Quantitative body fluid proteomics in medicine – a focus on minimal invasiveness

Éva Csősz¹, Gergő Kalló¹, Bernadett Jakob¹, Eszter Deák¹ ², Adrienne Csutak² and József Tőzsér¹*¹

1) Proteomics Core Facility, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, Egyetem ter 1., 4032 Debrecen, Hungary,

2) Department of Ophthalmology, Faculty of Medicine, University of Debrecen, Egyetem ter 1., 4032 Debrecen, Hungary

*CORRESPONDING AUTHOR

József Tőzsér, Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Debrecen, 4010 Debrecen, Egyetem ter. 1. Tel. +36-52-416432, Fax. +36-52-314989, Email: tozser@med.unideb.hu

Keywords: Body fluid, quantitative proteomics, targeted proteomics, biomarker, AMP

Abstract

Identification of new biomarkers specific for various pathological conditions is an important field in medical sciences. Body fluids have emerging potential in biomarker studies especially those which are continuously available and can be collected by non-invasive means. Changes in the protein composition of body fluids such as tears, saliva, sweat, etc. may provide information on both local and systemic conditions of medical relevance. In this review, our aim is to discuss the quantitative proteomics techniques used in biomarker studies, and to present advances in quantitative body fluid proteomics of non-invasively collectable body fluids with relevance to biomarker identification. The advantages and limitations of the widely used quantitative
proteomics techniques are also presented. Based on the reviewed literature, we suggest an ideal pipeline for body fluid analyses aiming at biomarkers discoveries: starting from identification of biomarker candidates by shotgun quantitative proteomics or protein arrays, through verification of potential biomarkers by targeted mass spectrometry, to the antibody-based validation of biomarkers. The importance of body fluids as a rich source of biomarkers is discussed.
Introduction

With the advances of diagnostic techniques there is an increasing demand for non- or minimally invasive methods in medical diagnosis. The constantly produced and continuously available body fluids which can be collected by non-invasive means such as tears, saliva, sweat etc. can provide a feasible alternative to cerebrospinal fluid, amniotic fluid, synovial fluid, bronchoalveolar lavage fluid, serum etc. for diagnostic purposes. In some cases the analysis of body fluids collected by invasive means cannot be avoided but considering the need for well-trained medical workers, the possibility of infections and complications, and costs of medical interventions required for the collection of the body fluids increase the need for utilization of non-invasively collectable body fluids. In this review we will present the diagnostic utility of those body fluids which can be collected without medical intervention, such as tears, saliva, sweat, nasal secretion, cervicovaginal secretion and urine and we will discuss the biomarkers discovered with the administration of proteomics techniques.

Quantitative proteomics is a challenging part of proteomics, providing information not only on the presence or absence of proteins or protein isoforms, but also on their quantities. Typically, mass spectrometry, electrophoresis or immunological assays are utilized to give the relative or absolute quantities of proteins of interest. Absolute quantification determines the exact concentration of proteins, while in relative quantifications, only the relative change in protein quantity is compared in different conditions [1]. In most cases, relative quantification is sufficient to answer biological questions regarding changes in the amount of proteins following treatment, or when comparing two or more states. Relative quantification can give us an idea about protein-level changes in different states and conditions, and can show us differentially expressed proteins.
The semi-quantitative ELISA, one- or two-dimensional gel electrophoresis [2] and most of mass spectrometry-based techniques are suitable for relative quantification (Table 1).

For absolute quantification, a more elaborate experimental setup is needed. The simplest method for absolute quantification is the quantitative ELISA and the multiplex immunobead assay [3], however, these are rather expensive methods. Regarding mass spectrometry-based techniques, SRM-based targeted proteomics using various dilutions of purified, stable isotope labeled peptides is the method of choice [4]. Absolute quantification is required when the exact protein amount has biological significance, as in the case of stoichiometry analysis of protein complexes, biopharmaceutical applications for quality control or drug safety, inter- or intra-laboratory cross-validations, biomarker studies, or when the amount of various proteins within the same sample is required [1].

It is very hard to find the ideal quantitative proteomics technique applicable for the body fluid analysis; researchers willing to administrate such techniques should find the optimal solution matching their needs, considering the biological question, sample availability, costs, workload, and availability of mass spectrometers (Table 1). The presence of highly abundant proteins, such as α-amylase in saliva, dermcidin is sweat, lactotransferrin, lysozyme-C, etc. in tears and the low amount of available sample, especially when posttranslational modifications are to be studied, may require the administration of protein depletion and/or enrichment [5,6]. However the possible loss of quantitative information may hinder the wide application of these methods in the examination of non-invasively collectable body fluids. The administration of antibodies and the use of well-defined standards to monitor for example phosphoprotein enrichment [7] make possible the utilization of quantitative data. Another variation of protein enrichment is the utilization of antibodies against the digested peptides in the samples [8]. The Stable Isotope
Standards with Capture by Anti-Peptide Antibodies (SISCAPA) approach can be utilized to enrich specific peptides from different samples by using anti-peptide antibodies to capture endogenous peptides and spiked stable isotope-labeled internal standard peptides [9].

**Biomarkers**

The two major fields utilizing quantitative proteomics in medicine are biomarker studies and identification of therapeutic targets; however, these two fields sometimes overlap. In some conditions, proteins are considered biomarkers when identified as having a central role in the normal or pathological function of the cells/tissues, and their presence or absence causes a malfunction leading to disease [10]. Consequently, biomarkers are at the same time targets for drug design and therapy [11]. In other cases, although the biomarkers can give information regarding alterations related to disease, they are not suitable targets for therapeutic intervention; their presence, absence or differential expression may be the consequence and not the cause of the pathological condition, hence, they cannot be used as target for therapies, nor can they explain the pathophysiological phenomena [12]. In this review, our aim is to show advances in the biomarker field aided by quantitative proteomics. According to the National Institutes of Health Biomarkers Definitions Working Group, a biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [13]. In recent years, studies aiming to identify biomarkers specific for different pathological conditions emerged, and hundreds of proteins were shown to be potential biomarkers specific for various diseases [14].
In this review the potential biomarker term will be used for proteins whose level is significantly different between the disease and control groups, while the validated biomarker will refer to those proteins which were validated using an independent antibody-based technique.

Despite the fact that advances in research techniques and intensive efforts in biomarker studies had led to identification of high numbers of potential biomarkers in the discovery phase, the number of FDA approved biomarkers is much less [15–21]. One reason for drastically decreasing the number of validated biomarkers as they go through the verification and validation processes can be the lack of standardized workflows [14]. Comparative studies examining protein profiles usually include less than ten samples per group, with different analytical methods detecting many differentially expressed proteins between the diseased and control groups. The number and identity of the differentially expressed proteins can vary from study to study, depending on the sensitivity of the applied analytical method, the size of the groups, and the patients enrolled in the study. Ideally, the only difference between the diseased and control groups should be the presence or absence of the disease [22]. In small groups, this is possible, however, when larger patient cohorts are enrolled, the presence of secondary disorders serves as hindrance to the analysis.

The Clinical Proteomic Technologies for Cancer (CPTC) initiative established by the National Cancer Institute recommends a workflow for cancer biomarker development [14]. They suggest the use of SRM-based targeted proteomics method for verification of hundreds of potential biomarkers identified in the discovery phase. For verification, a large sample number is required. Ideally, hundreds of samples are analyzed, and after verification, the biomarkers can be subjected to the next step, the validation step. For validation, only few proteins should be chosen and tested on thousands of samples. Statistical analysis, including ROC curves [23] for biomarkers proved
to be convenient in the verification process can help identify the proteins with the highest specificity and sensitivity in discriminating between the control and diseased groups.

Immunoassays are also very suitable for biomarker validation, allowing for analysis of large sample numbers in a short period of time. When multiple proteins have to be validated, the utilization of multiplex immunobead assays are more cost effective [3].

One of the bottlenecks of biomarker studies is sample availability. From pathophysiological point of view, it is of critical importance to analyze the tissue or the biological material in which the alteration happens, as well as the proximity of the material to the site of pathological alteration. In many cases however, these types of samples are not available for biomarker studies. For example, in case of pathological conditions affecting the retina; such as diabetic retinopathy, proliferative vitreoretinopathy, etc., the study of proteome changes of the retina would be important to understand the pathological mechanisms underlying the disease. On the other hand, for diagnostic purposes, the retina; which can be obtained only with highly invasive surgical intervention, is not an acceptable source for minimally- or non-invasive biomarker studies.

In spite of the availability of advanced imaging methods, there is a high demand for laboratory diagnostic assays using biological samples or body fluids acquired by non-invasive or minimally invasive methods. Advances in proteomics and metabolomics techniques had led to improved sensitivity, and provided the possibility to detect protein or metabolite changes in body fluids which are not necessarily in proximity to the diseased area. For example, studies of tear or serum composition could identify potential biomarkers specific for retinal diseases [24,25].

The emerging “omics” technologies provide new possibilities for identification of biomarkers form the continuously available body fluids that can be collected by non-invasive means, shifting
the focus from the study of tissue material obtained by autopsy and biopsy to the study of easily collectable and continuously available body fluids such as tears, saliva, sweat, etc.

**Tear fluid as a source of biomarkers for ocular and systemic diseases**

Tear fluid is a complex mixture of proteins, lipids, salts and other organic molecules produced by the lacrimal glands. Normally, the tear production rate is approximately 2 µl/min [26] and its typical protein concentration is 5-7 µg/µl [27]. Functions of the tear film are lubrication of the eye, delivery of nutrients and maintenance of the refractivity of the cornea [28]. Beside these roles, tear provides an effective chemical barrier on the surface of the eye via the secreted antimicrobial and immunomodulatory proteins (AMP), which provide protection against infections [29]. Currently, more than 1500 tear proteins have been identified by proteomics techniques [30–32]. Major tear proteins; such as lactotransferrin, lysozyme-C, prolactin-inducible protein, lacritin etc., are involved in the defense against pathogens [31], and their relatively high abundance makes these proteins the major antimicrobial proteins of tear. While many of the tear proteins are produced by the lacrimal glands, some of them originate from epithelial cells; such as dermcidin, defensins, etc., and there are also proteins filtered from the blood such as albumin [33,34].

The human tear has become one of the most investigated body fluids, being a possible source of biomarkers [30,31]. Tear fluid is one of the non-invasively obtainable body fluids that is relatively easy to collect, and its study may help understand the pathogenesis of ocular and some systemic diseases, in addition to aiding in the prediction of the outcome and scheduling follow-ups for therapeutic sessions. Based on localization, tear can primarily reflect pathological conditions related to the anterior segment of the eye, but it can also provide information on the
retinal or vitreal status, and more broadly, information related to systemic changes can also be acquired by analysis of tear proteins (reviewed by [35,36]). The protein composition of tear was examined in various pathological conditions, and proteome-level changes were observed in case of contact lens wearers as well [37]. Analysis of tear protein profile can provide useful biological information on understanding the molecular mechanisms of ocular diseases; such as dry eye syndrome [38–41], blepharitis [42], climatic droplet keratopathy [43], diabetic retinopathy [25], keratoconus [44,45], mycotic keratitis [46] and vernal keratoconjunctivitis [47]. Changes in the tear proteome can also reflect systemic diseases such as multiple sclerosis [48] and Alzheimer’s disease [49]. SELDI-TOF data are available on tears of patients with Sjögren’s syndrome [50] and breast cancer [51,52], but the exact protein composition of the peaks characteristic for the pathological conditions is missing. Biomarkers for these pathological conditions identified by quantitative proteomics are listed in Table 2.

Various immunological tests exist to study the concentration of pro- or anti-inflammatory markers in tears, in order to assess the inflammatory status in ocular or systemic diseases [53–55]. Inflammatory changes were shown to be associated with glaucoma and neurodegenerative disorders such as Parkinson’s disease and multiple sclerosis [48,56–59]. The 47 potential and the 10 validated biomarkers for various ocular or systematic diseases identified by quantitative proteomics experiments predict the clinical application of some of the identified biomarkers in the near future.

Saliva – the easily accessible, continuously available source for biomarkers

Saliva is a complex mixture secreted from major and minor salivary glands and from the gingival crevice [60]. It is composed of more than 99% water, making it a very dilute body fluid. The
remaining part of saliva is composed of various electrolytes, proteins, mucins and nitrogenous products such as urea [61]. Saliva contains more than 2000 proteins [62], the most abundant of which are α-amylase [63], mucins [64], proline rich peptides [65], cystatins [64] and serum albumin [66]. Although high variability in protein content was observed depending on collection time, sex, age, pathological conditions, the typical protein concentration of saliva is 0.7-2.4 mg/ml [67-69]. Similarly to tear fluid, the abundant salivary proteins are part of the innate immune system, due to their antimicrobial activity, and their protective role from microbial proteases. The non-invasive collection and the continuous availability of the saliva make it an excellent source of biomarker studies; unsurprisingly, the protein composition of saliva has been analyzed by several workgroups, indicating its relevance to medical applications. A classical 2D gel electrophoresis-based study of saliva obtained from children with autism spectrum disorder revealed 8 down-regulated and 8 up-regulated salivary proteins as potential biomarkers for the disorder [70]. The advantage of multiplex measurements by chemical labeling with iTRAQ and label free quantification mass spectrometry of salivary proteins have been used for analysis of patients with breast cancer [71,72], malignant oral lesions [73], oral squamous cell carcinoma [74,75], chronic graft-versus-host disease [76,77], Sjögren's syndrome [78] and bisphosphonate-related osteonecrosis of the jaw [79]. Differentially expressed proteins have been identified in these studies (Table 3), providing potential biomarkers to aid diagnosis and possibly enable for a deeper understanding of the molecular mechanisms involved in the pathogenesis of the studied diseases. SRM-based quantitative proteomics method for the analysis of O-glycosylated forms of salivary protein MUC7 in saliva of patients with rheumatoid arthritis has been developed and used to demonstrate the altered O-glycosylation pattern in rheumatoid arthritis, with a proposed role in diagnosis and treatment follow-up [80]. SELDI-TOF mass spectrometry method was used to identify biomarkers for fibromyalgia [81], oral squamous cell carcinoma [82,83] and Sjögren's
syndrome [84] (Table 3), and peaks showing significant differences are available for breast cancer [85] and oral mucositis [86]. From the so far identified 347 potential biomarkers for various diseases only 10 were validated by antibody-based techniques and are indeed suitable for population screening. This is highly relevant, especially in case of oral cancers, given the high prevalence of oral squamous cell carcinoma and its increasing tendency in the younger population [87,88].

**Sweat biomarkers for local and systemic conditions**

The skin acts as an effective barrier against pathogens in the first line of host defense. Besides providing a physical barrier, the skin also creates a chemical barrier via antimicrobial and immunomodulatory proteins secreted by keratinocytes, sebocytes and epithelial cells [89]. Like saliva, sweat is a continuously secreted and highly diluted body fluid, its protein content provides an effective defense against pathogens, and is involved in tissue regeneration after injury [90]. The most abundant human sweat protein is dermcidin, though clusterin, prolactin-inducible protein, apolipoprotein D and serum albumin are highly expressed in sweat as well [91]. These proteins are essential to the formation of the chemical barrier of the skin, due to their antimicrobial activity [90,92], chaperone function [93] and antioxidant effect [94,95]. Sweat samples from patients with skin or systemic diseases such as ectodermal dysplasia, cystic fibrosis, atopic dermatitis and schizophrenia were analyzed using quantitative proteomics [96–99]. It was demonstrated that in case of skin disorders, the level of proteins involved in the host defense and tissue regeneration were reduced in the samples of patients with ectodermal dysplasia and atopic dermatitis [96,99]. The typical concentration of sweat proteins is 0.1-0.4 mg/ml [97,100] but it can vary depending on the exterior and interior conditions such as temperature, stress etc. [101,102]. It was proposed that reduction in the amount of proteins
involved in the immune homeostasis may contribute to the development of ectodermal dysplasia; therefore, they are classified as potential biomarkers and possible therapeutic targets in this disease. In case of systemic conditions, analysis of the sweat proteome has revealed proteins which were differentially expressed in schizophrenia (Table 4) [97]. In spite of the 180 potential and two validated (dermcidin and RNase7) biomarkers identified, many open questions remain in connection to the properties of sweat. Only limited information is available regarding (1) the normal rate of the sweat production, (2) the basal protein composition of the sweat collected at various body surfaces, (3) possible diurnal variations of the protein content, (4) the protein composition depending on the collection method etc., highlighting the role of proteomics in utilizing this continuously available, but very dilute and sometimes not so easy-to-collect body fluid for diagnostic purposes.

**Nasal secretion as a valuable source of potential biomarkers**

The nasal secretion has a protective role in the airways; its production rate and the protein content can vary substantially. Protein amounts produced between 0.8 and 32.7 mg/ml were observed depending on the collection methods [103,104]. Nasal discharge contains a lot of antimicrobial proteins; such as lysozyme-C, lactotransferrin, several types of defensins [105], also in addition to components of the adaptive immune system, such as immunoglobulins [106]. Compared to the other body fluids, a few number of publications report nasal mucus as a source of biomarkers; by quantitative proteomics only six potential biomarkers have been identified and only one of them has been validated so far (Table 5) [107]. As a result of various studies, more than 451 proteins were identified in the nasal mucus, and many of them are related to the immune system [107–109]. Because of the anatomical localization, nasal mucus is the easily collectable body fluid with the closest proximity to the nervous system. This property is exemplified by the nasal
administration of drugs, enabling their direct delivery through the olfactory neurons or trigeminal nerve endings to the central nervous system [110,111]. Considering the fact that parts of the olfactory neurons are embedded into the nasal mucosa, their secretions should be detectable in the mucus. Based on the high sensitivity and versatility provided by the state-of-the-art quantitative proteomics methods, the easy-to-collect nasal discharge is predicted to be a very valuable source for further biomarker studies.

**Cervicovaginal fluid as source for biomarkers for obstetrics and gynecology**

Vaginal fluid is a non-invasively collectable body fluid which may secrete from vagina, cervix or upper genital tract [112]. Vaginal discharge has an important function in the innate immunity and homeostasis and pathological conditions can significantly influence the balance in normal vaginal milieu leading to qualitative and quantitative protein alterations in vaginal secretion [113–115]. Many factors can influence the protein composition of the vaginal fluid, for example the varying levels and ratios of estrogens and progesterone resulting in variable amounts of protein yield, therefore it is practically impossible to quantify the normal amount of total protein compound in healthy cervicovaginal fluid [116]. Currently, more than 680 proteins have been identified in vaginal fluid and the majority of the identified proteins have a role in host defense [117]. The proteomic analysis of vaginal fluid can provide more information about the pathophysiological conditions affecting the female genital tract [118,119]. 63 potential biomarkers for pregnancy-related problems such as intra-amniotic inflammation [112,120] and preterm labour/preeclampsia [121–125], cervical cancer [126] and HIV infection [127] were identified (Table 6) using quantitative proteomics techniques and six biomarkers were validated.

**Urine biomarkers for systemic and urogenital diseases**
Urine is formed in the kidneys as a result of ultrafiltration of the plasma to eliminate waste products; such as urea and metabolites. Serum proteins are filtered based on their size and charge at the glomeruli, and are thereafter reabsorbed in the proximal renal tubules [128,129], thus, protein concentration in urine under physiological conditions is very low, usually between 0 and 0.2 mg/ml [130] and the normal protein excretion is less than 150 mg/day [131]. This is about 1000 times less compared to other body fluids such as plasma [132]. Excretion of more than 150 mg/day protein is defined as proteinuria, and is indicative of glomerular or metabolic dysfunction [129].

Urine can be collected in large quantity and non-invasively. Therefore, despite the low protein concentration, adequate amounts of proteins can be collected from a single sample [131,133]. Another advantage of using urine as a body fluid for diagnosis is the possibility of collection of samples repeatedly, even over long periods of time. Under normal conditions, urine contains proteins originating from the blood and kidneys [134–139], making urine a good source for analysis of diseases affecting the kidney or the urogenital tract; such as kidney failure resulting from high blood pressure and diabetic nephropathy [140,141], prostate cancer [142,143], polycystic kidney disease [144], kidney chronic allograft dysfunction [145], chronic allograft nephropathy [146], congenital obstructive nephropathy [147], lupus nephritis [148], urolithiasis [149], in addition to urinary, renal and bladder cancer [150–161]. Besides urogenital and kidney dysfunctions, urinary proteomics has a great potential in biomarker studies of coronary artery atherosclerosis [162,163], obstructive sleep apnea [164], ovarian cancer [165], breast cancer [166] and sepsis [167,168].

Urine has become one of the most attractive body fluids in clinical proteomics [155], nevertheless, variability in protein concentration poses a hindrance to the analysis of samples.
This can be compensated for by standardization based on urinary creatinine [169] or urinary housekeeping peptides, which are present almost ubiquitously in human urine independently of age, sex, health, or drug administration [170]. Details on urine sample preparation have already been discussed in some recently published reviews [171–175]. However, urine samples require handling through a set of methods; such as centrifugal filtration, ultrafiltration, lyophilization or precipitation. Removal of albumin is sometimes required to improve identification of low-abundance proteins [176].

Urine proteome analysis may potentially unravel markers for cancers of urogenital or systemic origin including bladder [150–154,156,157,160,161], prostate [170], renal [158,159], breast [166] and ovarian cancers [165] (Table 7). There has been an increasing interest in developing urine biomarkers for the detection of renal allograft rejection as an alternative to percutaneous needle biopsy, which is costly and associated with significant patient morbidity and mortality [79]. The 245 identified potential and 10 validated biomarkers for kidney-related and systemic diseases make urine a rich source of easy-to-collect, and continuously available source of biomarkers.

**Body fluids as part of the innate immune system**

The localization of the above mentioned body fluids correlates with the possible entry sites for microorganisms, implying the presence of a well-defined defense system. Besides mediating their physiological functions; body fluids also partake in the protection of the organism by providing a chemical barrier [29]. This chemical barrier is made up of the secretions of various glands and epithelial cells, and the characteristic composition of antimicrobial and immunomodulatory peptides (AMP) makes the AMP cocktail specific for different body fluids [177]. The composition of the AMP cocktail may determine which microbes can colonize our body,
providing unfavorable conditions for pathogen growth under physiological conditions. It was demonstrated that skin microbiome changes in response to various stimuli such as food, drugs, age, gender, physiological status etc., implying changes of the AMP cocktail [178]. These changes can be examined by the sensitive quantitative proteomics techniques, which are well suited to monitor AMP levels at various sites of the body in different physiological or pathological conditions [179].

Proteomic analyses have revealed that highly abundant proteins observed in some body fluids; such as tears and sweat, are part of the first line defense. Biomarkers identified from body fluids by different research groups have enriched the library of proteins possessing antimicrobial or immunomodulatory properties (proteins with italics in the Table 2-7). More than half of the identified tear biomarkers for eight pathological conditions have antimicrobial or immunomodulatory functions. A similar phenomenon can be observed in case of vaginal fluid and nasal mucus as well, though only one proteomic study is available for the nasal mucus. In case of saliva and sweat, the number of the proteins having a role in host defense is high, making up in; most of the cases, more than half of the identified biomarkers, in some instances, where more than 130 proteins were identified approx. only 30% of the proteins belonged to the AMP family (Table 8). From an analytical point of view, this can also mean that utilizing mass spectrometers with high resolution mass analyzers or extensive sample fractionation, enrichment or depletion and identifying more than 100 proteins can reveal the proteome masked by the highly abundant proteins. The urine is unique in a sense that its distinct anatomical and physiological properties imply the presence of a different chemical barrier as compared to that of other body fluids [180]. The number of AMPs among the identified urine biomarkers is high but fall far less than those of other body fluids (Table 8).
Analysis of the chemical barrier components can provide valuable information; our group has identified lipocalin 1, lactotransferrin, lacritin, lysozyme-C, lipophilin A and immunoglobulin lambda chain as potential biomarkers for the proliferative stage of diabetic retinopathy [25]. Lipocalin 1, lacritin, lysozyme-C and dermcidin for Alzheimer’s disease [49]. Lactotransferrin, lipocalins, lysozyme-C, lacritin and dermcidin are involved in the immunologic and inflammatory processes and defense against pathogens [30,31,35,36]. It has been shown that lactotransferrin found in all body fluids is an active agent against microbes and parasites, and has been implicated in protection against cancer [90]. Lysozyme-C is a hydrolytic enzyme with muraminidase activity required for degradation of the bacterial cell wall [90], has antifungal activity [181] and protects against HIV infection [182]. Lipocalins are a family of lipid binding proteins, they play a role in host defense because of their protease inhibitor activity [183] and by sequestrating iron, they limit bacterial growth [184]. Lacritin is a secreted protein found in tears and saliva, it promotes secretion [185], epithelial cell proliferation [186], corneal wound healing [187], additionally, its C-terminal fragment has bactericidal activity [188]. Changes of tear AMP levels are likely not the cause but rather the consequence of diabetic retinopathy progression, hence, it can be used to develop bedside test for screening purposes. Currently, photographic methods and retinal image analysis to track neovascularization and increased vascular permeability in the retina can be used for screening [189–191], this, however, requires proper instruments, software, and well-trained human graders, making diagnosis available only in big screening centers. In order to make the screening widely accessible, easy-to-implement tests; which can be used by general practitioners, are needed. Potential biomarkers for different stages of diabetic retinopathy were identified, and their validation is in progress, making tear biomarkers novel tools for population screening, based on the pre-screening of a large number of patients and involving human graders only in positive or ambiguous cases [192].
The composition of the AMP cocktail changes constantly, implying both qualitative and quantitative changes of the cocktail components, in order to adapt to various conditions. Most probably, composition of the chemical barrier is characteristic to the stimulus to which the organism has to adapt. Thus, changes of the composition of the host defense proteins in body fluids as a response to pathological conditions provide a feasible source for biomarker studies, which can be implemented using cutting edge quantitative proteomics analyses.

Extracellular vesicles – future perspectives

Different types of membrane vesicles of endosomal and plasma membrane origin are released in an evolutionally conserved manner into the extracellular environment by cells. These circulating microvesicles found in many types of body fluid and in extracellular space play an important role in cell signaling and intercellular communication [193–195]. According to their biogenesis, the extracellular vesicles can be grouped as apoptotic blebs, exosomes and microparticles, but the nomenclature is not clear for the latter two [196].

Exosomes, the membrane vesicles of endocytic origin are fragments from 40 to 100 nm secreted by most cell types. There are studies regarding exosome analysis of plasma [197], urine [198], amniotic fluid [199], saliva [200], bronchoalveolar lavage fluid [201], breast milk [202] etc.

Microparticles are of cellular membrane origin shed from stressed or damaged cells. Their diameters ranges from 0.1 to 1.0 μm [203]. Almost 3700 proteins were identified in total plasma microparticles using proteomic methods [204,205].

Proteomics has traditionally focused on the study of extracellular vesicle (EVs) proteins found in body fluids, but the proteomic profiles are highly dependent on how EVs were isolated [206]. Alterations in EV protein levels observed in pathological conditions make exosomes a good
source for biomarker investigations [207]. More than 35000 proteins were found in the extracellular vesicles (http://www.microvesicles.org/) and using quantitative proteomics techniques 32 potential biomarkers for bladder and prostate cancer, kidney diseases, deep vein thrombosis and brain tumors were identified in body fluids. 10 of the potential biomarkers were validated using antibody-based methods (Table 9) [208–215].

Conclusion

Based on the reviewed literature, hundreds of biomarker candidates have been identified, mainly from shotgun quantitative proteomic studies. Experiments using 2D electrophoresis, SELDI-TOF, label-free quantification, iTRAQ and TMT labeling on various body fluids have already provided more than 800 soluble potential biomarkers specific for different diseases. The number of verified potential biomarkers using targeted; preferentially SRM-based, approach in examined body fluids is 40, and the number of biomarkers validated by antibody-based methods is 34, highlighting the need for verification and validation studies. In our opinion, the NCI-CPTC Biomarker Development Pipeline suggested for cancer biomarker identification [14] should be used in all biomarker studies involving body fluids, regardless of the nature of the examined pathological condition. In some diseases, there are hundreds of potential biomarkers identified; in such cases the verification step using targeted proteomics followed by validation step using antibodies should be carried out (Figure 1). In those diseases where there are no data available, shotgun proteomics studies followed by targeted methods have to be applied in order to identify new biomarkers with high diagnostic value. In the field of body fluid proteomics, the need for non-invasively obtained, readily available biomarkers with diagnostic significance imposes
pressure on state-of-the-art quantitative proteomics to utilize more body fluids, a rich source of yet undiscovered biomarkers.

Acknowledgement

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences for Éva Csősz and Adrienne Csutak, TÁMOP-4.2.2.D-15/1/KONV-2015-0016, TÁMOP 4.2.4.A/2-11-1-2012-0001, TÁMOP 4.2.2.A-11/1/KONV-2012-0045 and OTKA PD116817. Gergő Kalló is receiver of Astellas Pharma Ltd. Fellowship and Adrienne Csutak is receiver of Lajos Szodoray Postdoctoral Fellowships. The work of Dr. Mohamed Mahdi for English proofreading of the manuscript is greatly acknowledged.
References


[90] J. Wiesner, A. Vilkenskas, Antimicrobial peptides: the ancient arm of the human immune


[127] G. Van Raemdonck, G. Zegels, E. Coen, B. Vuylsteke, W. Jennes, X. Van Ostade, Increased Serpin A5 levels in the cervicovaginal fluid of HIV-1 exposed seronegatives suggest that a subtle balance between serine proteases and their inhibitors may determine


1267038.


Figure legends

Figure 1. Body fluid analysis workflow for biomarker discoveries
<table>
<thead>
<tr>
<th>Major consideration on method selection</th>
<th>2DE</th>
<th>Metabolic labeling</th>
<th>Chemical labeling</th>
<th>Label-free quantification</th>
<th>SRM</th>
<th>HR-SRM</th>
<th>PRM</th>
<th>SWAT</th>
<th>Multi-ex immunobead assay</th>
<th>SELDI-TOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative quantification</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>not suitable</td>
<td>good</td>
<td>very good</td>
<td>good</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>Absolute quantification</td>
<td>not suitable</td>
<td>not suitable</td>
<td>possible</td>
<td>not suitable</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not suitable</td>
<td>not suitable</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>fair</td>
<td>very good</td>
<td>good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>good</td>
<td>very good</td>
<td>fair</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>good, depends on the type of dye applied</td>
<td>good, depends on the type of instrument applied</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>very good</td>
<td>good</td>
</tr>
<tr>
<td>Possibility to analyze multiple samples in one run</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
<td>not possible</td>
</tr>
<tr>
<td>Possibility to analyze multiple analytes in one run</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
<td>possible</td>
</tr>
<tr>
<td>Type of mass spectrometer required</td>
<td>no specific requirement</td>
<td>no specific requirement</td>
<td>no specific requirement</td>
<td>triple quadrupole, QTRA P</td>
<td>Q-TOF</td>
<td>quadrupole-Orbitrap</td>
<td>preferably Q-TOF</td>
<td>not applicable</td>
<td>SELDI-TOF</td>
<td></td>
</tr>
<tr>
<td>Manual work requirement</td>
<td>very high</td>
<td>high</td>
<td>high</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>medium</td>
<td>low</td>
<td></td>
</tr>
</tbody>
</table>

Table 1
| **Special advantages** | • Global changes of protein profiles can be examined  
• Post-translational modifications can be detected  
• Any kind of sample can be analyzed  
• 100% label incorporation efficiency  
• Any kind of sample can be analyzed  
• Wide dynamic range  
• Availability of databases containing SRM transitions  
• High resolution  
• All transitions can be monitored  
• High resolution  
• High mass accuracy  
• High throughput  
• Allows the recording of the complete fragment ion spectrum  
• Allows retrospective protein analysis  
• High specificity provided by the administration of antibodies  
• Separation on the surface of the protein chip |
| **Limitations** | • In case of body fluids with low sample volume sample pooling may be needed  
• Highly abundant proteins can overload the gel  
• Proteins with low abundance may be undetected  
• Typical labeling efficiency is conditional and requires optimization  
• The presence of byproducts may cause problems during data evaluation  
• Many technical replicates are needed which increases instrument time and costs  
• Availability of peptide MS/MS spectra is essential to develop SRM methods  
• SIL peptides may be required  
• Availability of m/z of peptide precursors and their retention times are essential to develop HR-SRM methods  
• Availability of m/z of peptide precursors and their retention times are essential to develop PRM methods  
• Low level of automation  
• Data analysis  
• Although there is possibility to construct custom kits, the circumstances such as cost, time, optimization difficulty due to the highly different concentration of the individual analytes in the samples,  
• Lack of information on the identified proteins  
• Has not been supported by the manufacturer for many years |
<table>
<thead>
<tr>
<th>Utility in body fluid analysis</th>
<th>suitable</th>
<th>not suitable</th>
<th>suitable, ideal for screens</th>
<th>suitably, ideal for validations</th>
<th>suitable, ideal for screens</th>
<th>suitable, ideal for validations</th>
<th>suitable with serious limitations</th>
</tr>
</thead>
</table>

favor the utilization of already available kits, in this way analysis may be limited to what is provided by the vendors.
Table 2: Tear protein biomarkers identified by quantitative proteomics methods.

The proteins in italic have antimicrobial and immunmodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline, asterix) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s disease</td>
<td>SRM</td>
<td><em>LACRT, LCN1, LYZ, DCD</em></td>
<td>49</td>
</tr>
<tr>
<td>Blepharitis</td>
<td>2D SDS-PAGE</td>
<td>ALB, *CST2, IGKCVIII, LACRT, LYZ, PIP, PKLR, SERPINA1</td>
<td>42</td>
</tr>
<tr>
<td>Climatic droplet keratopathy</td>
<td>iTRAQ</td>
<td>DKFZp686M08189, HP, JCHAIN, LACRT, PIGR</td>
<td>43</td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>iTRAQ</td>
<td>*IGLC1, LACRT, LCN1, LTF, LYZ, SCGB1D1</td>
<td>25</td>
</tr>
<tr>
<td>Dry eye syndrome</td>
<td>iTRAQ</td>
<td>ALDH1A1, APOA1, <em>AZGP1,C3, CST4, ENO1, EZR, GC, HP, IGHA1, IGHA2, IGHG3, IGLC1, JCHAIN, LACRT, LCN1</em>, LTF, LYZ*, ORM1, PIGR, PIP, PRDX1, PRP1, <em>PRR4</em>, S100A11, S100A4 , S100A8*, S100A9, SCGB2A1, SERPINA1*, SMR3B*, TCN1, ZG16</td>
<td>38,39,40,41</td>
</tr>
<tr>
<td>Keratoconus</td>
<td>Label-free quantification</td>
<td>AZGP1, B2M, CLU, CTSB, CST1, CST4, FGA, IGHA1, IGHA2, IGKC, IGLC1, JCHAIN, KRT14, KRT5, LACRT, LCN1, LCN1P1, LGALS3BP, LTF, LYZ, PIGR, PIP, PROL1, PRR4, SLPI, ZG16B</td>
<td>44,45</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>TMT</td>
<td>SERPINA3</td>
<td>48</td>
</tr>
<tr>
<td>Mycotic keratitis</td>
<td>2D SDS-PAGE</td>
<td>ALB, CST1, CST4, GLRX5, LCN1, PIP</td>
<td>46</td>
</tr>
<tr>
<td>Vernal keratoconjunctivitis</td>
<td>iTRAQ</td>
<td>ALB, HPX, SCGB1D1, SCGB2A1, TF</td>
<td>47</td>
</tr>
</tbody>
</table>
Table 3: Salivary proteins identified as biomarkers revealed by quantitative proteomics.

The proteins in italic have antimicrobial and immunomodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline, asterix) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autism spectrum disorder</td>
<td>2D SDS-PAGE</td>
<td>ALB , AMY1A, ATP6V1C1, <strong>BPIFA2</strong>, CA6, <strong>CST5</strong>, FRAT1, GRTP1, <strong>IGHA1</strong>, <strong>JCHAIN</strong>, LCN1P1, MUC16, PIP, TF</td>
<td>70</td>
</tr>
<tr>
<td>Bisphosphonate-related osteonecrosis of the jaw</td>
<td>iTRAQ</td>
<td>A1BG, <em>A2ML1</em>, ACTG2, ADH7, <strong>AKR1B10</strong>, ALDH3A1, ALDH9A1, <strong>ANXA1</strong>, ANXA2, ANXA4, APOA2, <strong>ARPC3</strong>, <strong>ARPC4</strong>, ATP5A1, ATP5B, <strong>BASP1</strong>, C1R, C5, CA2, CAB39L, CALML5, CAPN1, CARHSP1, CCT5, CCT8, CEACAM5, CHAD, CHI3L2, CLTC, CRABP2, CRNN, <strong>CSTA</strong>, <strong>CTSA</strong>, CUTA, DSP, ECM1, EEF1A1, EEF1G, EIF5A, EVPL, FAM49B, FKBP1A, FLG, GBP6, GNB2, GSTP1, H2AFY, HBD, HBG2, <strong>HIST1H1B</strong>, <strong>HIST1H2AA</strong>, <strong>HIST1H4A</strong>, <strong>HIST2H2BE</strong>, <strong>HIST2H2BF</strong>, <strong>HIST3H2A</strong>, <strong>HIST3H3</strong>, HSP90AA1, HSP90AB1, HSP90B1, HSPA1L, HSPB1, HSPD1, <strong>HTN1</strong>, <strong>IGHD</strong>, <strong>IGHG1</strong>, <strong>IGLC2</strong>, <strong>IL36A</strong>, IQGAP1, ITIH2, IVL, JUP, KRT10, KRT13, KRT14, KRT15, KRT16, KRT19, KRT25, KRT3, KRT4, KRT5, KRT6A, KRT73, KRT76, KRT78, LGALS7, LMNA, MGP, MMP9, MYH14, MYH9, MYL6, NPEPPS, <strong>ORM1</strong>, PCBP2, PDA6, PFKL, PKP1, PLA2G2A, PLEC, PLS3, PPL, <strong>PRB3</strong>, PRDX1, PRDX6, RAB10, RDX, RHOA, RHOB, RNASE4, RNASET2, RNB1, RPLP0, <strong>RPS12</strong>, <strong>RPS25</strong>, <strong>RPS8</strong>, S100A14, S100A16, SAA4, SCEL, SCGB2A1, SERPINA1, <strong>SERPINA4</strong>, <strong>SERPINB2</strong>, <strong>SERPINB5</strong>, SERPIN11, SFN, TGM1, TGM3, TPM2, <strong>TPT1</strong>, TUBB1, UGP2, VCP, YWHAZ</td>
<td>79</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>iTRAQ</td>
<td><strong>AKR1E2</strong>, ALB, ALDOA, <strong>AMY1A</strong>, ANXA1, ANXA3, APOA1, <strong>ARHGDI1B</strong>, AZGP1, <strong>B2M</strong>, <strong>BPIFA2</strong>, <strong>BPIFB2</strong>, C3, CA1, CA6, <strong>CALM1</strong>, CAT, <strong>CRISP3</strong>, <strong>CST1</strong>, <strong>CST2</strong>, <strong>CST3</strong>, <strong>CST4</strong>, <strong>CST5</strong>, <strong>CSTA</strong>, <strong>CSTB</strong>, <strong>DMBT1</strong>, ENO1, FABP5, FAM25A, FAM25B, FAM25C, FAM25G,</td>
<td>71, 72</td>
</tr>
<tr>
<td>Chronic graft-versus-host disease</td>
<td>iTRAQ, Label-free quantification</td>
<td>A2ML1, ACTB, ALDH3A1, ALDOA, AMY1A, ATP6AP1, AZGP1, B2M, BP1FA1, BP1FA2, BP1FB1, BP1FB2, C4orf40, C6orf58, C6A, CALM1, CALML5, CAMP, CEACAM5, CHI3L2, CRISP3, CRNN, CST1, CST2, CST3, CST4, CST5, CST6, CTSB, CTSD, CYCS, DBL, DEFA1, DFKZp686C15213, DFKZp686K18196, DMBT1, DSC2, DSG1, DSG3, ELANE, ENO1, ERO1A, EZR, FAM3B, FAM3D, FCGBP, FGA, FGB, FGG, GAPDH, GOLM1, GRN, GSTP1, HBA2, HBB, HP, HRP, HSPA5, HTN1, HTN3, IGHA1, IGHA2, IGHM, IGHV3-23, IGHV3-49, IGKC, IGKV4, IGL@, ILIRN, IGLC2, IGLL5, JCHAIN, KLK1, KRT1, KRT10, KRT16, KRT2, KRT6A, KRT9, KTR1, LCNI, LCN2, LEG1, LGALS3BP, LOC652694, LOC654188, LPO, LTA4H, LTF, LYPD3, LYZ, MPO, MUC5B, MUC7, NUCB2, P4HB, PGD, PGK1, PI3, PIGR, PIP, PKM, PKM2, PLTP, PPIB, PRB2, PRB3, PRB4, PRDX5, PRDX6, PRH1, PRR27, PRR4, PRTN3, PSAP, S100A7, S100A9, SCGB1A1, SERPINA3, SERPINB3, SERPINB5, SFN, SLPI, SMR3B, SPARCL1, SPINK7, SPRR2B, SPRR3, TALDO1, TCN1, TGM3, TIPM1, TKT, TTN, TXN, TXNDC17, Vκ3, YWHAZ, ZG16B</td>
<td>76, 77</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>SELDI-TOF</td>
<td>ARHGDIB, S100A8, S100A9</td>
<td>81</td>
</tr>
<tr>
<td>Malignant lesions</td>
<td>iTRAQ</td>
<td>ACTB, MRLC2</td>
<td>73</td>
</tr>
<tr>
<td>OSCC</td>
<td>Label-free</td>
<td>A1BG, A2M, AHSG, APOA1, APOA2, APOA4, AZGP1, BP1FB2, C3, C4B, C4BPA, C6, C9, CD109,</td>
<td>74, 75, 82,</td>
</tr>
<tr>
<td>Condition</td>
<td>Technique</td>
<td>Proteins</td>
<td>Cases</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>SRM</td>
<td>CFB, CHIT1, CRISP3, CST1*, CST2, CST4, CSTB, DSG1, DSG3, EGF, EPHX1, FABP4, FAM49B, FGB, FGG, FSCN1, GSTM1, HBA1, HEXA, HIST1H2AA, HK3, HP, HPR, HPX, HSD17B4, ICAM3, IGHA1, IGHA2, JCHAIN, ILV102, ITIH2, ITIH4, IκVI, IκVIV, KLG1, KNGL, LCN2, LCP1, LPA, MIF, MUC5B, ORM1, PPIA, PRCP, PRTN3, RAC2, RETN, RPL7, SERPINA1, SERPINA6, SERPIND1, SLC4A1, SPINK5, TF, TLN1, TMEM132A, TPI1, TTR, TUBA1C, VIM, VTN, YWHAZ</td>
<td>83</td>
</tr>
<tr>
<td>Sjögren's syndrome</td>
<td>2D SDS-PAGE Label-free quantification SELDI-TOF*</td>
<td>ALB, AMY1A, AMY2B, ANXA1, APOA1, AZGP1, B2M*, BPFA2, BPFB2, C3, CA6*, CFB, CLU, CST1, CST2, CST3, CST5, DMBT1, HBA1, HBB, HTN1, HTN3, IGHA1, IGHA2, IGHG1, IGHG3, IGHM, IGKC, IGLCL2, IGLL5, JCHAIN, KLG1, KRT1, KRT10, KRT13, KRT16, KRT4, KRT5, KRT6C, KRT9, LCN2, LGAL3BP, LPO, LTF*, LYZ, MGP, MUC5B, NUCB1, NUCB2, PIGR*, PIP, PLTP, PRB1, PRB2, RR27, PSAP, S100A8, S100A9, SERPINA1, SFN, SLPI, SMR3B, STATH, TF, ZG16B</td>
<td>78, 84</td>
</tr>
</tbody>
</table>
Table 4: Sweat proteins identified as biomarkers by quantitative proteomics analyzes.

The proteins in italic have antimicrobial and immunmodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline, asterix) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atopic dermatitis</td>
<td>SELDI-TOF</td>
<td>DCD</td>
<td>99</td>
</tr>
<tr>
<td>Ectodermal dysplasia</td>
<td>Label-free quantification</td>
<td>A1BG, A2ML1, ACTN4, AHSG, ALAD, AMY1A, ANXA2, AP1B1, ARG1, ARHG1, ASAH1, ATP5B, ATRN, AZU1, BP1A1, BP1B1, BRN, CALM2, CAPG, CAT, CD44, CEACAM5, CHIT1, CLU, CNTRL, COPB2, COTL1, CPA4, CPE, CPM, CPNE3, CREG1, CRISP3, CTSA, CTSB, CTSD, CTSV, CYB5R2, CST1, CST2, CST4, CST5, CSTB, DCD, DDAH1, DDAH2, DDB1, DMD, DSC1, DSG1, DTD1, ECM1, EEF2, ELANE, ENDO1, FABP5, FAHD1, FASN, FCGBP, FLNA, GBA, GGCT, GM2A, GPl, GSS, GSTP1, HARS2, HBA1, HEBP2, HRSP12, HSD17B4, HSPA5, IDS, IGHA2, IGHM, IGLL5, IL1RN, IL37, INTS7, JUP, KLK10, KLK11, KLK5, KLK7, KLK8, KRT33B, KRT35, KRT4, KRT73, KRT77, KRT82, KRT83, KRT85, LACRT, LAMP1, LCN1, LCN15, LDHA, LDHC, LEG1, LGALS1, LRG1, LYNX1, LYPD3, LYPD5, LYPLA1, LZC, MDH1, MDH2, MMP8, MMP9, MUC5AC, MUC7, MYDGF, MYH9, NQO2, NUDT5, NUTF2, ORM2, PEBP1, PFN1, PGAM1, PGK1, PGM2, PIP, PITH1, PLBD1, PM20D1, PNP, PPIB, PRCP, PSMA7, QPCT, RAB7A, REXO2, RNASE7, RNASET2, S100A18, S100A6, S100A8, S100A9, SBSN, SCGB1D2, SCGB2A2, SCPEP1, SERPINA12, SERPINB1, SERPINB13, SERPINB2, SERPINB8, SERPINC1, SIAE, SLPI, TALDO1, TF, TGM1, TOLLIP, VPS26A, ZG16B</td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Label-free quantification, SRM*</td>
<td>ANXA5*, ARG1*, AZGP1*, BLMH*, CALML3, CALML5*, CASP14*, CDSN*, CSTA*, DCD, DSG1*, FLNB, GAPDH*, GOT1, KLK11*</td>
<td>97</td>
</tr>
<tr>
<td>KRT10, KRT1B, KRT2, KRT6A, KRT9, MDH2, NUCB1, PARK7*, PEBP1*, PIP*, PKM, PRDX1*, S100A7*, TPI1, TXN*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Biomarkers in nose water revealed by quantitative proteomics

The proteins in italic have antimicrobial and immunmodulatory effects, while the bold face refers to validated biomarkes.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic rhinosinusitis</td>
<td>Label-free quantification</td>
<td>BP1A, BPIFB1, CHI3L1, LCP1, <strong>MUC5B, SERPINB10</strong></td>
<td>107</td>
</tr>
</tbody>
</table>
Table 6: Vaginal fluid protein biomarkers identified by quantitative proteomics methods.

The proteins in italic have antimicrobial and immunmodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline, asterix) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervical cancer</td>
<td>Label-free quantification</td>
<td>ACTN4, ACTR3, CFH, CRABP2, EEF1A1P5, GC, NAMPT, PGK1, SERPINB13, SERPINB3, YWHAE</td>
<td>126</td>
</tr>
<tr>
<td>HIV-infection</td>
<td>iTRAQ</td>
<td>PRTN3, SERPINA5</td>
<td>127</td>
</tr>
<tr>
<td>Intra-amniotic inflammation</td>
<td>Label-free quantification SELDI-TOF*</td>
<td>A1BG, C3, CSTA, DEFA1*, DEFA2*, DEFA3*, DSG3, DSP, FABP5, FN1, GC, HP, IGFBP1, IFL, KLK13, LMNA, LTE, LYZ, MMP9, ORM1, PPL, S100A7, S100A8*, S100A9*, SERPINA1, SERPINB3, SERPINB13, SPRR3, VCL</td>
<td>112, 120</td>
</tr>
<tr>
<td>Preterm labour/ preeclampsia</td>
<td>2D-SDS-PAGE SELDI-TOF SRM*</td>
<td>ACBP, ACTB, ALB, AMBP, ANXA1, ANXA3, APOA1, COL, CSTA, DSP*, FABP5, FGB, GC, GGCT, GSTP1, HP, IL1RN, MICA, ORM1, PRDX1, PRDX2, RBP, S100A7, S100A9, SERPINA1, SERPINB1, SERPINB3, SFN*, SOD1, TCEA2, TF, THBS1*, TPM1, TXN</td>
<td>121, 122, 123, 124, 125</td>
</tr>
</tbody>
</table>
Table 7: Protein biomarkers identified in urine by quantitative proteomics techniques.

The proteins in italic have antimicrobial and immunmodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline, asterix, etc.) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder cancer</td>
<td>2D SDS-PAGE</td>
<td>A2M, ALB, AFM&lt;sup&gt;a&lt;/sup&gt;, AGT&lt;sup*&gt;&lt;/sup&gt;, AHSG&lt;sup&gt;a&lt;/sup&gt;, APOA1, APOA2&lt;sup&gt;a&lt;/sup&gt;, APOA4, APOB, APOC2, APOC3, APOE, APOH, APOL1&lt;sup&gt;a&lt;/sup&gt;, AXL, AZGP1, BCAN, C2, C3, C9&lt;sup&gt;a&lt;/sup&gt;, CADM1, CALR, CD44&lt;sup*&lt;/sup&gt;, CDH2, CEL, CLEC4G, CNTN1, COL15A1, CP, CPQ, CST3, CTSA, EGF, F2&lt;sup&gt;a&lt;/sup&gt;, FABP5, FBN1, FGA, FGB, FGG, FN1, GAA, GC, GGH, GUCA2A, GUSB, HPX&lt;sup*&lt;/sup&gt;, HSPG2, IGFBP7, ITIH4&lt;sup&gt;a&lt;/sup&gt;, GALS7, MGAM, MMP2, MMP9, NID1, ORM1, PGA5, PLG&lt;sup&gt;a&lt;/sup&gt;, PRCP, PRG4, PROZ, PSAP&lt;sup*&lt;/sup&gt;, QPCT, QSOX1, RBP4, ROBO4, S100A7, S100P, SAA4, SDC4, SERPINA1&lt;sup&gt;<em>#&lt;/sup&gt;, SERPINC1, SOD3, SPP1&lt;sup</em>&lt;/sup&gt;, TF&lt;sup&gt;<em>&lt;/sup&gt;, THBS1&lt;sup&gt;a&lt;/sup&gt;, TNXB, TTR, UMOD&lt;sup</em>&lt;/sup&gt;, YIPF3</td>
<td>150, 151, 152, 153, 154, 156, 157, 160, 161</td>
</tr>
<tr>
<td></td>
<td>iTRAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Label-free quantification&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SRM&lt;sup&gt;#&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Label-free quantification</td>
<td>ALB, AGRN, ANXA1, AHSG, APOA4, C3, C9orf131, CA1, CSTA, DNAH8, HBA1, IGHG2, IkV, ITIH4, LCN2, LRRC36, LTF, MAST4, NEGR1, ORM1, PGA3, SPARCL1, SULF2, VTN</td>
<td>166</td>
</tr>
<tr>
<td>Chronic allograft nephropathy</td>
<td>SELDI-TOF</td>
<td>B2M</td>
<td>146</td>
</tr>
<tr>
<td>Congenital obstructive nephropathy</td>
<td>Label-free quantification</td>
<td>ARG1, AMY1A, AMY2B, CASP14, CDH1, CUBN, EGF, FLNB, IGFBP7, KLK1, LRRC15, PDCD1LG2, PTGDS, SERPINA5, TGM3, VCAN</td>
<td>147</td>
</tr>
<tr>
<td>Coronary artery atherosclerosis</td>
<td>Label-free quantification</td>
<td>COL1A1, COL1A3</td>
<td>163</td>
</tr>
<tr>
<td>Diabetic nephropathy</td>
<td>2D SDS-PAGE</td>
<td>APOA4, CP, EGF, PIP, SERPINA1, TTR</td>
<td>140, 141</td>
</tr>
<tr>
<td></td>
<td>iTRAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney chronic allograft dysfunction</td>
<td>Label-free quantification</td>
<td><em>KNG1</em>, <em>UMOD</em></td>
<td>145</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------------------------</td>
<td>-----------------</td>
<td>-----</td>
</tr>
<tr>
<td>Lupus nephritis</td>
<td>2D SDS-PAGE</td>
<td>AMBP, <em>AZGP1</em>, <em>IGKC</em></td>
<td>148</td>
</tr>
<tr>
<td>Obstructive sleep apnea</td>
<td>2D SDS-PAGE</td>
<td>AMBP, EGLN2, KAT2B, <strong>KLK1</strong>, <strong>ORM2</strong>, TNC, TRIB2, <strong>UCN3</strong>, <em>UMOD</em>, ZFP36, ZNF81</td>
<td>164</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>2D SDS-PAGE</td>
<td><strong>RNASE2</strong></td>
<td>165</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>iTRAQ SRM</td>
<td>A2M, ADGRF5, ADGRL1, <em>ADM</em>, AGT, AFM, ALB, AMY1A, AMY2A, ANGPTL2, ANPEP, ANXA11, APOA1, APOA2, APOA4, APOB, APOD, APOE, ART3, <em>AXL</em>, B2M, B3GNT2, B4GAT1, BHMT, BLMH, <em>C1RL</em>, C3, C4B, C5, C7, CA1, CADM1, CASP14, CD14, CD248, CDH11, CDH13, CDH2, CDH6, CDHR5, CEL, CETP, CFB, CFD, CFH, CHL1, CILP, CILP2, CNTN1, COL15A1, COL6A1, CPN2, CRB2, CRYAB, CRYM, CST6, DPEP1, DSC1, EGF, FAT4, FBN1, FBP1, <em>FGGR3B</em>, FGB, FGG, FLG, FLRT2, FN1, FREM2, GAA, GALNS, GAS6, GC, GGCT, GOLM1, GP5, GPC1, GPC3, GPC4, GSTM3, GSTT1, GUCA2B, <em>HP</em>, HPX, <em>HRG</em>, <em>HSPB1</em>, <em>HSPG2</em>, <em>HYAL1</em>, ICAM2, IGFBP6, IGFBP7, ISLR, ITFG1, ITIH1, ITIH2, KRT4, KRT6A, LCAT, <em>LGALS3BP</em>, LRRC15, MAN1A1, MB, MCAM, MFI2, MME, MMRN2, MUC20, MXRA8, MYOC, OGN, OLR1, <strong>ORM1</strong>, OSCAR, PCDH1, PCDHC3, PDCD6, PFN1, PHPT1, PON1, PROM2, PSAP, PSG11, PSG9, PVALB, PVR, PVRL1, PVRL3, QPCT, QSOX1, RBP4, REG1A, ROBO4, SDF4, SDK1, SERPINA1, SERPINA4, SERPINA5, SERPINC1, SERPIND1, SERPINF1, SERPINI1, SH3BGRL, SPRR3, <strong>SUSD2</strong>, TF, TGM4, TNXB, TPP1, TTR, VASN</td>
<td>144</td>
</tr>
<tr>
<td>Cancer Type</td>
<td>Technique</td>
<td>Proteins/Enzymes</td>
<td>References</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>2D SDS-PAGE</td>
<td>S100A9</td>
<td>143</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td>SELDI-TOF</td>
<td>ALB, IGLC, MASP2, SECTM1, VMO1</td>
<td>158, 159</td>
</tr>
<tr>
<td>Sepsis</td>
<td>iTRAQ</td>
<td>A1BG, AFM, AMY2A, CDH1, CEACAM8, DPP4, GM2A, HP, HSPG2, IκV, KRT1, KRT10, KRT16, KRT5, KRT9, LAMP1, LCN1, MUC1, SELENBP1</td>
<td>167, 168</td>
</tr>
</tbody>
</table>
Table 8: Number of antimicrobial peptides/proteins among the identified potential biomarkers

<table>
<thead>
<tr>
<th>Disease</th>
<th>Body fluid</th>
<th>Technique</th>
<th>Number of potential biomarkers</th>
<th>Number of AMPs</th>
<th>MS instrument type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blepharitis</td>
<td>Tear</td>
<td>2D SDS-PAGE</td>
<td>8</td>
<td>5</td>
<td>Q-TOF</td>
</tr>
<tr>
<td>Climatic droplet keratopathy</td>
<td>Tear</td>
<td>iTRAQ</td>
<td>5</td>
<td>4</td>
<td>Q-TOF</td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>Tear</td>
<td>iTRAQ</td>
<td>6</td>
<td>5</td>
<td>4000QTRAP</td>
</tr>
<tr>
<td>Dry eye syndrome</td>
<td>Tear</td>
<td>iTRAQ, SELDI-TOF</td>
<td>33</td>
<td>21</td>
<td>Q-TOF, LTQ-Orbitrap, PBS-IIc ProteinChip Reader</td>
</tr>
<tr>
<td>Keratoconus</td>
<td>Tear</td>
<td>Label-free quantification</td>
<td>26</td>
<td>19</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Tear</td>
<td>TMT</td>
<td>1</td>
<td>0</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Mycotic keratitis</td>
<td>Tear</td>
<td>2D SDS-PAGE</td>
<td>6</td>
<td>4</td>
<td>MALDI-TOF</td>
</tr>
<tr>
<td>Vernal keratoconjunctivitis</td>
<td>Tear</td>
<td>iTRAQ</td>
<td>5</td>
<td>0</td>
<td>MALDI TOF-TOF</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>Saliva</td>
<td>2D SDS-PAGE</td>
<td>14</td>
<td>7</td>
<td>Q-TOF</td>
</tr>
<tr>
<td>Bisphosphonate-related osteonecrosis of the jaw</td>
<td>Saliva</td>
<td>iTRAQ</td>
<td>135</td>
<td>41</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Saliva</td>
<td>iTRAQ</td>
<td>92</td>
<td>50</td>
<td>Q-TOF</td>
</tr>
<tr>
<td>Chronic graft-versus-host disease</td>
<td>Saliva</td>
<td>iTRAQ, Label-free quantification</td>
<td>140</td>
<td>46</td>
<td>TOF-TOF, LTQ-Orbitrap</td>
</tr>
<tr>
<td>Fibromyalgia</td>
<td>Saliva</td>
<td>SELDI-TOF</td>
<td>3</td>
<td>2</td>
<td>ProteinChip SELDI system</td>
</tr>
<tr>
<td>Condition</td>
<td>Sample Type</td>
<td>Quantification Method</td>
<td>M</td>
<td>N</td>
<td>Mass Spectrometer</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------</td>
<td>----------------------------</td>
<td>---</td>
<td>---</td>
<td>------------------</td>
</tr>
<tr>
<td>Malignant lesions</td>
<td>Saliva</td>
<td>iTRAQ</td>
<td>2</td>
<td>0</td>
<td>LTQ-LIT</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Saliva</td>
<td>SRM</td>
<td>1</td>
<td>0</td>
<td>5500QTRAP</td>
</tr>
<tr>
<td>OSCC</td>
<td>Saliva</td>
<td>Label-free quantification, SELDI-TOF</td>
<td>76</td>
<td>30</td>
<td>LTQ-Orbitrap ProteinChip biology system</td>
</tr>
<tr>
<td>Sjögren's syndrome</td>
<td>Saliva</td>
<td>Label-free quantification</td>
<td>65</td>
<td>35</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>Sweat</td>
<td>SELDI-TOF</td>
<td>1</td>
<td>1</td>
<td>Protein Chip Array</td>
</tr>
<tr>
<td>Ectodermal dysplasia</td>
<td>Sweat</td>
<td>Label-free quantification</td>
<td>159</td>
<td>42</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Sweat</td>
<td>Label-free quantification, SRM</td>
<td>30</td>
<td>9</td>
<td>LTQ-Orbitrap TSQ Quantum Ultra</td>
</tr>
<tr>
<td>Chronic rhinosinusitis</td>
<td>Nose water</td>
<td>Label-free quantification</td>
<td>6</td>
<td>3</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Bladder cancer</td>
<td>Urine</td>
<td>2D SDS-PAGE, iTRAQ, Label-free quantification, SRM</td>
<td>79</td>
<td>14</td>
<td>MALDI-TOF LTQ-Orbitrap 4000QTRAP</td>
</tr>
<tr>
<td>Chronic allograft nephropathy</td>
<td>Urine</td>
<td>SELDI-TOF</td>
<td>1</td>
<td>1</td>
<td>Protein Chip Array</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Urine</td>
<td>Label-free quantification</td>
<td>24</td>
<td>9</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Congenital obstructive nephropathy</td>
<td>Urine</td>
<td>Label-free quantification</td>
<td>16</td>
<td>1</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Coronary artery atherosclerosis</td>
<td>Urine</td>
<td>Label-free quantification</td>
<td>2</td>
<td>1</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Condition</td>
<td>Sample Type</td>
<td>Technique</td>
<td>n</td>
<td>m</td>
<td>Instrumentation</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>---</td>
<td>---</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Diabetic nephropathy</td>
<td>Urine</td>
<td>2D SDS-PAGE</td>
<td>6</td>
<td>1</td>
<td>SELDI-TOF 5600Triple-TOF</td>
</tr>
<tr>
<td>Kidney chronic allograft dysfunction</td>
<td>Urine</td>
<td>Label-free quantification</td>
<td>2</td>
<td>1</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Lupus nephritis</td>
<td>Urine</td>
<td>2D SDS-PAGE</td>
<td>3</td>
<td>2</td>
<td>TOF-TOF</td>
</tr>
<tr>
<td>Obstructive sleep apnea</td>
<td>Urine</td>
<td>2D SDS-PAGE</td>
<td>11</td>
<td>1</td>
<td>MALDI-TOF</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>Urine</td>
<td>iTRAQ</td>
<td>151</td>
<td>19</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Urine</td>
<td>2D SDS-PAGE</td>
<td>1</td>
<td>1</td>
<td>MALDI-TOF</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td>Urine</td>
<td>SELDI-TOF</td>
<td>5</td>
<td>3</td>
<td>ProteinChip Reader PCS 4,000</td>
</tr>
<tr>
<td>Sepsis</td>
<td>Urine</td>
<td>iTRAQ</td>
<td>19</td>
<td>4</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>Vaginal fluid</td>
<td>Label-free quantification</td>
<td>11</td>
<td>3</td>
<td>LTQ-Orbitrap</td>
</tr>
<tr>
<td>HIV-infection</td>
<td>Vaginal fluid</td>
<td>iTRAQ</td>
<td>2</td>
<td>1</td>
<td>MALDI-TOF-TOF</td>
</tr>
<tr>
<td>Intra-amniotic inflammation</td>
<td>Vaginal fluid</td>
<td>Label-free quantification</td>
<td>29</td>
<td>12</td>
<td>LTQ-Orbitrap Protein Chip Array</td>
</tr>
<tr>
<td>Preterm labour/ preeclampsia</td>
<td>Vaginal fluid</td>
<td>2D-SDS-PAGE SELDI-TOF SRM</td>
<td>34</td>
<td>13</td>
<td>LTQ-LIT MALDI-TOF Linear ion trap</td>
</tr>
</tbody>
</table>
Table 9: Extracellular vesicle biomarkers identified by quantitative proteomics methods.

The proteins in italic have antimicrobial and immunomodulatory effects, while the bold face refers to validated biomarkers. The other font settings (normal, underline) refer to the method by which the biomarkers were identified.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Method</th>
<th>Potential biomarkers (Gene symbol)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder cancer</td>
<td>Label-free quantification, Dimethyl labeling</td>
<td><strong>TACSTD2, GSA, RETN, GPRC5A</strong></td>
<td>211, 212</td>
</tr>
<tr>
<td>Brain tumor</td>
<td>2D-DIGE</td>
<td><strong>EEF1B, PSMA1, 3DPGH, ARRDC2, NANS, CENP-P, EIF3B, PCNA, MTAP, GPNMB, EGFR, HSC70, HSPD1</strong></td>
<td>215</td>
</tr>
<tr>
<td>Deep venous thrombosis</td>
<td>2D SDS-PAGE, iTRAQ</td>
<td><strong>Gal3BP, A2M</strong></td>
<td>213,214</td>
</tr>
<tr>
<td>Nephropathy/nephrotic syndrome</td>
<td>Label-free quantification</td>
<td><strong>ANPEP, VASN, SERPINA1, CP</strong></td>
<td>209, 210</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Label-free quantification</td>
<td><strong>FASN, XPO1, PDCD6IP, CD9, ENO1</strong></td>
<td>208</td>
</tr>
</tbody>
</table>
Figure 1
Graphical abstract