood, making ongoing processes of tumor development detectable [9]. Combinations of markers that are indicative for the specific interactions of the tumor tissue microenvironment will achieve higher specificity and higher sensitivity than the application of any single marker. Candidate biomarkers are expected to exist at very low concentrations diluted in blood plasma with highly abundant proteins such as albumin, which exist in billion-fold excess. At early stages of disease, cancer-specific proteins will always constitute an evanescent subfraction of the proteome representing a true analytical challenge. Noteworthy, early-stage disease lesions such as carcinoma in situ represent tumor cell numbers hardly exceeding several thousand cells. However, the affected microenvironment comprises many more cells compared to the number of tumor cells. Thus proteins derived from tumor associated stroma cells will be produced by more cells and may accumulate to higher amounts. Consequently it can be expected that such proteins will be better accessible for diagnostic purposes than proteins derived from cancer cells themselves. Secretome analysis is applicable to cultured cells as well as tissue specimens [9]. The most comprehensive analysis results, however, are obtained in case of isolated and cultured cells.

In contrast to secreted proteins as new candidates for blood biomarkers, specific proteins identified in the cytoplasm rather represent biomarker candidates accessible to immunohistochemical analysis. Cytoplasmic proteins also comprise specific indicators of functional cell states and cell activities. Combining the information of both secreted and cytoplasmic proteins further supports the detailed understanding of complex patho-physiological processes.

21.3.3 Development of Rational Therapy Design by Secretome Analysis

For many years, the main principle in the treatment of metastatic cancer has been the cyclic administration of high-dose chemotherapy, which is a unselective

strategy based on cytotoxic effects [20]. Chemotherapy uses the small window between killing of rapidly dividing cancer cells and sparing healthy tissues. All tissues with a high proliferation rate are affected by chemotherapy leading to severe and dose limiting side effects such as myelosuppression, damage of the intestinal mucosa and severe skin reactions. Due to this issue, cycles of therapy have to be interrupted by drug-free periods to allow normal tissue to recover. Although the initial effects of chemotherapy are often quite impressive in terms of depleting tumor mass, the duration of remission is often short and resistance may be induced. This risk of selecting chemoresistant cell clones can be linked to the genetic instability and the high mutational rates and heterogeneity of tumor cells. In order to overcome this drug resistance, doses of chemotherapy can either be increased; intervals shortened or chemotherapeutic combination strategies can be chosen. All these options are subsequently potentiating side effects [9].

For an accurate, individualized assessment of risk of disease progression it was suggested to classify disease subgroups and rationally select treatments to substantially affect the outcome of advanced disease. Sekulic et al. [11] discuss that the low overall response rates observed in clinical trials that rely on clinical disease features for patient selection might simply reflect a relatively low percentage of patients with the disease susceptible to a given therapeutic agent or combination. As a consequence, patient selection for clinical trials and selection of therapy on the basis of individual molecular attributes might be necessary to improve response rates to any kind of therapy. Sekulic et al. propose that the detailed consideration of each single patient will overcome the problems of heterogeneity and may lead to a new classification by genomic techniques [11]. Newer individual sequencing data, however, suggest that the heterogeneity of genetic aberrations even within a single patient is by far too large to enable patient stratification. Another stratification option may be derived from the specificity of protein expression profiles which are largely dependent on functional states of cells. Cells make proteins in order to fulfil specific tasks. Functional activation, therefore, inevitably results in the expression of a protein cluster dedicated to fulfil the newly requested functions. Specific pathologic processes may, therefore, be characterized by functional protein signatures. These proteins, here designated as functional protein signatures, may thus enable the identification of relevant functional cell states. In contrast to the genomic techniques focusing on hereditary predisposition, proteome analysis is able to detect when and to what extent the risks have become manifest. For characterisation of diseases, functional aberrations causative for the disease have to be distinguished from aberrations resulting from these primary functional aberrations. To give an example, uncontrolled proliferation is a common process characteristic for neoplasia. The detection of a common process will not support disease sub-classification. Different kinds but characteristic stressors such as inflammatory activation, oxidative stress, DNA damage or ER stress, however, may be causative for disease states such as uncontrolled proliferation. Each kind of stressor is specifically detectable by a defined protein signature providing the basis for functional disease classification. Understanding and detecting the variety of mechanisms leading to a common pathology may serve patient stratification aiding rational therapeutic

concepts better than the consideration of downstream consequences of pathological processes. As a consequence, protein clusters rather than single proteins will serve as biomarkers. Such application may be more feasible than individual genetic profiling to support optimal therapeutic decisions.

In search for alternative strategies for the treatment of advanced cancer, targeting the tumor stroma seems to be a promising tool since this approach is not cytotoxic but interferes with the cooperativity of tumor and tumor stroma cells. This concept is based on the improving understanding that tumor development is associated with the transformation of normal stroma into an “activated” stroma phenotype. Tumor cells are able to establish a permissive and supportive environment for survival and cell growth and to facilitate invasion and metastasis by modulating the stromal host compartment. Targeting this interference between tumor and tumor stroma may consistently lead to a reduction of tumor growth and metastasis. The targets in this approach are genetically normal activated cells which will not be able to escape therapy due to genetic instability and clonal selection. Therefore, targeting these cells should lead to a reduction of development of resistance. This strategy is also considered to be less toxic and thus allows sustaining the therapeutic pressure continuously over longer time periods [9]. Considering that the stroma provides proteins supporting tumor survival, a blockage of this process might chemosensitize the tumor. Therefore, this approach might serve as an efficient combination therapy with chemotherapeutic agents. The enhanced knowledge generated by secretome analysis of molecular aberrations involving important cellular processes, such as cellular signaling networks, regulation of cell cycle and cell death, will contribute to better diagnosis, accurate assessment of prognosis, patient stratification and rational design of effective therapeutics.

21.3.4 Clinical Application

Secretome analysis aims to address three important features of clinical proteomics [9]:

1. Tumor cells may recruit stromal cells for the secretion of growth factors which serve as powerful survival factors. The onset of these characteristic events seems to precede tumor progression. These secreted proteins may have a good chance entering the bloodstream, due to the leaky nature of newly formed blood vessels and the increased hydrostatic pressure within the tumors. Stroma cell secretion of bioactive molecules, which may serve as **diagnostic biomarkers**, are early events in carcinogenesis and may thus enable the early detection of cancer progression.
2. Proteome profiling may identify molecular signatures of processes which promote metastasis. Secretome analysis of defined cell populations offers the opportunity to identify the contribution of the involved cell types and thus the underlying **pathomechanisms**. These pathways rather than single proteins should be monitored and targeted.

3. Transformation of cancer cells is an irreversible process which may be corrected only by apoptotic cell death. Tumor therapy usually targets cancer cells; modern therapy concepts include targeting the stroma in an anti-angiogenic and anti-inflammatory fashion. Cooperativity contributed by stromal cells is reversible and thus directly accessible to **therapeutic intervention**. Most importantly, stroma derived survival factors shall be decreased resulting in a higher chemosensitivity of the tumor cells. Detailed understanding of the responsible processes may thus enable the design of completely new therapeutic strategies.

21.4 Methods

To gain reliable insights into the cancer secretome it is obligatory to prepare samples which are clearly defined and as pure as possible. Secreted proteins occur in body fluids, the direct analysis of potential marker proteins from such samples is hindered by the high complexity and dynamic range of resident plasma proteins. A cell is the smallest independent protein synthesis unit, therefore a reduction of sample complexity to single cell types greatly improves the chances to identify low abundant proteins. It has been observed that proteins secreted by tumor cells in vitro may very well reflect the proteins secreted by tumors in vivo [21]. Therefore, the routine method used is to analyze the secreted of tumor cells or tumor stroma cells in vitro [21]. Mbeunkui et al. [22] performed a comprehensive study of the secretome of three metastatic cancer cell lines and demonstrated that an incubation time of 24 h and 60–70% cell confluence were considered as optimal cell incubation conditions (Fig. 21.1). Due to the low abundance of secreted proteins, the contamination by non-secreted proteins may mask the proteins of interest. The discrimination of genuine secreted proteins from non-secreted proteins is a major issue that needs to be answered in every single experiment [21].

In addition, secreted proteins present in the culture media usually occur at low concentrations, which is often below the ng/mL range. These proteins should be concentrated before proteomics analysis [21]. Ultrafiltration can be used for the concentration of the secretome [21]. Alternatively, precipitation can be performed with acetone or ethanol.

21.4.1 2D-gel Electrophoresis

Zwickl et al. [23] have established a metabolic labeling-based technology with [³⁵S]-labelled methionine and cysteine which allows for the sensitive and selective detection of secreted proteins. They demonstrated the applicability of this method by a study on secretome profiles of a hepatocellular carcinoma-derived cell line. These cells were incubated in the presence of [³⁵S]-labelled methionine and cysteine. Subsequently, the cell supernatant was filtered, precipitated and subjected to

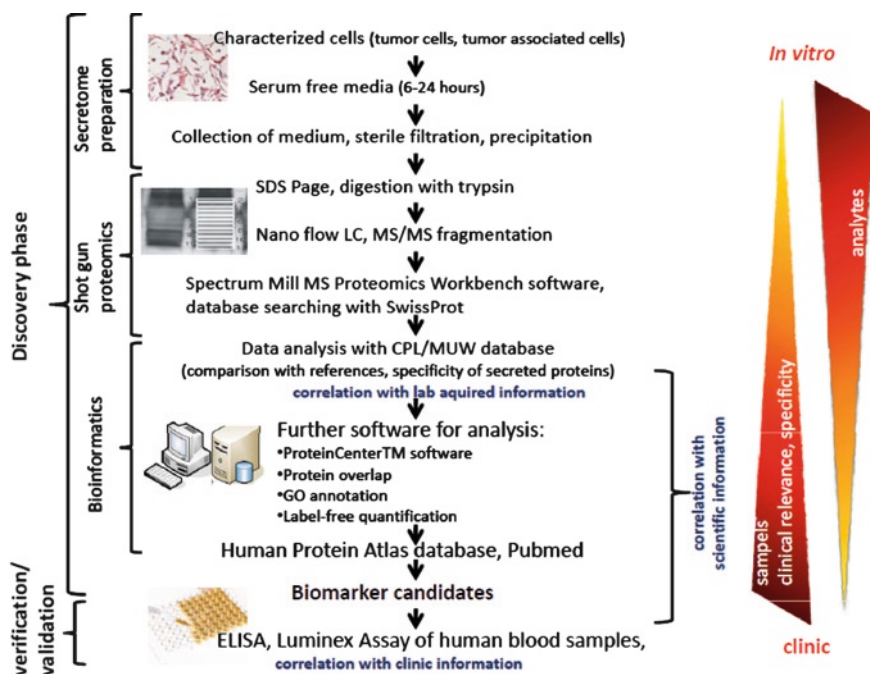


Fig. 21.1 Workflow of secretome proteomics. Secretome preparation is performed with well-characterized tumor or tumor associated cells. Supernatant collection, sterile filtration and precipitation is performed after 6–24 h incubation of the cells in special formulated serum free media. For shot gun proteomics the protein samples are separated by SDS-gel electrophoresis followed by tryptic in-gel digestion and peptide separation by nano-flow LC. Peptide identification is accomplished by MS/MS fragmentation analysis and the MS/MS data are interpreted by the Spectrum Mill MS Proteomics Workbench software and searched using the UniProt Database. Biomarker candidates are selected considering own laboratory and public available expert information. In the verification and validation phase performing ELISA studies in human blood samples these candidates are correlated with clinic information. Specificity and clinical relevance is increased starting from *in vitro* to clinic while the number of analytes is decreased

two-dimensional gel electrophoresis. After staining proteins were detected by fluorescence staining and autoradiography. Fluorescence staining detects all proteins, in contrast autoradiography detected only those proteins synthesized and secreted by living cells during the metabolic labeling period. All identified 16 protein spots in autoradiography were found to be authentic secreted proteins.

The disadvantages of 2-DE are the low sensitivity in the detection of proteins in low concentrations, the poor representation of hydrophobic membrane proteins in 2D-gels, furthermore the technique is time-consuming, labor-intensive and has a relatively low efficiency in protein detection due to limited amenability to automation [21]. To circumvent some of these inherent problems of the standard 2-DE procedure, a modified method, differential in-gel electrophoresis (DIGE) has been developed by GE Healthcare [24], where three charge and mass-matched fluorescent dyes (Cy2, Cy3 and Cy5), are utilized. These dyes can primarily combine covalently with lysine.

Different protein samples are differently labeled by these fluorescent dyes, then mixed and visualized in one gel. DIGE reduces the experimental variations using one gel for three samples [19]. Instead this method is not applicable to those proteins without lysine (in case of minimal dyes) or cysteine (in case of saturation dyes).

21.4.2 Mass Spectrometry

A mass spectrometer consists of three components: (a) an ion-producing source, (b) a mass analyzer to measure the mass-to-charge ratio (m/z) of the ionized molecule, and (c) a detector that registers the number of ions. A typical shotgun proteomic experiment generally consists of five stages: (1) proteins present in cell lysates, tissue or body fluids are separated by fractionation or affinity selection to define the subproteome, (2) enzymatic degradation of proteins to peptides by trypsin, (3) peptides are separated by reversed phase nano-flow HPLC and eluted into an electrospray ion source where they become charged single molecules in the gas phase which may enter the MS. Isotope-labeling methods, such as isotope coded affinity tag (ICAT) and stable isotope labeling by amino acids in cell culture (SILAC), can be used to introduce quantitative aspects in cancer secretome analysis [25]. These label based approaches are expensive, time-consuming and not always feasible due to the limitation of available tags for primary human materials [25]. We have started to systematically analyze secretomes of various primary and cultured human cells [9,26]. Therefore we have standardized a procedure to bioinformatically filter the truly secreted proteins from contaminant proteins regarding the known main contaminants, i.e. cytoplasmic proteins and serum proteins and as well regarding signal peptides characteristic for secreted proteins. Secreted proteins are then classified with respect to cell type specificity and their relation to functional cell states which are investigated in vitro by functional activation. The relation of identified proteins to the most plausible cells of origin as supported by the CPL/MUW database [27] greatly facilitates the interpretation of complex proteome profiles as derived from human serum samples (Figs. 21.1 and 21.2).

The applied standard procedure to analyse secretomes is detailed in the following (Fig. 21.1). For the accumulation of secreted proteins cells are incubated in serum-free specialized media formulations for 6–24 h at 37°C. For isolation of the secreted protein fraction, the cell supernatant is collected, sterile filtrated to remove cellular debris and precipitated by the addition of ethanol. For the isolation of the corresponding cytoplasmic proteins, all buffers are supplemented with protease inhibitors. Cells are lysed in hypotonic lysis buffer and pressed through a 26 g syringe in order to open the cells by rupture. The cytoplasmic fraction is separated from the nuclei by centrifugation and precipitated by the addition of ethanol. All protein samples are dissolved in sample buffer (7.5 M urea, 1.5 M thiourea, 4% CHAPS, 0.05% SDS, 100 mM DDT) and separated by SDS-gel electrophoresis followed by tryptic in-gel digestion. For shotgun analysis, peptides are separated by nano-flow

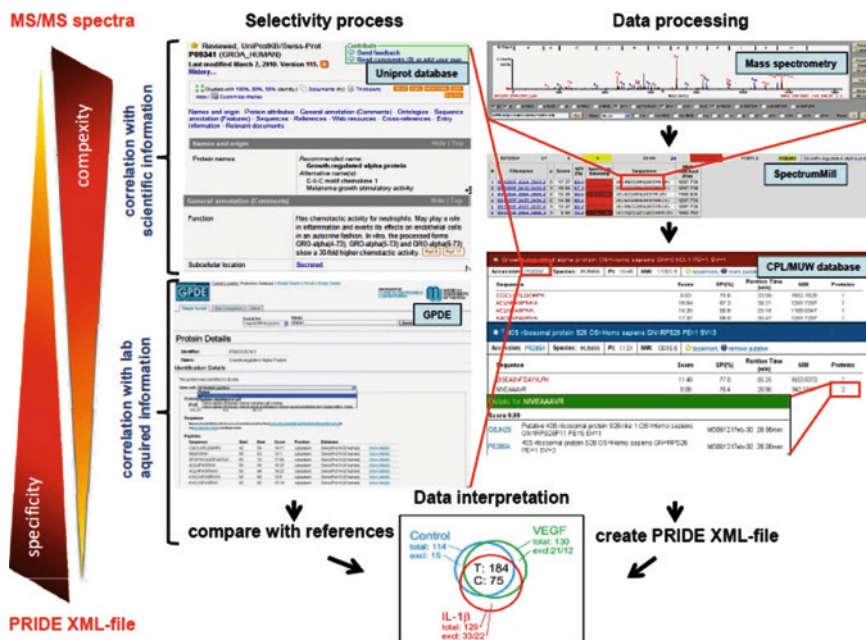


Fig. 21.2 All proteome identification data are based on peptide fragmentation spectra. Blast search of each peptide reveal the corresponding proteins. All peptides related to a single protein become sorted accordingly. Ambiguity may arise due to partial sequence similarities of different proteins, which may not allow to assign a peptide to a single protein only. UniProt and the CPL/MUW database assist in the selection of the most plausible candidate. Data of various experiments are combined to obtain reference maps of single cell types at specific states. The specificity of any single protein expression with respect to cell types can be retrieved using the GPDE. Overlap and specificity of proteome maps can be visualized by accurate Venn diagrams. During this process specificity is increased while complexity is decreased

LC (1100 Series LC system, Agilent, Palo Alto, CA) using the HPLC-Chip technology (Agilent) equipped with a 40 nl Zorbax 300SB-C18 trapping column and a 75 $\mu\text{m} \times 150\text{ mm}$ Zorbax 300SB-C18 separation column at a flow rate of 400 nl/min, using a gradient from 0.2% formic acid and 3% ACN to 0.2% formic acid and 50% ACN over 60 min. Peptide identification is accomplished by MS/MS fragmentation analysis with an ion trap mass spectrometer (XCT-Ultra, Agilent) equipped with an orthogonal nanospray ion source. The MS/MS data are interpreted by the Spectrum Mill MS Proteomics Workbench software (Version A.03.03, Agilent) and searched against the SwissProt Database (UniProt Version 15.4 containing 20,328 protein entries) (Figs. 21.1 and 21.2) allowing for precursor mass deviation of 1.5 Da, a product mass tolerance of 0.7 Da and a minimum matched peak intensity (%SPI) of 70%. Due to previous chemical modification, carbamidomethylation of cysteines is set as fixed modification. The reliability of peptide identifications from MS/MS spectra relates to spectral quality indicated with specific scores. The scores are essentially calculated from sequence tag lengths, but also mass deviations are

considered. To assess the reliability of the peptide identifications, searches are performed against the corresponding reversed database. Further details are accessible via www.meduniwien.ac.at/proteomics.

A protein fraction may be contaminated with keratins derived from dust and comprise identifications with questionable identification quality. To make appropriate decisions, we make use of lists of common contaminants as well as reference lists dependent on the kind of sample comprising “expected” proteins. Only those putative identifications are included, which are present in the according reference list, while all other are discarded. The resulting protein profile is classified using the CPL/MUW database to support subsequent data interpretation (Fig. 21.1). Classification considers common housekeeping proteins, cell type-specific proteins and proteins related to the exertion of specific functions. Furthermore, other public available data as the gene ontology (GO) can be included. Protein expression data derived from methods other than mass spectrometry such as Protein Atlas and gene expression data may support the final decision for expression specificity and thus choice of biomarker candidates. Such biomarker candidates have to be verified and validated performing ELISA studies with human blood samples and by correlation with clinic data. Specificity and clinical relevance is increased starting from in vitro to clinic while sample size is decreased (Fig. 21.1).

21.5 Bioinformatics

Proteomes of biological samples typically consist of thousands of different proteins with a concentration range spanning nine or more orders of magnitude [28]. Only technically demanding high-throughput technologies such as mass spectrometry may actually cope with such an analytical challenge [29]. Modern machines produce more than 10,000 peptide fragmentation spectra per hour, piling up to huge amounts of data for each experiment. As a consequence, there is no proteome profiling without the assistance of well-performing computers and sophisticated bioinformatics tools.

A typical workflow to analyse proteomics data would consist of several independent but interrelated steps. These include interpretation of spectra, subsequent protein identifications and quantifications as well as the assignment of specifically expressed proteins based on comparative analysis. While several different and powerful software packages exist to support these steps such as Mascot [30], SEQUEST [31] and Spectrum Mill [32], there is still urgent demand for further improvements. In the following, the implications of each step will be presented in more detail.

To begin with more technical aspects, there is still a broad variety of data formats and protein sequence databases which complicate the exchange and comparison of data generated by different laboratories. It was the initiative of the European Bioinformatics Institute to establish a common data format, PRIDE-XML, which can be realized starting from almost any kind of existing data format. To support the dissemination of complex proteome data, the public data repository

PRIDE (PRoteomics IDentifications database, <http://www.ebi.ac.uk/pride>) was installed [33]. The Global Proteome Machine Organisation (gpm) at gpmdb.rockefeller.edu was established to improve the quality of proteome analysis data relying on tandem mass spectrometry, to make results portable and to provide a common platform for testing and validating proteomics results [34]. These important tools provide access to thousands of proteome analysis experiments and supports documentation of published data.

To summarize, clinical proteomics needs standard operating procedures and guidelines for data generation, data analysis and validation of datasets [35] since the biomarker discovery has suffered in the past from inconsistent data acquisition, statistical interpretation and validation [36]. These standards are represented by (1) the use of standards in the data format and storage (mzXM/mzData), (2) by public data repositories (Peptide Atlas, PRIDE, SwissProt/Uniprot and (3) the integration of a complex database including biological information and different bioinformatic programs using to link different protein lists for instance to specific pathways [2].

Data mining strategies fall into two categories: unsupervised (analogous to clustering) and supervised (analogous to classification) such as classification and regression trees and support vector machines (SVM) [36]. Each algorithm has inherent strengths and weaknesses, which must be matched to the different statistical problems [36]. Some of these softwares are (Fig. 21.1):

1. ProteinCenter software, a proteomics data mining and management software, can be used to predict the function of the identified proteins based on universal GO annotation terms. Here a comparison of cell line secretomes with each other and a functionally categorization can be performed [36,37].
2. The SignalP program can be used to determine the presence of secretory signal peptide sequences and thus predict potential secretion.
3. The SecretomeP program offers the possibility to predict non-signal peptide-triggered protein secretion and to distinguish between protein secretion pathways-the classical and non classical pathway [37].
4. MetaCore (GeneGo, St. Joseph, MI) is used for biological network building and describe millions of relationships between proteins, according to publications on proteins and small molecules including direct protein interactions, transcriptional regulation, binding or enzyme-substrate interactions [37].

In the process of biomarker discovery, a single biomarker may hardly provide sufficient specificity; often several biomarkers have to be combined. Here a two-step process is required:

1. Biomarkers have to be identified employing statistics for multiple testing.
2. They are combined in a predictive model using some of the algorithms [36].

Support Vector Machines (SVMs) offer a cross-validated predictive statement, which is an important issue in biomarker combination. In the case of making a predictive diagnosis through the combination of biomarker, it is possible to calculate the level of confidence with a classification algorithm. Two basic considerations

have to be applied: (1) the number of independent variables should be kept minimal and (2) a blinded validation set should be included [2]. Diagnostic accuracy establishes how accurately the test discriminates between those with and without the disease and is determined by calculating the test's sensitivity, specificity, likelihood ratio and receiver operating characteristic (ROC) curve [36].

One inherent problem of the high throughput technology mass spectrometry becomes evident upon consideration of statistical aspects [38]. A confidence level of 99.5% for the assignment of peptide sequences to fragmentation spectra suggests very high validity of data which is currently hardly realised. Modern equipment may allow the researcher to identify thousand different peptide sequences per hour. A confidence interval of 99% implies that five out of the thousand peptides are not correct. A typical experiment consists of around ten injections, summing up to 50 or more false peptide assignments. Comparative analysis of two groups of experiments summarizing five independent experiments would already sum up to 500 false peptide assignments. Complex analyses may require the consideration of hundreds of experiments. In such a case, a confidence rate of 99.5% per peptide identification may result in a chance to receive false results from a database query higher than 50%.

The only way out of this dilemma will be the consideration of expert knowledge in data analysis [27]. Currently, only quality features of individual spectra are considered for the assignment of amino acid sequences. Each decision is made independent of any other data. Actually, there are chances to make use of other data. We know that a given peptide has characteristic and reproducible chromatographic mobility as well as ionization and fragmentation characteristics. Therefore, the accessible knowledge of successfully identified peptides may facilitate the decision of peptide assignments in case of uncertainty. Furthermore, consideration of knowledge of the origin of the sample may greatly improve data consistency. To give an example: analysis of a mitochondrial fraction may allow some contaminating proteins derived from the endoplasmatic reticulum, but hardly from the cell nucleus. The analysis of a liver sample may include proteins from e.g. immune cells but hardly proteins specific for the heart. Although these implications seem trivial, they require complex expert system programming in order to be automatically implemented in the high throughput analysis of data. The systematic assessment of ontologies may, however, enable the implementation of such strategies.

The processing of data as realized in case of the CPL/MUW-database is outlined in the following. Actually, all protein identifications are based on peptide fragmentation spectra (mass spectrometry) (Fig. 21.2). Amino acid sequences are derived from the spectra and all related peptides identified during a LC-MS/MS run are sorted according to proteins they are derived from (SpectrumMill software). Actually, there are peptides which may be allocated to more than one protein, which need to be nominated in an easily accessible fashion (Fig. 21.2). In such a case, several considerations have to take place. The ambiguity may be solved by consideration of gene expression data and previously determined protein expression data. Consequently, established knowledge made available via the SwissProt-database needs to be accessed, while laboratory-owned data may as well aid the

decision process (Fig. 21.2). On the other hand, known potential contaminants such as keratins should be known to avoid misassignments. After the decision process resulting in protein lists comprising all relevant experimental and peptide identification data as realized via PRIDE XML-files, interpretation of data may be enabled by comparative analysis (Fig. 21.2). To provide an example: we have analysed secretomes of primary human endothelial cells at normal, angiogenic and inflammatory cell states. Accurate Venn diagrams displays the relation between these protein fractions (Fig. 21.2). Out of a total of 184 different proteins identified, 75 were found in all three kinds of cells. 114 proteins were secreted by untreated cells, 14 of which were not identified at the other two functional states. One twenty-nine proteins were identified in IL-1 β -treated cells, 33 of those were not identified at the other two functional states. Actually, some of them were found as well secreted by e.g. inflammatory activated macrophages, leaving 22 proteins apparently specific for inflammatory activated endothelial cells. This kind of comprehensive comparative analysis may strongly support the interpretation of complex data.

While data acquisition and protein identification may be considered as relatively simple tasks, there is still obvious demand for tools supporting data interpretation. These processes organize the data with respect to experiments and cell types, but not to functional aspects. Currently there is still obvious demand for further tools supporting data interpretation. The application of -omics techniques often leave the researcher with very long lists of identified genes and proteins which are impossible to comprehend. Current strategies try to relate expression data to signaling pathways in order to support biological interpretation [39–41]. There are still major limitations to these approaches. In many cases, the known involvement of a gene or a protein in a specific signaling or metabolic pathway would highlight the protein as such. Comparative analyses, however, record up- or down-regulation of proteins. Switching on a specific pathway does not necessarily mean that relative amounts of proteins involved in the pathway would be regulated. In many cases, however, the activation of a specific pathway would result in the up-regulation of proteins which are not at all involved in the exertion of the signaling or metabolic event. For the identification of the involvement of pathways, which is evidently desirable, databases would be required which exhibit consequences of pathway activation rather than involvement in pathways. There is still a demand for such databases.

Another shortcoming of current analysis strategies is the preferential assignment of tissue-specific expression patterns rather than cell type-specific expression patterns. Actually it is obvious that tissues are made of different kind of cell types. Some cell types such as immune cells occur in all tissue types, other cell types specifically occur in a single organ. It is the specific functional characteristics of hepatocytes which give raise to liver-specific specific proteins, liver cells other than hepatocytes do not express liver-specific proteins. Therefore, it would be more accurate to talk about hepatocyte-specific proteins rather than liver-specific proteins. There are databases listing organ-specific protein expression but no databases listing cell type-specific protein expression.

For this reason we established the following data analysis strategy. First of all the proteome profiles of isolated organelles which commonly occur in cells, such

as nuclei, mitochondria, ribosomes and proteasomes were determined. Such analyses obviously allow for the fact that cell type-specific proteins may as well occur in organelles such as nuclei but very much account for the fact that the basic protein composition of these organelles is highly similar. A proteome profile of a cell may thus already be structurally sorted according to the belonging to an organelle. As a consequence, a long protein list may already become much easier to be interpreted as related groups of proteins are identified.

The next step of systematic analyses focuses on cell types. We have already determined proteome profiles of lymphocytes, monocytes, dendritic cells, neutrophils, fibroblasts, endothelial cells, various epithelial cells and many others and classified both commonly expressed proteins as well as cell type-specific proteins. Some of these data have been made available to the public via the CPL/MUW database at www.meduniwien.ac.at/proteomics/database [27]. The expression specificity of several thousand proteins with respect to cell types can thus be immediately determined.

The SQL database (CPL/MUW – database of the Clinical Proteomics Laboratories at the Medical University of Vienna) facilitates (i) quality management of protein identification data, which are based on MS, (ii) the detection of cell type-specific proteins and (iii) of molecular signatures of specific functional cell states [27].

Proteome analyses of clinical materials constitute a big challenge for investigators due to its great complexity. Exact planning and documentation of each analysis step is crucial to enable meaningful data interpretation. This is why we strictly follow the established rules of the “minimum information about a proteomics experiment” (MIAPE) [35]. According to highest international standards, submit all relevant proteome analysis data to the international repository for proteome analysis data, the PRIDE database. We have already successfully implemented a program which automatically translates experimental data out of our database to a standardized PRIDE-XML format using international standardized ontology-terms to describe all experimental details (<http://www.ebi.ac.uk/ontology-lookup/>) [41]. Furthermore, we have programmed a proteome analysis database referring to the investigation of cross-cell type and cross-species comparisons of proteome analysis data derived from both, 2D-PAGE and shotgun analysis [27].

Proteins fulfil biological functions. If a cell enters a characteristic functional state it may need proteins not expressed under normal conditions. Such proteins may be specifically expressed only when the cells enter the functional state. As a consequence, the identification of such specifically expressed proteins may identify the corresponding cell state. Any disease-related symptom is a consequence of aberrant cell activities associated with the disease. Identification of aberrant cell activities may thus identify diseases. When investigating disease biomarkers we should consider the fact that proteins were designed by evolution to exert functions rather than to indicate diseases to medical doctors. Therefore, there are no protein biomarkers specific for a disease; there are only, actually plenty of, biomarkers specific for biological functions. If such an aberrant function is specifically associated with a certain disease the corresponding protein may be considered as a disease biomarker.

We have started to systematically assess protein expression profiles of cells at characteristic functional states. As expected, we were able to identify several specifically expressed proteins. These include proteins specifically related to functional states such as cell proliferation or inflammatory activation which may be entered by different kinds of cells. Actually, there are proteins which we found to be exclusively expressed by a single cell type at a specific cell state but not by any other cell. Therefore, these proteins are classified into organelle-derived, cell type-specific, cell state-related and cell type cell state-specific proteins. Comparisons of normal and diseased tissue proteome sample therefore result in the consideration of alterations in the abundance of organelles (indicative for, e.g. rate of mitochondrial respiration compared to glycolysis), the consideration of alterations of the occurrence of cell types (indicating e.g. invasion of immune cells or increase in the number of fibroblasts), the consideration of cell states (assessment of cell proliferation, cell stress, apoptosis, inflammatory activation of myofibroblast formation) and finally the occurrence of specific cell entities (e.g. type II macrophages). The knowledge of disease-associated aberrations in one or several of these aspects may thus allow us to design highly specific marker panels.

21.6 Identification of Biomarker Candidates by Secretome Analysis

Secretome analysis is an upcoming field of cancer research. This chapter gives a brief overview of the latest key secretome studies:

Recently, secretome analysis based on a LC-MS/MS label-free quantitative proteomics approach was used to compare the secretome of a primary cell line SW480 with its lymph node metastatic cell line SW620 from the same colorectal cancer patient [25]. They identified a total of 910 proteins from the conditioned media and 145 differential proteins between SW480 and SW620 (>1.5-fold change). Among them, trefoil factor 3 and growth/differentiation factor 15, two proteins upregulated in the metastatic cell line SW620, were analyzed in a large cohort of clinical tissue and serum samples and confirmed as biomarker candidates for the prediction of colorectal cancer metastasis [25]. Here secretome analysis allowed new insights into the pathophysiology of tumor progression.

An important study for a systematic identification of unique markers for colorectal cancer was performed by Wu et al. [42]. Secretomes of 21 cancer cell lines derived from 12 cancer types (colon cancer, leukemia, bladder cancer, lung cancer, NPC, hepatocellular carcinoma, cervical carcinoma, epidermoid carcinoma, ovary adenocarcinoma, uterus carcinoma, pancreatic carcinoma and breast cancer) were compared. Collapsin response mediator protein-2 (CRMP-2) was only secreted by the colorectal cell lines (Colo205 and SW480) but not any other cell lines tested and was therefore selected for further evaluation. Initially CRMP-2 was identified as a mediator required for semaphoring triggered growth cone collapse and was associated with carcinogenesis by p53 regulation. ELISA analyses of plasma

samples from colorectal patients and healthy controls were performed to examine the levels of CRMP-2 and CEA revealing that the sensitivities of plasma CRMP-2 and CEA were found to be 60.5% and 42.9%, respectively. This secretome analysis led to a novel marker, CRMP-2, which may be a colorectal marker superior to CEA. However, a large cohort study is required to validate the utility of plasma CRMP-2 levels for CRC screening and diagnosis.

In addition these authors analyzed proteins released by most cancer cell lines (pan-cancer marker candidates) and assigned these to specific secretion mechanisms. In the conditioned media of cancer cells proteins may be released via various cellular mechanisms, including classical secretion and nonclassical secretion pathways, as well as secretion via exosomes. The exocytosis of membranous vesicles called exosomes was initially described in antigen-presenting cells such as B-lymphocytes and dendritic cells, and was later found to also occur in tumor cell lines. The authors assigned some identified proteins to characteristic constituents of exosomes including ubiquitously expressed molecules such as intracellular metabolic enzymes (pyruvate kinase and alpha enolase), cytoskeletal proteins (actin, cofilin, tubulin, and moesin), and chaperones (HSP90 and HSP70). To determine whether some proteins may have been released into the medium by cell death, cell viability has to be measured.

To get panels of serum biomarkers for lung cancer, Xiao et al. [43] compared the secretome of primary cultures of lung cancer cells and the adjacent normal bronchial epithelial cells of six lung cancer patients using one-dimensional PAGE and nano-ESI MS/MS. They demonstrated that a panel of four proteins, CD98, fascin, polymeric immunoglobulin receptor/secretory component and 14-3-3 η had a higher sensitivity and specificity than any single marker.

To characterize extracellular events such as cell-to-cell interactions and cell-to-extracellular matrix interactions associated with breast cancer progression on the genomic level, gene profiles of secreted proteins were investigated in a cell line of human proliferative breast disease. Differentially expressed genes were searched for genes encoding secreted proteins in three public databases. The analysis displayed two clusters of secretome genes with expression changes correlating with proliferative potential [44].

Celis et al. [45] employed 2-DE and MALDI-TOF-MS to analyze the tumor interstitial fluid (TIF), which was collected of freshly dissected invasive breast carcinomas. From TIF, which perfuses the breast tumor microenvironment, they identified 267 primary translation products, involved in cell proliferation, invasion, angiogenesis, metastasis and inflammation.

A novel technology for investigating *in vivo* cancer secretome was recently developed by Huang and colleagues [46]. They collected the samples for further secretome analysis by implanting capillary ultrafiltration (CUF) probes into tumor masses of a live mouse at the progressive and regressive stages. Five of the detected proteins, including cyclophilin-A, S100A4, profilin-1, thymosin beta 4 and 10, which previously correlated to tumor progression, were identified at the progressive stage. They also identified specifically secreted proteins at the regressive stage called fetuin-A, alpha-1-antitrypsin 1–6, and contrapsin.

Very recently, a secretome analysis of 23 human cancer cell lines derived from 11 cancer types using one-dimensional SDS-PAGE and nano LC-MS/MS (GeLC-MS/MS) was performed on LTQ-Orbitrap MS to generate a comprehensive cancer cell secretome [37]. The identified proteins were selected as potential marker candidates according to three categories: (i) proteins apparently secreted by one cancer type but not by others (cancer-type-specific marker candidates), (ii) proteins released by most cancer cell lines (pan-cancer marker candidates), and (iii) proteins putatively linked to cancer-relevant pathways [37]. This analysis yielded 6–137 marker candidates selective for each tumor type and 94 potential pan-cancer markers. Among these, the monocyte differentiation antigen CD14 (for liver cancer), stromal cell-derived factor 1 (for lung cancer), cathepsin L1 and interferon-induced 17 kDa protein (for NPC) were selected for validation as potential serological cancer markers.

Immunohistochemistry revealed that bile salt sulfotransferase, ornithine carbamoyltransferase, monocyte differentiation antigen CD14, and isoform 1 of asialoglycoprotein receptor 2 were less immunoreactive in tissues of other cancer types, while multidrug resistance protein 1 and vitamin K-dependent protein C were over-expressed in hepatocellular carcinoma versus other cancers. Bladder cancer tissues reacted more strongly with proteins such as cadherin-6, squalene synthetase, ribophorin II, and 15-hydroxyprostaglandin dehydrogenase while the levels of neurogenic locus notch homolog protein 3 and trefoil factor 1 were higher in breast cancer tissues versus tissues of other cancers [37]. The stromal cell-derived factor 1 (CXCL12) reacted more strongly with lung cancer tissues. In addition, Wu et al. confirmed the significantly elevated plasma levels of two candidates (CD14 and SDF-1/CXCL12) in hepatocellular carcinoma and lung cancer patients [37].

In our recent study, we analyzed the secretomes of primary melanocytes, cultured melanoma cells and representatives of the most prominent stroma cells including fibroblasts, endothelial cells and dendritic cells by shotgun proteomics [9]. We consider the assessment of cell type-specific secretion characteristics as a prerequisite before potential relevant alterations of tumor-associated stroma cells can be recognized. In case a tumor-associated fibroblast secretes a protein not secreted by normal fibroblasts, but secreted e.g. by normal endothelial cells, such a protein would hardly be useful as biomarker. This is why we systematically analyzed the most important representatives of tumor-associated stroma cells. This strategy enables us to identify proteins which are aberrantly expressed by tumor-associated fibroblasts but not in any normal counterparts isolated from healthy background [9]. We performed secretome and proteome profiles generated from normal human skin fibroblasts in comparison to melanoma-associated fibroblasts isolated from mouse xenografts and fibroblasts from bone marrow of multiple myeloma patients. Further mutual comparisons were enabled including proteome profiles of melanocytes and M24met melanoma cells. All shotgun proteomics data have been made accessible via the PRIDE database. Amongst others, the candidate biomarkers GPX5, secreted by melanoma cells, in addition to periostin and stanniocalcin-1, which are expressed by melanoma-associated fibroblasts, were identified. Due to this data we started to investigate tumor associated fibroblasts of

primary melanoma and primary melanoma cells in a more systematic fashion by rtPCR, comparative genomic hybridization and cytoplasmic proteome and secretome analysis. This information will enable us to better understand cellular processes of the tumor and tumor associated cells in order to define new therapeutic agents and rational concepts for melanoma treatment and to detect biomarkers.

Secretome analysis is a novel research area offering new opportunities for biomarker discovery and drug development. However, despite promising results highlighted in this chapter, more systematic and hypothesis driven studies are needed. As primary cells are highly sensitive living units, any alteration in culture condition may result in aberrant protein secretion. Therefore, for clinical proteomics supporting biomarker discovery it is inevitable to refer to a SOP driven data resource of secretomes to enable an appropriate correlation of scientific with patient-derived information.

21.7 Conclusion

The identification of potential marker proteins is not trivial. Comparative analysis of serum samples and tissue specimen is hindered by the natural complexity of protein expression. Diseases like cancer mean a variety of de-regulated cell processes all of which eventually causing characteristic aberrant protein expression. Different kinds of patho-physiological processes may be associated with tumor development, such as involvement of the immune system, alterations of the microenvironment and characteristic processes in the cancer cells themselves. This complexity is further enhanced by the individual heterogeneity in disease in addition to heterogeneities introduced by the involved experimental procedures. Low abundant proteins may be hard to identify as long as they are present in a complex protein mixture together with other proteins, several at million fold higher concentrations. Dependent on the protein mixture, positive identification of actually present, but low abundant proteins may thus fail. Statistical evaluation of comparative proteome analysis data may thus not be able to identify the truly relevant proteins. One possible concept to overcome this inherent heterogeneity is based on the functional analysis of cell types in advance. It is predicated on the characterization of smallest independent units and tries to find a combination of independent units to match the molecular profile of an individual sample. This smallest unit capable of protein synthesis, the cell, decides whether or not to produce proteins with specific activity which may become related to a disease.

In mathematics the strategy to refer to independent functions is called Fourier transform which makes a complex function amenable for further analysis. The smallest independent and potentially predictable protein synthesis machinery unit is a cell. Since every functional cell aberration is associated with aberrations of protein expression when compared to normal, the cell is an optimal starting point for biomarker discovery. Like Fourier transform in physics, the establishment of profiles of the smallest autonomous protein production units in the body, i.e. cells,

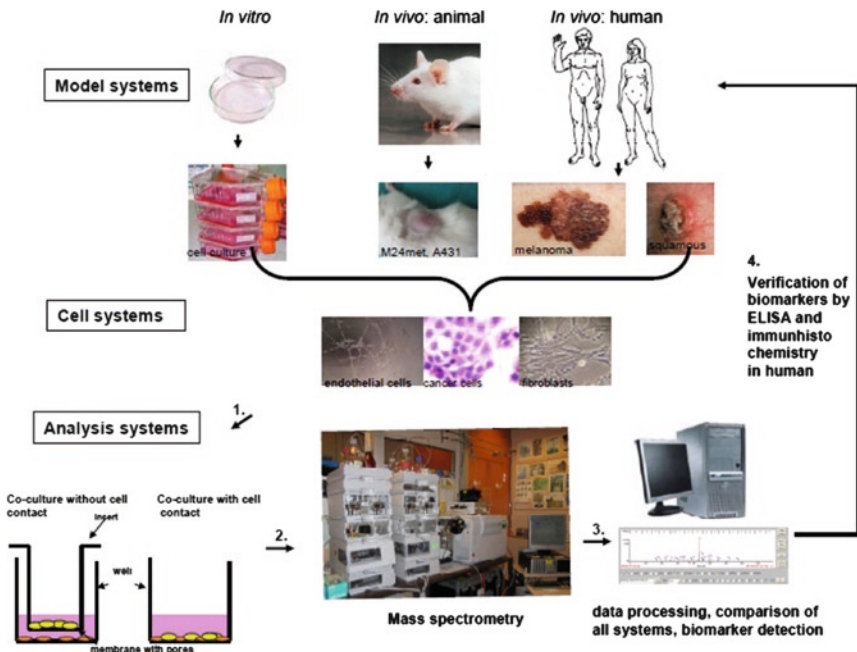


Fig. 21.3 The novel approach detecting biomarkers and defining potential therapeutic targets. The basic strategy for biomarker discovery is visualized. As model systems cultured cell lines, animal models for melanoma and squamous skin cancer and biopsy specimens of human skin cancer are presented. In all cases the secretome of the same isolated cell types (i.e. cancer cells, endothelial cells and fibroblasts) is analyzed. In further steps it is envisaged to analyze for specific cell-cell interactions mimicking characteristic tissue states for example by applying different co-cultures starting from in vitro to in vivo models. In a last step these results shall then be evaluated in the human background in the tissue and blood profile [9]

may greatly facilitate the interpretation of complex proteome profiles as derived from human serum or tissue samples (Figs. 21.2 and 21.3). All proteomes, i.e. protein mixtures, should it be from tissues, blood, plasma or other body fluids can be expressed as a function of cellular proteomes. The assignment to cellular proteome reference maps will lead to a massive reduction of apparent complexity (Fig. 21.2). Therefore possible candidates can be extracted by defining the involved cell systems such as cancer cells and distinguished cell of the environment including fibroblasts and endothelial cells in a first step. With the aid of specialized databases, for instance the CPL/MUW-database [27], specificities and commonalities of protein expression profiles of such different cells can be quickly assessed. Therefore, early teamwork between the clinical level, bioinformatics, medical informatics, and proteomic scientists is needed to overcome the current limitations.

One key question relates to our ability to draw appropriate conclusions for (short-, mid-, or long-term) therapeutic approaches and consequences from the highly dynamic proteome profiles. Specific cellular systems and subsystems and

functional components have to be defined prior to the analyses of a complex organism influenced by various states of disease. Integration of proteomics and cell-based technologies will allow the description of the molecular setup of normal and abnormal cell systems leading to the standardized discrimination of abnormal cell states in disease permitting for instance the design of individualized therapies, the prediction of further disease course in patients, the identification of new pharmaceutical targets, and establishment of a standardized framework of relevant molecular alterations in disease [2].

We make use of three different model systems (cell culture, tissue in vivo and human being), all have their strength and weakness starting from in vitro to human. The complexity but also relevance is increased from in vitro to human being. Therefore we combine all these systems (Fig. 21.3).

Our strategy is composed of seven independent steps (Fig. 21.3) [9]:

1. Establishment of relevant model systems mimicking various functional cell states including characteristic in vitro cell activation experiments and (non-) contact co-cultures
2. Standardization of protein isolation
3. Standardization of MS-procedures
4. Generation of proteome reference maps for human primary cells
5. Data organization via database
6. Interpretation of data from diseased tissues by the use of multiple reference maps
7. Verification of biomarkers or possible therapeutic targets by i.e. ELISA, immunohistochemistry, Western blot

In a last step these results shall then be evaluated in the human background in the tissue and blood profile (Fig. 21.3). ELISAs for instance the Luminex system [47] are to be established for the most promising candidates (including the specifically expressed proteins mentioned above). These assays will then be used to assess protein levels of candidate biomarkers in serum samples of patients. For validation we begin with assaying patients whose fibroblasts were found in vitro to secrete large amounts of candidate biomarker proteins. These data are then compared to serum samples derived from patients whose fibroblasts were found not to secrete these factors. This step of analysis will allow us to assess whether serum protein levels of these marker proteins are indeed related to the in vitro fibroblast expression levels as anticipated. The secretion specificity of the cancer associated fibroblasts has to be assessed by comparison to the secretomes of fibroblasts, endothelial cells, tumor cells and macrophages, which contribute to tissue remodeling and repair [9,26,48]. Here, we present a novel technical approach to better understand the mechanisms of tumor progression and metastasis by involving the microenvironment. The approach is of tremendous importance since it will allow us new insights in the pathophysiology of tumor progression, leading to the identification of novel biomarkers for early detection and prognosis and may lead to the identification of new therapeutic targets. The plethora of data will offer new opportunities to develop biomarker sets for ELISA analysis for the clinical routine [9]. The combination of a set of relevant

markers will yield an improvement of sensitivity and specificity of the screenings. By focusing on secreted proteins which are early shed by the microenvironment into the blood, specific information about the actual status of the patient and define a fingerprint of the tumor status in the patient can be gained. This strategy may enable early diagnosis of metastatic processes and offers an opportunity for a rational therapy selection. Candidate biomarkers shall be evaluated in clinical studies by correlation with the progression free and overall survival. This concept may be able to establish novel classifications, to define patient subgroups and to consequently allow us to enhance the often low overall response rates observed in clinical trials.

References

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL et al (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13:19–50.
2. Apweiler R, Aslanidis C, Deufel T et al (2009) Approaching clinical proteomics: current state and future fields of application in cellular proteomics. *Cytometry A* 75:816–832.
3. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol* 24:971–983.
4. Ahmed FE (2009) The role of capillary electrophoresis-mass spectrometry to proteome analysis and biomarker discovery. *J Chromatogr B Analyt Technol Biomed Life Sci* 877:1963–1981.
5. Jemal A, Siegel R, Ward E et al (2008) Cancer statistics, 2008. *CA Cancer J Clin* 58:71–96.
6. Schamberger CJ, Gerner C, Cerni C (2004) bFGF rescues 423-cells from serum starvation-induced apoptosis downstream of activated caspase-3. *FEBS Lett* 573:19–25.
7. Culp WD, Neal R, Massey R et al (2006) Proteomic analysis of tumor establishment and growth in the B16-F10 mouse melanoma model. *J Proteome Res* 5:1332–1343.
8. Liotta LA, Kohn EC (2001) The microenvironment of the tumour-host interface. *Nature* 411:375–379.
9. Paulitschke V, Kunstfeld R, Mohr T et al (2009) Entering a new era of rational biomarker discovery for early detection of melanoma metastases: secretome analysis of associated stroma cells. *J Proteome Res* 8:2501–2510.
10. Ohsie SJ, Sarantopoulos GP, Cochran AJ (2008) Immunohistochemical characteristics of melanoma. *J Cutan Pathol* 35:433–444.
11. Sekulic A, Haluska P, Jr., Miller AJ et al (2008) Malignant melanoma in the 21st century: the emerging molecular landscape. *Mayo Clin Proc* 83:825–846.
12. Omenn GS, States DJ, Adamski M et al (2005) Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* 5:3226–3245.
13. Prakash A, Piening B, Whiteaker J, Zhang H et al (2007) Assessing bias in experiment design for large scale mass spectrometry-based quantitative proteomics. *Mol Cell Proteomics* 6:1741–1748.
14. Diamandis EP (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. *Mol Cell Proteomics* 3:367–378.
15. Iau PT, Macmillan RD, Blamey RW (2001) Germ line mutations associated with breast cancer susceptibility. *Eur J Cancer* 37:300–321.
16. Welsh JB, Sapinoso LM, Kern SG et al (2003) Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum. *Proc Natl Acad Sci U S A* 100:3410–3415.

17. Tjalsma H, Bolhuis A, Jongbloed JD (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* 64:515–547.
18. Greenbaum D, Luscombe NM, Jansen R et al (2001) Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Res* 11:1463–1468.
19. Volmer MW, Stuhler K, Zapatka M (2005) Differential proteome analysis of conditioned media to detect Smad4 regulated secreted biomarkers in colon cancer. *Proteomics* 5:2587–2601.
20. Bundscherer A, Reichle A, Hafner C et al (2009) Targeting the tumor stroma with peroxisome proliferator activated receptor (PPAR) agonists. *Anticancer Agents Med Chem* 9:816–821.
21. Xue H, Lu B, Lai M (2008) The cancer secretome: a reservoir of biomarkers. *J Transl Med* 6:52.
22. Mbeunkui F, Fodstad O, Pannell LK (2006) Secretory protein enrichment and analysis: an optimized approach applied on cancer cell lines using 2D LC-MS/MS. *J Proteome Res* 5:899–906.
23. Zwickl H, Traxler E, Staettner S et al (2005) A novel technique to specifically analyze the secretome of cells and tissues. *Electrophoresis* 26:2779–2785.
24. Marouga R, David S, Hawkins E (2005) The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem* 382:669–678.
25. Xue H, Lu B, Zhang J, Wu M et al (2010) Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J Proteome Res* 9:545–555.
26. Gundacker NC, Haudek VJ, Wimmer H et al (2009) Cytoplasmic proteome and secretome profiles of differently stimulated human dendritic cells. *J Proteome Res* 2009.
27. Wimmer H, Gundacker NC, Griss J et al (2009) Introducing the CPL/MUW proteome database: interpretation of human liver and liver cancer proteome profiles by referring to isolated primary cells. *Electrophoresis* 30:2076–2089.
28. Thadikkaran L, Siegenthaler MA, Crettaz D et al (1995) Recent advances in blood-related proteomics. *Proteomics* 5:3019–3034.
29. Kislinger T, Emili A (2003) Going global: protein expression profiling using shotgun mass spectrometry. *Curr Opin Mol Ther* 5:285–293.
30. Perkins DN, Pappin DJ, Creasy DM, Cottrell et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20:3551–3567.
31. Yates JR, 3rd, Eng JK, McCormack AL et al (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67:1426–1436.
32. Kapp EA, Schutz F, Connolly LM et al (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* 5:3475–3490.
33. Jones P, Cote RG, Martens L (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res* 34:D659–663.
34. Craig R, Cortens JP, Beavis RC (2004) Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* 3:1234–1242.
35. Taylor CF, Paton NW, Lilley KS et al (2007) The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* 2007, 25:887–893.
36. Ahmed FE (2008) Application of MALDI/SELDI mass spectrometry to cancer biomarker discovery and validation. *Curr Proteom* 5:224–252.
37. Wu CC, Hsu CW, Chen CD et al (2010) Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol Cell Proteomics* 2010.
38. Nesvizhskii AI, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 4:787–797.
39. Malik R, Dulla K, Nigg EA, Korner R (2010) From proteome lists to biological impact- tools and strategies for the analysis of large MS data sets. *Proteomics* 2010.

40. Jimenez-Marin A, Collado-Romero M, Ramirez-Boo M et al (2009) Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proc* 2009, 3 Suppl 4:S6.
41. Griss J, Gerner C (2009) GPDE: a biological view on PRIDE. *J Proteom Bioinform* 2:167–174.
42. Wu CC, Chen HC, Chen SJ et al (2008) Identification of collapsin response mediator protein-2 as a potential marker of colorectal carcinoma by comparative analysis of cancer cell secretomes. *Proteomics* 8:316–332.
43. Xiao T, Ying W, Li L et al (2005) An approach to studying lung cancer-related proteins in human blood. *Mol Cell Proteomics* 4:1480–1486.
44. Dombkowski AA, Cukovic D, Novak RF (2006) Secretome analysis of microarray data reveals extracellular events associated with proliferative potential in a cell line model of breast disease. *Cancer Lett* 241:49–58.
45. Celis JE, Gromov P, Cabezon T et al (2004) Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol Cell Proteomics* 3:327–344.
46. Huang CM, Ananthaswamy HN, Barnes S et al (2006) Mass spectrometric proteomics profiles of in vivo tumor secretomes: capillary ultrafiltration sampling of regressive tumor masses. *Proteomics* 2006, 6:6107–6116.
47. Huber J, Kiefer FW, Zeyda M et al (2008) CC chemokine and CC chemokine receptor profiles in visceral and subcutaneous adipose tissue are altered in human obesity. *J Clin Endocrinol Metab* 93:3215–3221.
48. Mohr T, Haudek VJ, Gundacker CN et al (2009) Proteome and secretome profiles of angiogenic and inflammatory activated human endothelial cells: contribution to a systematic biomarker discovery approach. *Mol Cell Proteomics* 2009:submitted.