Delineating the synovial fluid proteome: Recent advancements and ongoing challenges in biomarker research

Daniela Cretu1,2, Eleftherios P. Diamandis1,2,3, and Vinod Chandran4,5

1Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada, 2Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada, 3Department of Clinical Biochemistry, University Health Network, Toronto, Ontario, Canada, 4Centre for Prognosis Studies in the Rheumatic Diseases, Toronto Western Hospital, Toronto, Canada, and 5Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, Canada

Abstract

There is an urgent need for identifying novel serum biomarkers that can be used to improve diagnosis, predict disease progression or response to therapy, or serve as therapeutic targets for rheumatic diseases. Synovial fluid (SF) is secreted by and remains in direct contact with the synovial membrane, and can reflect the biochemical state of the joint under different physiological and pathological conditions. Therefore, SF is regarded as an excellent source for identifying biomarkers of rheumatologic diseases. The use of high-throughput and/or quantitative proteomics and sophisticated computational software applied to analyze the protein content of SF has been well-adopted as an approach to finding novel arthritis biomarkers. This review will focus on some of the potential pitfalls of biomarker studies using SF, summarize the status of the field of SF proteomics in general, as well as discuss some of the most promising biomarker study approaches using proteomics. A brief status of the biomarker discovery efforts in rheumatoid arthritis, osteoarthritis and juvenile idiopathic arthritis is also provided.

Abbreviations: 2D-PAGE: two-dimensional polyacrylamide gel electrophoresis; ACL: anterior cruciate ligament; ACPA: anti-citrullinated protein antibody; ADAMTS: a disintegrin and metalloproteinase with thrombospondin motifs; ANA: anti-nuclear antibody; COL2: type II collagen; COMP: cartilage oligomeric matrix protein; CPP: cyclic citrullinated peptide; CRP: C-reactive protein; CSF: cerebrospinal fluid; CV: coefficient of variation; CYP1A1: Cytochrome P450, family 1, subfamily A, polypeptide 1; DIGE: difference gel electrophoresis; EDTA: ethylenediaminetetraacetic acid; ELISA: enzyme-linked immunosorbent assay; ESI: electrospray ionization; GAG: glycosaminoglycan; HA: hyaluronic; HLA: human leukocyte antigen; ICAT: isotope coded affinity tag; IGF: insulin growth factor; IL: interleukin; iTRAQ: isobaric tags for relative and absolute quantitation; JIA: juvenile idiopathic arthritis; LC: liquid chromatography; LFQ: label-free quantification; LIF: leukemia inhibitory factor; MALDI: matrix-assisted laser desorption/ionization; MGP: matrix Gla protein; MMP: matrix metalloproteinase; MRM: multiple reaction monitoring; MRP: myeloid-related protein; MS: mass-spectrometry; MS/MS: tandem mass spectrometry; MudPIT: multidimensional protein identification technology; NGAL: neutrophil gelatinase-associated lipocalin; OA: osteoarthritis; pI: isoelectric point; PRG4: proteoglycan 4; PsA: psoriatic arthritis; QqQ: triple quadrupole; RA: rheumatoid arthritis; RF: rheumatoid factor; RP: reverse phase; SAA: serum amyloid A; SCX: strong cation exchange; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; SELDI: surface-enhanced laser desorption/ionization; SERPIN: serine protease inhibitor; SF: synovial fluid; SID-SRM: stable isotope dilution-selected reaction monitoring; SLE: systemic lupus erythematosus; SM: selected reaction monitoring; TGF: transforming growth factor; TIMP: tissue inhibitor of metalloproteinase; TOF: time of flight; TUB: tubulin; VDBP: vitamin D binding protein; VIME: vimentin

Referee: Professor Michael Glocker, Institute for Immunology, Faculty of Medicine, University of Rostock, Rostock, Germany.
Address for correspondence: Dr Vinod Chandran, Toronto Western Hospital, 399 Bathurst Street, Room 1E 416, Toronto, ON, M5T 2S8, Canada. E-mail: vchandra@uhnresearch.ca

Introduction

To understand and study joint diseases, we must have a thorough understanding of three joint components: synovial fluid (SF), the synovial membrane and the articular cartilage. The synovial membrane is a layer of cells (macrophages and
synovial fibroblasts) one to three cells deep\(^1\), embedded in a collagen and hyaluronan-rich matrix\(^2\). Although it lacks a basement membrane, the intimal matrix of the membrane behaves as a semipermeable coating as it comes in contact with the blood contents in the superficial capillary network. The surface of articular joints is covered by articular cartilage, which is mostly comprised of chondrocytes embedded in a matrix of collagen and proteoglycans (Figure 1). SF is secreted by the synovial membrane, and is in direct contact with both the synovial membrane and the articular cartilage. It is a hyaluronic acid-rich fluid and, under normal conditions, it lubricates and provides articular cartilage with the essential nutrients necessary for chondrocyte metabolism. It also serves as the intermediate carrier of proteins shed by the articular cartilage and transferred to the systemic circulation\(^3\). The blood-joint barrier has been modeled as a double barrier, in series, consisting of synovial interstitial space that limits diffusion of small molecules, and microvascular endothelium that limits transport of proteins\(^1\). SF is normally a clear, straw-colored, viscous liquid present at volumes of \(\sim 1\) mL in normal joints. The cellular and molecular constituents within SF give rise to its unique properties and functions in maintaining joint homeostasis. The total protein concentration of normal SF is 19–28 mg/mL, which includes blood plasma dialysate and molecules secreted by cells lining the synovial joint space. The composition and function of SF is altered in joint injury and disease due to changes directly to the SF, as well as to the tissues lining the synovial joint\(^4\). Changes in the cellular metabolism and structure of these tissues as they occur in a disease state may be reflected by changes in SF function and composition. We can exploit this particular characteristic of SF when investigating potential biomarkers of joint disease.

A biomarker is defined as a measureable indicator of a specific biological state – in particular, one that indicates information about the risk, presence or stage of a disease\(^5\). These biomarkers can be used in the clinic to diagnose (diagnostic), predict disease progression (prognostic), monitor activity of the disease, assess therapeutic response (screening) or guide molecular targeted therapy\(^5\). Biomarkers for joint diseases may come in many forms: they may be clinical, histological or imaging parameters, as well as specific molecules, or molecular patterns\(^6\). Molecular biomarkers include genomic, proteomic and transcriptomic biomarkers. Table 1 contains a list of biomarkers currently used in the diagnosis and treatment of joint diseases.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Molecular class</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creatinine</td>
<td>Metabolite</td>
<td>Drug toxicity(^{130})</td>
</tr>
<tr>
<td>CRP</td>
<td>Protein</td>
<td>Identify acute inflammation(^{131})</td>
</tr>
<tr>
<td>ANA</td>
<td>Autoantibody</td>
<td>Diagnostic of SLE(^{132})</td>
</tr>
<tr>
<td>RF</td>
<td>Autoantibody</td>
<td>Diagnosis for RA(^{133,134})</td>
</tr>
<tr>
<td>ACPA</td>
<td>Autoantibody</td>
<td>Diagnosis and prognosis for RA(^{135,136})</td>
</tr>
<tr>
<td>Anti-dsDNA</td>
<td>Autoantibody</td>
<td>Diagnosis and monitoring for SLE(^{137})</td>
</tr>
<tr>
<td>HLA-DRB1 shared epitope alleles</td>
<td>Genomic</td>
<td>Prognosis for RA(^{138,139})</td>
</tr>
</tbody>
</table>

Figure 1. Representation of the structure and pathology of the synovial joint. A comparison is made between the normal and arthritic joint to highlight changes occurring during inflammation. (A) The thin synovial membrane lines the joint space and is composed of macrophage-like and fibroblast-like synoviocytes. The arthritic synovial joint is characterized by inflammation and thickening of the synovial membrane and a consequent influx of lymphocytes and macrophages. (B) The resulting SF inflammatory environment stimulates degradation of the articular cartilage.

Table 1. List of biomarkers currently used in the diagnosis and treatment of joint diseases.
However, plasma or serum analysis presents with several challenges. These include high complexity in the number of proteins and protein isoforms, a 12-fold dynamic range of concentration between high-abundance and low-abundance proteins and changes in concentration, structure and function as a result of physiological and pathological processes. Therefore, the discovery of biomarkers from serum by shotgun proteomic analysis becomes a challenging task. As a result, and taking into consideration the role of SF in joint physiology and its close proximity to the diseased joint, SF is the ideal fluid to mine for potential disease markers. Proteins differentially expressed in the inflamed SF can be more readily mined. This is due to the fact that they are present in significantly higher concentrations in the inflamed tissues than in serum, which facilitates their identification by an unbiased discovery approach. The most promising of these identified proteins can then be sought in the serum of patients using more targeted approaches.

Four different phases exist in the discovery of novel biomarkers: discovery phase, qualification stage, validation stage and the verification stage. In the discovery phase, a small number of well-characterized, high-quality samples are compared using fractionation and quantitative proteomics methods to generate an extensive list of protein components. All the phases that follow in the biomarker development pipeline replace the unbiased experimental design with target-driven qualitative strategies relying mainly on immune analytical methods such as ELISAs, as will be discussed later. In the qualification stage, stringent selection criteria are applied to potential markers, and these are retested using targeted strategies in the original starting samples. During the verification phase, the specificity of the candidate markers is investigated in a larger number of samples, and in samples that closely represent the sample type in which a final clinical test would be employed. To advance in development, all biomarker candidates also require validation, which is undertaken on only a subset of verified candidates and is performed in, ideally, thousands of samples. Potential biomarkers showing good sensitivity and specificity are considered for further clinical evaluation.

In this review, we will focus on some critical pre-analytical factors that should be considered prior to performing proteomic biomarker experiments on SF. In addition, a selection of the most frequently used proteomics pipelines for biomarker discovery and validation will be discussed. Finally, a brief summary is given on some of the recent biomarker discovery work reported on rheumatoid arthritis (RA), osteoarthritis (OA) and juvenile idiopathic arthritis (JIA).

### The protein composition of synovial fluid in health and disease

SF contains a large number of proteins originating from the synovial membrane, cartilage and serum. The protein composition in SF may reflect the pathophysiological conditions affecting the synovial tissue and articular cartilage.

#### Plasma proteins

A major component of SF is proteins derived from plasma. Blood plasma and SF share many similarities in their protein compositions irrespective of the synovium, selectively hindering large plasma proteins from entering the joint space from vasculature. Total protein concentration in normal SF is 19–28 mg/mL – one-third of that found in plasma. The size of plasma proteins determines their filtration properties through the synovial membrane and their entry into SF: large molecular weight plasma proteins such as fibrinogen are...
present at low concentration in normal SF; in contrast, albumin and transferrin are present in relatively high abundance\(^{10,11}\). Albumin represents the major protein species in SF (\(\sim 12 \text{ mg/mL}\)) along with \(\beta_1, \lambda, \gamma_1\) and \(\gamma_2\) globulins (each at a concentration of 1–3 \(\text{mg/mL}\))\(^{10}\).

The protein content and concentrations in SF alter with inflammation\(^{10}\). Total protein concentration in SF from patients with OA, RA and psoriatic arthritis (PsA) is higher than normal, which indicates the presence of both structural and functional changes in the synovial membrane as a result of disease and joint injury. More specifically, synovial inflammation deters the ability of the synovial membrane to selectively retain and filter proteins. For example, SF from RA patients contains high levels of globulins and glycoproteins, which are high molecular weight proteins not found in normal SF\(^{10}\). The protein distribution in RA SF is also altered and resembles that of blood plasma, in which \(\beta_2\)-microglobulin, fibrinogen, \(\beta_1\)-lipoprotein, \(\alpha_2\)-macroglobulin and \(\alpha_2\)-glycoprotein are present at increased concentrations\(^{12,13}\).

**Lubricant molecules**

Lubrication of articular cartilage by SF is mediated by several lubricant macromolecules synthesized and secreted by the synovial fibroblast cell populations, which are therefore present in SF. Hyaluronan (HA)\(^{14}\) and PRG4\(^{15}\) represent the primary lubricant macromolecules in SF and are present in normal SF (\(\sim 3.2–4.1\) and \(\sim 0.035–0.24 \text{ mg/mL}\), respectively)\(^{16,17}\). HA is a non-sulfated glycosaminoglycan composed of repeating disaccharide units of D-glucuronic acid and D-N-acetylgalactosamine present in polydisperse populations with an average molecular weight of 6–7 MDa\(^{18}\). HA contributes to the viscosity of SF and provides outflow buffering. Products of the PRG4 gene include superficial zone protein (SZP) and lubricin, which are mucinous glycoproteins with multiple O-linked \((1-3)\)Gal-GalNAc oligosaccharides, functioning in boundary lubrication of articular cartilage\(^{19}\). SZP is synthesized and secreted by chondrocytes in the cartilage superficial zone, while lubricin is expressed by synovial fibroblasts\(^{20}\).

As expected, OA, RA, PsA and other forms of joint injury are associated with changes in SF lubricant macromolecules. Specifically, the mean HA concentration in pathology-associated SF is lower than that of normal SF (\(\sim 1.2–2.2 \text{ mg/mL}\) in OA SF, and \(\sim 0.7–2.7 \text{ mg/mL}\) in RA SF)\(^{21,22}\). The molecular weight distribution of HA is also altered, with a shift toward lower molecular forms (<4 MDa)\(^{23}\). Furthermore, cellular PRG4 immunostaining and mRNA levels were decreased in degenerative cartilage from an ovine meniscectomy model of OA when compared to levels found in normal cartilage\(^{24}\). SF analysis revealed an association between decreased lubricin and boundary-lubricating ability, increased elastase activity and increased cartilage degradation\(^{25}\).

**Cytokines and growth factors**

Cytokines and growth factors present in SF are important regulatory molecules for the cell populations within the synovial joint space, such as chondrocytes and synovial fibroblasts\(^{26,27}\). Regulatory molecules in SF may be derived from plasma through the selective filtration mechanisms, as previously described, or as secreted products of chondrocytes, synovial fibroblasts and other cells present within the SF or surrounding tissues. Cytokines are categorized as pro- or anti-inflammatory according to their immediate tissue-specific effects. Proinflammatory cytokines present in SF include IL-1\(\alpha\), IL-1\(\beta\), TNF-\(\alpha\), leukemia inhibitory factor (LIF), IL-6, IL-8, IL-17 and IL-18\(^{3,28–30}\). Anti-inflammatory cytokines include IL-4, IL-10 and IL-13. Several growth factors and their respective binding proteins are also found in SF, including TGF-\(\beta\) and insulin growth factor (IGF-1), which play important roles in cell regulation\(^3\).

In arthritis patients, the cytokine profile in SF is also altered. While normal SF contains cytokines and growth factors at low concentrations, pathology-associated SFs, have markedly increased levels\(^3\). Cytokines, in particular, play important roles in disease pathogenesis and joint destruction, and have attracted attention as putative therapeutic targets; several cytokine-directed therapies are already in clinical use, while others are in clinical trials\(^{31–33}\). Specifically, blocking TNF-\(\alpha\) effect with Etanercept resulted in increased amounts of articular cartilage-bound lubricin and decreased sGAG release in a rat ACL injury model of posttraumatic arthritis\(^{34}\).

**Proteolytic enzyme activity**

Degradative processes in the joint are mediated by proteolytic enzymes, which are carefully regulated\(^{35,36}\). Matrix-degrading enzymes such as MMP-1 and MMP-3 are present in normal SF and are elevated in RA, PsA and OA\(^{35,37,38}\). MMPs are secreted from chondrocytes aszymogens, and are activated following propeptide cleavage. A distintegrin and metalloproteinase with thrombospondin motifs (ADAMTS) proteinases, known to degrade aggrecan, are also secreted aszymogens, requiring subsequent activation\(^{39,40}\). Other proteinases such as serine and cysteine proteinases are also present, and are involved in the activation ofproMMPs (plasmin, kallikrein and cathepsins)\(^{36}\). Tissue inhibitors of metalloproteinases (TIMPs) and inhibitors of serine proteinases (SERPINs) known to activate proMMPs have also been identified in SF. In RA, infiltrating neutrophils and macrophages also contribute to protease levels in the joint\(^{40}\).

Changes in the levels and activities of matrix-degrading enzymes and their associated inhibitors and activators alter the fine balance between anabolism and catabolism in the joint, and these changes are indicated by elevated levels of degradation products\(^{41}\). For example, the concentration of fragments of aggrecan and type II collagen are elevated in the SF of patients with OA, PsA and RA\(^{42–45}\).

**Biological variables affecting SF protein content**

As discussed previously, the total protein concentration in SF is highly influenced by synovial membrane permeability. Molecular sieving by the synovial membrane matrix is primarily size-dependent, where high molecular weight proteins (such as HA and PRG4) are retained within the joint space in the SF, while low molecular weight species such as metabolic substrates, byproducts, growth factors and cytokines are not\(^{11}\). On the other hand, high molecular weight
species within plasma are unable to enter the joint space, and are therefore present at low concentrations in normal SF. Increased synovial inflammation in patients with various forms of arthritis is associated with proportionately greater increases in permeability to large proteins.

In RA, the synovial membrane dramatically increases in mass and in metabolic activity, resulting in infiltration of immune cells and edema. The synovial lining may increase up to 10-fold, an effect that is most often due to the increased number of synovial fibroblasts and macrophages. This also represents the site where most proinflammatory cytokines and matrix metalloproteinases are produced/activated. An increase in the number of blood vessels usually occurs, but this capillary network is most often disorganized in comparison to its normal state. Specifically in RA synovium, while permeability to large proteins is increased, permeability to small molecules (such as urea and glucose) is actually decreased. This is due to a combination of increased vascular permeability, synovial hyperplasia and cellular infiltration. Increased Starling pressure due to the reduced oncotic gradient as protein accumulates in SF, increased permeability of capillaries, and an increase in capillary pressure during inflammation contribute to the accumulation of fluid and protein in RA SF. HA contributes to outflow buffering in the synovial joint. In arthritis, the size of SF HA is reduced, and there is an increase in the rate of HA loss from SF. Although synovial membrane inflammation is also recognized as a key factor in OA pathophysiology, the transport characteristics of synovium in OA are less impacted than in RA, with reported permeabilities to proteins over a range of sizes.

Pre-analytical considerations

Several pre-analytical factors can affect the SF proteome, leading to inaccuracies in the quantitative protein measurements and the false discovery of biomarker candidates. As mentioned previously, the initial discovery step in the biomarker discovery pipeline (Figure 2) relies on the use of well-characterized, high-quality samples. It is ideal to collect and store samples following a standardized protocol, which is extremely important during the biomarker development process considering that a candidate biomarker needs to be repeatedly validated with large and independent sample sets. Although few efforts have been made toward the standardization of SF sample management, we suggest that collection and handling of SF should follow similar guidelines to those agreed upon by the Plasma Proteome Initiative. Implementation of this protocol across labs will also streamline future proteomics studies and will allow for the exchange of comparable samples between laboratories.

Sample handling

Normal SF does not clot because it does not contain sufficient fibrinogen. However, during inflammation, as discussed previously, there is an influx of clotting components from the peripheral blood supply. Therefore, pathological fluids clot, and the speed and size of clot formation typically relate to the severity of joint inflammation. SF storage does not normally require anticoagulants, though samples can also be collected in K-EDTA blood tubes for convenience. Clots usually form if samples are left for over 1 h at room temperature prior to centrifugation, which also compromises the stability of the SF proteome.

Furthermore, cellular material should be removed from samples by centrifugation, aliquots should be made and the resulting material should be stored at −80 °C. The importance of optimal storage conditions can be illustrated by a study performed by Carrette et al. where they identified a 12.5 kDa N-terminal cleavage product of the 13.4 kDa cystatin C resulting from storing cerebrospinal fluid (CSF) samples at −20 °C compared to −80 °C. Unfortunately, no such studies have been performed in SF, but a similar study in serum demonstrates the instability of several proteins at higher temperatures. Protease inhibitors can be added to the sample to protect against degradation during pre-fractionation, although some peptide components such as leupeptin in inhibitor cocktails can complicate mass spectrometric identifications.

Blood contamination

SF can be quite easily obtained from knee joints by aseptic aspiration, avoiding blood contamination, but this can still occur in up to 19% of cases. This leads to the addition of proteins to the SF proteome and can result in the discovery of false biomarker candidates. Analytically, suppression of ion signals can occur—an effect that can be minimized if the samples are centrifuged to remove red blood cells before freezing. Another suggestion, which may not always be feasible due to sample availability, includes the elimination of samples with visible blood contamination; unfortunately, significant blood protein additions to the SF proteome might be present irrespective of its visual appearance.

Biological and external factors

In addition to factors related to sample handling, biological and external factors should also be carefully regulated as they can affect the SF proteome. For instance, intake of food can alter protein content of plasma and, as previously mentioned, these effects are also translated to the SF through the passive protein exchange at the synovial membrane. Furthermore, several critical analytes, such as IL-6, exhibit a circadian rhythm. Therefore, sample collection should ideally occur in a fasting state.

A study by Shizu et al. demonstrated the effects of cigarette smoke derivatives on SF proteins. Expression levels of IL-1α, IL-1β, IL-6, IL-8 and CYP1A1, specifically, were...
found to be elevated in SF in response to cigarette smoke in RA patients. Heavy cigarette smoking should perhaps be used as an exclusion criterion when selecting SF donor patients.

Furthermore, sex, age and grade matching – common practices in biomarker discovery pipelines – are also crucial. For example, Cutolo et al. reviewed the effects of gender on synovial fluid composition and found that immune reactivity is greater in females than in males, possibly due to higher antigen-presenting activity and mitogenic responses. The conversion of dehydroepiandrosterone in macrophages to downstream effector hormones (including estrogens) is an important factor for local immunomodulation in RA.

A recent study by Rubenhein et al. demonstrated that age and OA grade also have effects on SF protein content, as IL-7 levels in SF are elevated in elderly patients but are depressed in patients with severe OA.

SF sample fractionation strategies

The dynamic range in protein concentration in SF extends from mg/mL for abundant proteins such as albumin and immunoglobulins to pg/mL for signaling molecules such as TNF-α. This is similar to plasma, where the dynamic range is about 12 orders of magnitude. The proteomic analytical methods currently available have a concentration range within 10^-2–10^5 orders of magnitude, leading to low detection of less abundant proteins. A common way to overcome this problem and reduce protein complexity is by implementing pre-fractionation methods and/or depleting the sample of the most abundant proteins using depletion columns.

Although all commercially available columns have been designed for serum and plasma, they are also valuable for SF because most abundant SF proteins correspond to those in serum/plasma. We should remark that, although high-abundance proteins can physically mask less abundant proteins with similar isoelectric points (pI) and molecular weight coordinates, depletion columns may also bind proteins in a nonspecific manner. Whether bound directly to a column or indirectly through secondary binding to immunoglobulins or albumin, depletion should only be used when whole sample integrity is not essential. According to a study by Chen et al., treatment of SF prior to fractionation will decrease reproducibility while increasing protein loss; therefore, it should also be avoided.

Instead, low molecular weight components of SF, which could be representative of putative biomarkers, can be enriched by size exclusion, strong cation exchange (SCX) or reverse phase (RP) chromatography. The term multidimensional protein identification technology (MudPIT) was introduced to describe the use of multidimensional liquid chromatography prior to MS analysis to identify proteins.

At the protein level, SDS-PAGE is the most common separation technique and is followed by gel cutting, protein digestion, online reverse-phase nano-LC separation and MS analysis of the resulting peptides from each gel band. With an increasing number of diverse separation methods that can be combined, improved proteome coverage is expected. This is at the expense of increased analysis time and decreased reproducibility, which in turn results in decreased detection of small changes in the proteome (when comparing physiological and pathological conditions). Therefore, careful consideration must be taken when deciding on pre-fractionation methods.

Proteomic biomarker discovery methods

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)

2D-PAGE involves the separation of proteins based on pI in the first dimension and on size in the second dimension. Following separation, the gels are generally stained with Coomassie Brilliant Blue, SyproRuby or silver to visualize the proteins as spots on the gel. The area and intensity of the spots indicate the relative abundance of the same protein across different gels (and therefore different samples). As a result of low-inter-gel reproducibility, proteins from different samples are labeled with different fluorescent dyes and are run on the same 2D-gel in a technique deemed difference gel electrophoresis (DIGE). This allows the simultaneous comparison of individual proteins in diverse samples using different laser wavelengths for differential detection. The 2D-PAGE analysis provides several types of information about the proteins investigated simultaneously, including molecular weight, pI and quantity, as well as possible posttranslational modifications. This method falls short in its reproducibility, inability to detect low-abundant and hydrophobic proteins and low sensitivity for proteins with pH values that are either too low (pH<3) or too high (pH>10) and molecular masses that are too small (<10 kD) or too large (>150 kD). Because of this, 2D-PAGE is extensively used, but is reserved predominantly for qualitative experiments. In complex proximal fluids like SF, there is a high degree of co-migration of proteins, leading to errors in both the quantification and identification of the differentially expressed proteins. The DIGE technique has dramatically improved the reproducibility, sensitivity and accuracy of quantitation. However, its labeling chemistry has various limitations: proteins lacking lysine cannot be labeled; fluorophores are very expensive and require special equipment for visualization. Furthermore, matching protein spots between DIGE gels and the corresponding preparative gels (used for protein identification) is a tedious and challenging task, and care needs to be taken in this process to minimize false discoveries. Irrespective of these disadvantages, 2D-PAGE is still a popular method for biomarker discovery in SF, and has led to the discovery of a number of disease-specific proteins in OA and RA SF.

MS-based techniques

Several MS-based approaches are central for biomarker discovery in various bodily fluids including proteome profiling, chemical labeling of proteins/peptides and label-free MS/MS quantification. Many groups have used surface-enhanced laser desorption/ionization (SELDI) as a fast and fairly reproducible method to analyze low molecular weight endogenous polypeptides in serum/plasma – a method that was later utilized in SF as well. Matrix-assisted laser desorption/ionization (MALDI) has also been utilized for
protein screening of SF\textsuperscript{82}, where only 50–150 protein peaks (out of a possible \(\sim 500\)) were detected in the 1–9 kDa mass range, thus overlooking many potential biomarkers. Using these techniques is also problematic due to the difficulty in carrying out protein identification of the putative markers, especially those in the higher end of the mass range. Additionally, there is also a dramatic decrease in sensitivity when detecting proteins of higher molecular mass, and the resolution of the instruments is limited, leading to overlapping peaks when analyzing complex samples such as SF\textsuperscript{76,83}. Analysis of the synovial fluid proteins using an Orbitrap mass spectrometer with online reverse phase nano-LC separation has not yet been performed. Such analysis would greatly increase the number of peptides and proteins identified (based on our unpublished data), as has also been previously demonstrated for CSF\textsuperscript{84}. In our opinion, MALDI and SELDI analysis provides only a limited overview of the SF proteome compared to what can be achieved using the Orbitrap approach.

Quantification of protein levels to achieve accurate differential protein profiling between samples has been a major challenge in proteomics\textsuperscript{85}. For LC–MS quantification, differential chemical labeling of the samples prior to analysis is usually required. An alternate protocol does not involve labeling and is termed label-free quantification (LFQ). The label-free, semi-quantitative approach is based on the peak intensity of the peptides in the MS scan\textsuperscript{76,86}, or on the number of observed spectra per peptide across different samples\textsuperscript{76,87}. In general, the advantage of label-free approaches over chemical labeling lies in the low cost and the high number of samples that can be easily included in the experiment. Potential disadvantages of label-free techniques include the lower reproducibility of results, which compromises detection of smaller quantitative changes between samples\textsuperscript{76,88}. The reproducibility of LFQ experiments is largely based on the timeframe of the experiment; therefore, concrete standardization of the entire pipeline (sample processing, instrument setup/calibration) is crucial. In terms of data analysis, spectral counting is the simplest label-free method but is also the least reliable because quantification accuracy drops when the number of spectra representing a protein becomes very low (i.e. two or less)\textsuperscript{88}. As a result, classifying smaller changes among the identified low abundance proteins, which typically have low numbers of observed spectra, will be difficult with this method. Alternatively, intensity-based quantification is, in theory, more capable of obtaining accurate values for lower abundance proteins\textsuperscript{88}. Computer algorithms such as MaxQuant\textsuperscript{89} require the identification of only one peptide in at least one of the samples being analyzed to extract the peak intensity information and to quantify the peptide in all the analyzed samples. Other algorithms such as PEPPeR\textsuperscript{90,91} do not require the identification of the peptide to report the intensity ratio of corresponding peptides/peaks across the analyzed samples. This approach allows for the quantification of low-abundance peptides/proteins that are not among the commonly chosen peptides for fragmentation during data-dependent acquisition. It also provides the option of identifying modified peptides that were otherwise unknown and not selected and, therefore, not reported following database searches. In such a case, the data analysis is more difficult, and software capable of this is not yet publicly available\textsuperscript{91,92}.

As an alternative to label-free approaches, chemical labeling of peptides/proteins prior to fractionation and MS analysis has also been used. There are a number of available labeling strategies: tandem mass tag (TMT)\textsuperscript{93}, iTRAQ\textsuperscript{94} and ICAT\textsuperscript{95} are the most popular commercial alternatives, while dimethylation\textsuperscript{96} and \(^{18}\text{O}/^{16}\text{O}\)\textsuperscript{97} represent non-commercial alternatives. All methods have particular advantages and disadvantages. In the case of iTRAQ (4-plex and 8-plex) and TMT (6-plex), free amines generated from trypsin digestion and present in all the peptides are labeled and, therefore, theoretically no information is lost\textsuperscript{95}. Reporter ions relating to the various samples being analyzed are released from the peptides during MS/MS fragmentation and are thus used to represent the originating sample from which specific proteins originate; no quantitative information is obtained from the MS scan\textsuperscript{93,94}. Utilizing these methods in complex samples such as SF may result in the partial suppression of the quantitative reporter ion signal\textsuperscript{98}. This is believed to occur as a result of precursors with similar mass to the precursor ion selected for MS/MS also entering the mass spectrometer during fragmentation (co-fragmentation). These precursors release their specific reporter ions, which then mix with those of the original precursor. The reporter ions from the co-fragmenting peptides will compress the ratios for the originally selected precursor, assuming that most peptides have similar levels across the samples\textsuperscript{98}. This suppression effect will vary depending on the width of the precursor isolation window. Low-abundance proteins/peptides are more vulnerable to this effect, which is something to consider when analyzing such data. ICAT has the clear limitation of only labeling cysteine residues, which is advantageous as it reduces the complexity of the sample\textsuperscript{95}. At the same time, proteins that do not have peptides containing cysteine, or only have cysteine-containing peptides that are unsuitable for MS (too large or too small), will not be quantified. In addition, the quantitative information from each protein with this method is sparse, and ICAT only appears as duplex labels\textsuperscript{98}. In our opinion, these labeling protocols are rather lengthy and involve many steps, resulting in compromised reproducibility.

In the case of dimethylation and \(^{18}\text{O}/^{16}\text{O}\) approaches, the quantification is done on the MS spectra, avoiding the suppression effect when using MS/MS-based reporter ion-dependent methods\textsuperscript{97}. However, sample complexity will be theoretically increased when the differences appear in the MS spectra, resulting in less unique peptide and protein identifications. The dimethylation and \(^{18}\text{O}/^{16}\text{O}\) methods label at the peptide level and appear to be cost-effective and good alternative strategies, labeling all tryptic peptides in the sample. However, due to the small mass difference between labeled and unlabeled peptide species, these approaches have usually been restricted to high-resolution mass spectrometers\textsuperscript{97}.

**Biomarker candidate qualification and verification**

During qualification and verification phases of candidate biomarkers, the unbiased experimental methods used during discovery are replaced by target-driven quantitative methods.
Until recently, the quantification and verification of biomarker candidates has mainly relied on the availability of antibodies to perform Western blotting, immunohistochemistry or ELISA, which has limited the verification of biomarkers lacking commercially available antibodies. If the development of high-quality immunoanalytical assays was fast, straightforward and inexpensive, the barriers to verification would be greatly reduced. Furthermore, these antibody-based methods are problematic when attempting to efficiently measure tens or hundreds of biomarker candidates arising from typical high-throughput discovery experiments. Recently, there has been a paradigm shift toward the use of MS-based proteomics in the qualification and verification steps by targeting and measuring selected signature peptides from the discovered biomarker candidates. This method is called selected reaction monitoring and it exploits the capabilities of triple quadrupole (QqQ) or Q-Trap instruments. For reliable quantification of a protein of interest, proteotypic peptides or peptides unique to a particular protein of interest are first selected. The corresponding predefined precursor masses of these peptides are selected in the first quadrupole and fragmented in the second quadrupole, with predefined fragmentation masses selected in the third quadrupole. This unique pair of precursor mass and fragmentation mass is termed a transition. The SRM method can be applied simultaneously to multiple proteins (MRM), a protocol that is further reviewed by Lange et al. Stable Isotope Dilution-SRM (SID-SRM) is based on the selection of three to five peptides resulting from tryptic digestion from each protein to be quantified. Synthetic peptides containing heavy lysine and arginine residues (which have incorporated 13C/15N atoms) are then added to all samples. These peptides serve as internal standards providing relative quantitative ratios for each proteotypic peptide corresponding to each protein of interest between all samples. The nature of this approach allows for very high-molecular selectivity, and if interference is present it can also be detected.

Compared to ELISAs that have been most often employed in candidate verification, the SID-SRM assay development timeline is short, the cost is low and the method can be highly multiplexed, monitoring 200 or more transitions in a single experiment. The median inter-laboratory coefficient of variation (CV) of quantification using SID-SRM for unfractionated plasma was found to be \( \leq 20\% \), with a limit of quantification in the low \( \mu \text{g/mL} \) range. In combination with high-abundance protein removal and SCX fractionation, limit of quantification was typically in the low ng/mL range in plasma. As far as we know, there are only two similar published studies on SF, and we believe that further optimization of this technique will highly impact protein biomarker qualification and verification in SF in the future. When it comes to validation of biomarker candidates after qualification and verification, ELISAs remain the gold standard as they can be conducted on numerous samples concurrently with low CVs and with very good sensitivity.

### Biomarker discovery for rheumatic diseases

Until now, different proteomic approaches have been utilized for the characterization of the human SF proteome. Several groups have combined 2D-PAGE with MALDI-TOF-MS and LC-ESI-MS; however, few have included any qualification and verification work, most likely due to the unavailability of ELISA assays or antibodies for Western blot analysis for the biomarkers of interest. New technologies such as SID-SRM have emerged over the last few years, making it possible to quantify, in a multiplexed fashion, many proteins with relatively high throughput. This holds great promise for the qualification/verification and validation of a larger portion of the biomarker candidates that have arisen from the many discovery experiments conducted in the last few years. In this section, we summarize the biomarker discovery efforts in SF for RA, OA and JIA using proteomic methods (Table 2).

### Proteomics biomarker discovery in RA

Rheumatoid arthritis is characterized by synovial inflammation and hyperplasia, autoantibody production (rheumatoid factor and anti-citrullinated peptide [CCP] antibody [ACPA]), cartilage and bone destruction, as well as systemic features...

### Table 2. Protein markers in joint diseases identified in SF using various proteomics methods.

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Methods</th>
<th>Differentially expressed proteins</th>
<th>Validated proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>SELDI</td>
<td>S100A8, CENP-E</td>
<td>S100A8</td>
</tr>
<tr>
<td></td>
<td>2-DE, MALDI-TOF</td>
<td>SAA, S100A9, S100A12, FGB</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>2-DE, MALDI-TOF</td>
<td>S100A9</td>
<td>S100A8, S110A9</td>
</tr>
<tr>
<td></td>
<td>MALDI-TOF</td>
<td>NGAL, TERA, CTSD, TG2</td>
<td>NGAL, TERA</td>
</tr>
<tr>
<td></td>
<td>DIGE, MALDI-Q-TOF; Q-TOF</td>
<td>HBP, C4B, APOB, PZP, DEF1, S100A8</td>
<td>S100A8</td>
</tr>
<tr>
<td></td>
<td>2-DE, NanoLC-MALDI-TOF/TOF</td>
<td>APOE, MMP3, FCN3, C8, PRG4, BGH3, MMP1</td>
<td></td>
</tr>
<tr>
<td>OA</td>
<td>2-DE, MALDI-TOF</td>
<td>SAA, S100A9, S100A12, FGB</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>DIGE</td>
<td>HP</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>2-DE, MALDI-TOF</td>
<td>S100A9</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>SDS-PAGE, LC-MS/MS</td>
<td>ALB, A2M, APOE, APOH, CP, HP, ORM1, C4B, PRG4, VDBP</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>UF, LC-MS/MS</td>
<td>COLI, CSPG4, SAA, VIM, TUB, MGP</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>2-DE, NanoLC-MALDI-TOF/TOF</td>
<td>COMP, ADAMTS12, A2AP, PEDF, HPG, CH3L1, FINC, GELS</td>
<td>FINC, GELS</td>
</tr>
<tr>
<td>JIA</td>
<td>DIGE, MALDI-TOF/TOF</td>
<td>C3c, APOAI/II, VDBP, ALB, TF, HG, APP, C3, C9, ITIH4, APOAI, HP, TF</td>
<td>C3c, APOAI, VDBP</td>
</tr>
<tr>
<td></td>
<td>DIGE, MALDI-TOF/TOF, Q-TOF</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
including cardiovascular, pulmonary, and musculoskeletal disorders (Table 1)\(^{109}\). The disease course in RA can range from mild and self-limiting to severe and progressive; therefore, effective treatment of RA has been hindered by the heterogeneity of the disease. “Early” diagnosis is made once erosion of cartilage and bone has already begun—a time at which the optimal treatment window may have already passed\(^{10,109}\); therefore, diagnostic biomarkers of RA are needed for patients presenting with undifferentiated arthritis.

The most significant progress in the diagnosis of RA over the last decade has been the development of assays for the detection of autoantibodies against cyclic citrullinated peptides\(^{110}\), which, unlike the traditional RA biomarker rheumatoid factor, are highly specific to RA\(^{110,111}\). However, the diagnostic sensitivity of ACPA positivity in cohorts of early synovitis is between 40% and 71%\(^{112,113}\) – partly because approximately 30% of RA patients never actually develop these antibodies\(^{114}\). Thus, the search for biomarkers that provide greater sensitivity and specificity in the diagnosis of early RA is ongoing.

Differential analysis of 2-DE protein patterns of SF from OA and RA patients has enabled the identification of markers specifically related to RA rather than OA such as S100A9 and SAA\(^{51,115}\). Analysis of SF from RA patients by 2-DE and MALDI identified myeloid-related proteins (S100A9, S100A8) in RA SF when compared to OA SF\(^{51,116}\), while a similar comparison of synovial tissue by 2-DE also revealed increased levels of MRPS (S100A8) in RA samples\(^{115}\). Furthermore, using quantitative proteomics, Liao et al. demonstrated a correlation between severity of joint erosion in RA and the levels of S100 proteins A8, A9 and A12 in SF\(^{104}\). They adopted a two-dimensional liquid chromatographic approach (LC-MS/MS) to generate protein profiles from erosive and nonerosive RA SF. Thirty proteins were selected due to their upregulation in erosive RA, including C-reactive protein (CRP), and were quantified in the sera of patients using MRMs. Once again, only the S100 proteins were significantly elevated in erosive versus nonerosive RA patients\(^{104}\). A different approach was used by Katano et al., where proteins derived from cytokine-stimulated neutrophils were analyzed by MALDI-TOF to identify cytokine-regulated genes\(^{117}\). NGAL protein, their most promising candidate, was then measured in the SF of OA and RA patients, where it was found to be significantly upregulated in the RA SF\(^{117}\).

Liquid chromatography-based approaches were most recently utilized to study SF and serum from RA and OA patients and the results revealed a high number of putative RA biomarkers. Various prognostic RA biomarkers were identified in SF, and were then validated in serum\(^ {106}\) (Table 2).

**Proteomics biomarker discovery in OA**

OA, the most frequent arthropathy, is associated with aging and is characterized by progressive degradation of the articular cartilage. It affects more than 10% of the population\(^ {118}\) and is the leading cause of permanent work incapacitation, as well as one of the most common reasons for visiting primary care physicians. A major objective for OA research is the development of early diagnostic strategies, because OA is clinically silent in its initial stages and, by the time of diagnosis, damage is already present. The current diagnostic method of OA relies on the description of pain and stiffness in the affected joints, and radiography is used as the reference technique in defining the grade of joint destruction\(^ {118,119}\).

New strategies for OA biomarker discovery and validation have emerged including genomic, proteomic and metabolomics methodologies. Many proteomic studies performed on SF have focused on RA and use samples of OA SF as controls\(^ {56,51,106,115,116}\). Two distinct proteomic approaches have been developed to gain insight into the OA SF proteome. In a study performed by Gobezie et al., researchers utilized SDS-PAGE and LC-MS/MS to map the SF proteome of healthy, early OA and late OA patient cohorts\(^ {108}\). From these groups they identified 135 SF proteins, 18 of which were altered in OA. Another group studied SF endogenous peptides using ultrafiltration and LC-MS/MS analysis\(^ {117}\) and noted six proteins which may serve as potential markers for OA: COL2, PRG4, SAA, TUB, VIME and MGP. Finally, the use of SELDI-MS led to the identification of several discriminatory biomarker candidates between RA and OA, one of which was MRPS-8 (S100A8)\(^ {119}\). More recently, Mateos et al. reported the identification of 136 SF proteins\(^ {106}\). In this data set, SF proteins from RA and OA were identified and quantified relative to each other to identify differentially expressed proteins between the two groups\(^ {106}\). Evidently, proteomic tools have already had a huge influence on biomarker discovery, as they have already aided in the identification of a number of molecules that might be related to arthritis. Some of these, including COMP, COL2 or MMPs, were previously detected in other studies, whereas others have been newly characterized only in proteomic analyses and need to be subjected to further qualification assays\(^ {107,120}\).

**Proteomics biomarker discovery in JIA**

Juvenile idiopathic arthritis (JIA) is a heterogeneous group of inflammatory diseases with varying sex distribution, genetic predisposition, clinical manifestations, disease course and prognosis. At present, there are no clinically useful prognostic markers to predict disease outcome in these patients. There are three main JIA subtypes: oligoarticular, the most frequent subtype, polyarticular, the more chronic subtype, and systemic, the severe subtype also associated with various extraarticular manifestations\(^ {105}\). Approximately 25% of children develop extended oligoarticular disease, which is much more resistant to therapies and harder to treat\(^ {121}\). Prognostic biomarkers are, therefore, essential for determining the risk of inflammation spreading to unaffected joints and helping to initiate the appropriate therapies.

Proteomic strategies, as previously discussed, can be used to identify and quantify proteins associated with a particular disease subset. Using 2-DE, MALDI-TOF and Q-TOF for protein identification, Rosenkranz et al. identified a subset of the synovial proteome, which could distinguish between oligoarticular, polyarticular and systemic forms of JIA. In this case, haptoglobin emerged as a particularly strong candidate biomarker\(^ {105}\). Ling et al. also identified a panel of seven plasma proteins using 2-DE DIGE, which can discriminate patients at risk of a disease flare with greater reliability than
CRP. Using similar methodologies, Gibson et al. performed proteomic characterization of SF from oligoarticular, extended oligoarticular and polyarticular patients. They identified specific clusters of proteins that differentiate between subtypes of JIA — more specifically, a truncated isoform of vitamin D-binding protein (VDBP) was present at significantly reduced levels in the SF of extended oligoarticular patients relative to other subgroups.

**Conclusion and future perspectives**

There are several factors we consider to be important to increase the chance that an SF-based proteomics biomarker project will be successful. Some of the key points are to include well-characterized, high-quality samples with documented and preferably standardized sample collection methods and handling history and with high-quality associated clinical information. Samples should be properly matched with regard to parameters like age, lifestyle, medication, time of day of SF collection and disease state.

Many technologies have been used in the field of SF proteomics including 2-DE and several MS-based methods. Quantitative methods such as spectral counting, iTRAQ, as well as MRMs have also been used to identify differences in SF proteins between different pathological states. The data generated from these various experiments consist of heterogeneous measurements; therefore, comparison across disease states is difficult. The ideal experiment should be performed using the same pre-analytical sample processing and fractionation techniques, and should be run on the same instrument using SF from multiple disease states. This will yield high confidence and extensive characterization of the different proteins expressed across various rheumatologic conditions.

It is obvious that the field of proteomics has advanced our understanding of diseases such as RA and OA, but in the field of biomarker research the current strategy most frequently employed is still transcriptomic analysis using microarrays, which allows the identification of candidate genes involved in the pathophysiology of the disease. Gene expression levels, however, do not always predict protein levels due to alternative transcriptional and translational regulatory steps, and the activity of protein degradation processes. The foremost advantage of proteomics is that the actual functional molecules of a cell are being studied, elucidating a reliable picture of what is occurring in the tissue. As such, proteomics complements genomics-based approaches by bridging the gap between what is encoded in the genome and what is occurring at the tissue level. It is well known that genomic and proteomic data sets have different sources of bias and variance, so combining them may lead to a more precise view of differential protein abundance. The key benefit of the integration of proteomic and transcriptomic data in the field of biomarker discovery is its potential for improving the selection of candidates to validate. If both transcriptomic and proteomic platforms agree on a strong differential expression between the groups of patients to be distinguished, the attractiveness of a candidate strengthens.

Furthermore, considering that one of the major clinical challenges in arthritis is the development of robust biomarkers for predictors of outcome and disease progression, clinicians and scientists need to work in tandem. Clinicians must ensure the rigorous categorization of patients’ disease according to internationally recognized criteria, and they must verify that samples are obtained and stored according to well-defined protocols. Scientists must optimize their methodologies to ensure that their techniques are sensitive and that their results are reproducible and relevant to the clinical questions.

**Declaration of interest**

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

**References**

62. D. Cretu et al.


