A Normative Study of the Synovial Fluid Proteome from Healthy Porcine Knee Joints

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.


Published Version: doi:10.1021/pr500587x

Citable link: http://nrs.harvard.edu/urn-3:HUL.InstRepos:21462451

Terms of Use: This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
A Normative Study of the Synovial Fluid Proteome from Healthy Porcine Knee Joints

Tue Bennike, †, † Ugur Ayturk, ‡, # Carla M. Haslauer, ‡ John W. Froehlich, § Benedikt L. Proffen, ‡ Omar Barnaby, † Svend Birkeland, † Martha M. Murray, ‡ Matthew L. Warman, ‡, ¶, †† Allan Stensballe, † and Hanno Steen∥, †††

†Department of Pathology and Proteomics Center, ‡Department of Orthopaedic Surgery, §Department of Urology, and ††Howard Hughes Medical Institute, Boston Children’s Hospital, Boston, Massachusetts 02115, United States
¶Department of Health Science and Technology, Aalborg University, Aalborg DK-9220, Denmark
∥Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, United States

ABSTRACT: Synovial fluid in an articulating joint contains proteins derived from the blood plasma and proteins that are produced by cells within the joint tissues, such as synovium, cartilage, ligament, and meniscus. The proteome composition of healthy synovial fluid and the cellular origins of many synovial fluid components are not fully understood. Here, we present a normative proteomics study using porcine synovial fluid. Using our optimized method, we identified 267 proteins with high confidence in healthy synovial fluid. We also evaluated mRNA expression data from tissues that can contribute to the synovial fluid proteome, including synovium, cartilage, blood, and liver, to better estimate the relative contributions from these sources to specific synovial fluid components. We identified 113 proteins in healthy synovial fluid that appear to be primarily derived from plasma transudates, 37 proteins primarily derived from synovium, and 11 proteins primarily derived from cartilage. Finally, we compared the identified synovial fluid proteome to the proteome of human plasma, and we found that the two body fluids share many similarities, underlining the detected plasma derived nature of many synovial fluid components. Knowing the synovial fluid proteome of a healthy joint will help to identify mechanisms that cause joint disease and pathways involved in disease progression.

KEYWORDS: Synovial fluid, synovium, plasma, porcine, human, proteomics, transcriptomics, origin, PTM, method optimization

INTRODUCTION

Synovial fluid is present in all joint cavities, where it protects the articular cartilage surfaces, in part by reducing friction. Synovial fluid, furthermore, facilitates the transport of nutrients and waste products including proteins and metabolites between the vascularized synovium and the avascular cartilage.1–3 Many components of synovial fluid are derived from blood plasma, and these two body fluids share many similarities in terms of protein composition.4,5 However, synovial fluid also contains proteins secreted from the surrounding tissue, including the articular cartilage and synovium.6 The protein concentration in synovial fluid from healthy knee joints is approximately 25 mg/mL, i.e., ∼1/5 of the concentration found in blood plasma, and albumin constitutes approximately 12 mg/mL.1,3,7–10 Joint diseases, in particular osteoarthritis (OA) and rheumatoid arthritis (RA), are the leading cause of disability in people over 55 years.6 It has been estimated that, as of 2005, 27 million adults in the United States have clinical OA, and in 2009, OA was the fourth most common cause of hospitalization.11,12 Furthermore, joint injuries, such as anterior cruciate tear, that predispose to precocious joint failure have become epidemic in young athletes.13,14 Although changes in the composition of synovial fluid have been described in patients with joint disease, there do not exist reliable biomarkers for early disease diagnosis or biomarkers that accurately depict response to therapy.1,15–17 As a consequence, OA is often not diagnosed before irreversible damage has occurred.13,18,19 Since synovial fluid is in direct contact with the joint tissues, it provides an attractive source of biomarkers candidates for monitoring joint health and for furthering the understanding of the disease mechanisms.3,18 Previous studies of synovial fluid have mainly been focused on various diseased states. However, the protein concentration, content, and volume of synovial fluid is known to change dramatically during active joint diseases, and few studies have been focused on synovial fluid in healthy...
state or the likely origins of the synovial fluid proteins.\textsuperscript{3,17} Balakrishnan et al.\textsuperscript{20} compared the human synovial fluid proteome from OA and RA patients. Synovial fluid was immune-depleted of the most abundant proteins, followed by liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis, which led to the identification of 575 proteins, of which 135 demonstrated a greater than 3-fold abundance change between the two groups. In a similar study, Mateos et al.\textsuperscript{25} demonstrated a greater than 3-fold abundance of plasma by the liver or by blood cells) and articular cartilage.

 Furthermore, a human synovial fluid sample was obtained from a RA patient according to an approved IRB protocol (IRB-P00006443) to evaluate the integrity of the UniProt Sus scrofa protein database.

Euthanasia of the animals was induced by intramuscular injection of atropine (0.04 mg/kg), Telazol (4.4 mg/kg), and xylazine (2.2 mg/kg) and finalized by intravenous injection of Fatal Plus (86 mg/kg). At the time of euthanasia, synovia from the knee joints of the hind limbs were harvested. Care was taken to sample only the synovial membrane without any subintimal structures, such as fat or blood vessels. Each tissue specimen was snap frozen in liquid nitrogen and stored at −80 °C.

### Protein Concentration

Total protein concentration for each sample (diluted 1:30 in water) was determined for normalization of sample material using a colorimetric (Bradford) protein assay kit (Bio-Rad, Hercules, CA) according to the manufacturer’s instructions, with bovine serum albumin used as the standard.

### SDS-PAGE

Thirty micrograms of total synovial fluid protein was prepared for sodium dodecyl sulfate (SDS)-PAGE in Laemmli sample buffer (Bio-Rad, Hercules, CA) according to the manufacturer’s instructions. SeeBlue Plus2 pre-stained standard (Invitrogen, Carlsbad, CA) was used as the protein molecular weight standard. The sample was fractionated using NuPAGE 4–12% Bis-Tris minigels (Invitrogen) at 150 V for 65 min in MOPS SDS-running buffer (Invitrogen). The gel was stained using Coomassie blue, SimplyBlue SafeStain (Invitrogen), according to manufacturer’s instructions.

### Synovial Fluid Protein Digestion

Three trypsin digest protocols were evaluated:

1. **Filter-Aided Sample Preparation (FASP) Digestion.** Performed using the FASP protein digestion kit (Protein Discovery, San Diego, CA) according to manufacturer’s instructions using 30 kDa cuto spin filters. Ninety micrograms of total synovial fluid protein was digested overnight at 37 °C with 2 μg of sequencing grade modified trypsin (Promega, Fitchburg, MA). To assess the need of gelatin removal when working with synovial fluid, 500 U peptide-N4-(N-acetyl-beta-glucosaminyl)-asparagine amidase (PNGase F) (New England BioLabs, Ipswich, MA) was added to these samples prior to the trypsin digestion step, and samples were incubated overnight at 37 °C, after which the normal FASP protocol was continued. After trypsin digestion, the samples were desalted with TARGA C18 columns (Nest Group, Southborough, MA) and resuspended in 5% acetonitrile (ACN) and 5% formic acid (FA) prior to analysis.

2. **Urea In-Solution Digestion.** Performed according to Gallien et al.\textsuperscript{31} Ninety micrograms of total synovial fluid protein was diluted with 8 M urea in 100 mM ammonium bicarbonate to a final volume of 25 μL. The sample was reduced with dithiothreitol at a final concentration of 12 mM for 30 min at 37 °C and alkylated with iodoacetamide at a final concentration of 40 mM for 1 h at room temperature in the dark. The samples were diluted with 100 mM ammonium bicarbonate to a total volume of 100 μL, 2 μg of trypsin was added, and the sample was digested overnight at 37 °C. The samples were desalted

---

**Experimental Procedures**

### Collection of Synovium and Synovial Fluid Samples

Six adolescent Yucatan minipigs (Coyote CCI, Douglas, MA), aged 12–15 months, were obtained for use in this study. All minipigs were housed and monitored by the Animal Resources at Boston Children’s Hospital (ARCH) and handled according to approved Institutional Animal Care and Use Committee (IACUC) protocols. Minipigs were acclimated to the ARCH environment for a minimum of 3 days prior to experimental handling.

Nonbloody synovial fluid was obtained from the joints by aspiration using a 21 gauge needle. Samples were centrifuged at 3000g at room temperature for 10 min to pellet and remove cells and cellular debris. In some cases, 3 mL of sterile saline was injected into the knee joint to facilitate fluid extraction; after saline injection the knee was bent 10 times to ensure homogeneous fluid distribution and mixing. The saline/
with TARGA C18 columns (Nest Group) and resuspended in 5% ACN, 5% FA prior to analysis.

(3). In-Gel Digestion. Three gel-lanes, each loaded with 150 μg total synovial fluid protein, were divided into 10 sections each and subjected to standard in-gel tryptic digestion as previously described, followed by analysis.

Human Plasma Protein Digestion

Human plasma was acquired as part of an ongoing method optimization study using a deidentified, discarded cord plasma sample and thus are not considered research of human subjects. One hundred micrograms of plasma protein was digested using the FASP protein digestion kit (Protein Discovery, San Diego, CA) with a modification to the recommended protocol. This modification involved the use of a 10 kDa MWCO filter instead of the stock 30 MWCO filters. A trypsin/LysC mix (Promega, Madison, WI) was added to the FASP filter at 1:25 ratio for digestion. The samples were then incubated at 37 °C overnight, and the resulting peptides were recovered as recommended by the manufacturer’s protocol. The peptides were desalted using Oasis HLB columns (Waters, Milford, MA) and resuspended in 2% ACN in 0.1% FA prior to analysis.

LC–MS/MS Measurement and Proteomics Data Analysis

Two different high-resolution/high-accuracy mass spectrometer systems were used for the shotgun proteomic analysis: (1) For post-translational modification (PTM) analysis and method optimization, synovial fluid samples were analyzed on a TripleTOF 6600 (AB Sciex, Framingham, MA) connected online with a nanoflow UPLC and a NanoFlex System (Eksigent/AB Sciex). The samples were loaded onto a 15 cm reversed-phase C18 200 μm chip with 2 μL/min in 100% solvent A (0.1% FA). The samples were then separated using a 15 cm reversed-phase C18 75 μm chip and eluted with a linear gradient of 2% solvent B (0.1% FA in ACN), which was raised to 35% solvent B over 120 min (60 min for in-gel digested samples) at a constant flow rate of 500 nL/min. (2) The six FASP digested synovial fluid samples used to determine the synovial fluid protein list and abundances as well as the trypsin-digested human plasma samples were analyzed on a Q Exactive (Thermo Scientific, Waltham, MA) connected online to an EASY-nUPLC 1000 (Thermo Scientific). The samples were loaded onto a 10.5 cm reversed-phase C18 PicoChip with a flow rate of approximately 1 μL/min in 98% solvent A and 2% solvent B and were eluted with eluent B using a linear gradient that was raised to 35% over 120 min at a constant flow rate of 300 nL/min.

The AB Sciex.wiff data files were analyzed using ProteinPilot 4.5 (rev. 1656, Paragon Algorithm 4.5.0.0). To identify the most commonly single observed PTMs, data files were searched in thorough mode with a focus on biological modifications in ProteinPilot to include more than 300 different PTMs.

The.raw data files from synovial fluid and human plasma analyzed on the Q Exactive were searched using MaxQuant 1.4.1.2. All standard settings were employed with carboxymethyl (C) as a static modification and deamidation (NQR), oxidation (M), and protein N-terminal acetylation included as variable modifications. Label-free quantitation of all proteins was performed in MaxQuant based on integrated precursor intensities. Protein abundances are represented as protein intensity-based absolute quantitation values (iBAQ) and are reported for all proteins having at least two quantifiable unique peptides in at least three LC–MS runs.

The human plasma sample was searched against the UniProt Homo sapiens reference proteome database with isoforms (downloaded 7/18/2014, containing 89,032 entries). The porcine synovial fluid data was searched against the UniProt S. scrofa reference proteome database (downloaded 11/09/2013, containing 26,070 entries). The human RA synovial fluid, used to evaluate the UniProt S. scrofa database, was searched against all reviewed H. sapiens UniProt proteins (downloaded 08/10/2013, containing 20,277 entries). All proteins and peptides are reported below a 1% false discovery rate (FDR) cutoff, and protein posterior error probability (PEP, equivalent to expectancy) was investigated to ensure only confident protein identifications. For the PTM analysis, the search results were analyzed using ProteinPilot Descriptive Statistics Template, version 3.001, and for the protein abundance analysis, the iBAQ values were analyzed using Perseus, version 1.4.1.3, and IBM SPSS Statistics (version 21). Venn diagrams were created with BioVenn and Venny.

Assignment of Formerly Glycosylated Asparagine Residues

Four criteria were required to assign N-glycosylation sites: (I) a 1% FDR cutoff to all peptide spectral matches (PSMs); (II) all site assignments required the presence of a consensus site (CS) for N-glycosylation, i.e., NX(S/T), where X may be any amino acid except proline; (III) once CS status was established for all peptide assignments, an asparagine deamidation at the asparagine within the CS was required; and (IV), finally, all true site assignments were required to come from sample preparations that were treated with PNGase F. The FDR of site assignment was estimated by evaluation of the random rate of site assignment among control samples that were not treated with PNGase F. In this way, the rate of PSMs leading to the identification of a deglycosylated peptide may be compared.

RNA Extraction

Total RNA was extracted from frozen synovium tissue using the PureLink RNA mini kit (Ambion, Austin, TX), treated with PureLink DNase I (Life Technologies, Grand Island, NY) according to the manufacturer’s protocol, and quantified. Briefly, frozen tissue samples were placed in tubes containing 5 metal lysing beads (Metal Bead Lysing Matrix, MP Biomedicals, Solon, OH) and 1 mL of TRIzol (Life Technologies). Tissue was homogenized using a FastPrep24 (Thermo Fisher Scientific, Santa Clara, CA). Each tissue sample was homogenized, and the RNA extraction was repeated two additional times, with samples briefly placed in liquid nitrogen between homogenization runs. Following total RNA extraction and quantification via spectrophotometry using a Nanodrop 1000 (Thermo Scientific, Wilmington, DE), samples were submitted for quality analysis on a Bioanalyzer model 2100 (Agilent Technologies, Santa Clara, CA). Each RNA sample had an RIN > 7.20, indicating that they were of sufficient quality for sequencing.

Library Preparation

A mRNA library for each tissue sample was prepared as previously described using the TruSeq RNA sample preparation kit, v2 (Illumina, San Diego, CA). Briefly, mRNA was enriched from total RNA, chemically fragmented, reverse-transcribed using random hexamers, and ligated to barcoded adapters per the manufacturer’s instructions. The cDNA fragments were amplified via PCR, and the libraries were washed with AMPure XP beads (Beckman Coulter Inc., Danvers, MA) to remove primer dimers. One microliter aliquots from each library were run on a 4%–20% TBE gel (Life
We compared two synovial fluid samples using FASP digestion, urea in-solution digestion, and in-gel digestion.

Additionally, 17% of the proteins unique to the in-gel digestion were keratins, supporting the notion that in-gel digestion protocols can be prone to keratin contamination. Besides this difference, no other bias was found with regard to which groups of proteins were identified uniquely with the different digestion methods. This indicates that the FASP protocol gives more consistent data than the in-gel digestion protocol. It is possible that depletion strategies can increase the number of identified proteins further; however, this was not investigated in this study.

The UniProt S. scrofa reference proteome is not as well annotated as the human proteome. Therefore, in order to determine whether our proteomic method is as sensitive for detecting pig synovial fluid proteins as it would be for human synovial fluid proteins, we performed the same FASP digestion protocol on a synovial fluid sample from a RA patient. In the four technical replicates, we identified, on average, 173 different proteins in the human sample, which seems to be comparable to the 179 different proteins found, on average, in the individual porcine synovial fluid FASP LC–MS runs. The protein composition of synovial fluid is known to be altered during active joint diseases, so the protein overlap was not investigated. Nonetheless, the similar number of identified proteins with the two databases indicates that the UniProt S. scrofa reference proteome is adequate for the analysis of the porcine synovial fluid.

Importance of Accounting for Post-Translational Modification in Protein Identification

Many PTMs occur in vivo or are introduced during sample preparation prior to MS analysis in vitro. Therefore, to identify which PTMs should be considered for the database search of MS data from FASP-digested synovial fluid samples, we analyzed the identified peptides from this method in ProteinPilot Descriptive Statistics Template (Table 1). Several PTMs were identified, and all were likely artifacts from the sample preparation. Chloroacetamide can be used as an alternative alkylation agent to iodoacetamide to reduce the number of alkylation related artifacts. However, due to the low number of detected peptides with these artificial modifications, this was not investigated further.

Synovial fluid is known to be rich in glycoproteins, which prompted us to evaluate the effect of removing N-linked glycans on the number of identifiable peptides and proteins. We compared two synovial fluid samples using the FASP digestion method evaluation. Number of proteins uniquely identified in at least two of the three synovial fluid samples, using FASP digestion, urea in-solution digestion, and in-gel digestion.
In order to identify synovial fluid proteins that are secreted by the synovium cells, i.e., transcribed and translated, we generated sequencing data of mRNA as active transcripts for pig synovium. We prepared and sequenced knee synovium from the right and left legs of 6 pigs. On average, we obtained 12.4 million sequencing reads per library, of which 90% mapped to the pig genome and 81% aligned uniquely. We then calculated RPKM values for individual genes and compared their expression level within each individual animal’s left and right knee (Figure 2a). Intra-animal Pearson correlation coefficients exceeded 0.85, indicative of reproducible data (Figure 2d). The top 20 abundant protein-coding mRNAs that did not originate from the ribosome or mitochondria exhibited minimal variation in abundance, as they ranked in the top 1% of all sample-specific RNA-seq data sets that we analyzed (Figure 2c).

Focusing on transcripts encoding proteins containing signal peptides, which are most likely secreted or targeted to cell surfaces, we observed high expression of known connective tissue proteins such as decorin (DCN), fibronectin (FN1), collagen type III (COL3A1), and clusterin (CLU) (Figure 2b). Importantly, the principal lubricating protein in synovial fluid, lubricin (PRG4) made by type B synoviocytes, was also highly expressed. From these data, we produced a pig synovium transcriptome database (Supporting Information Table 2), which is composed of transcripts encoding known and predicted proteins.

Table 1. Most Commonly Observed PTMs of the FASP-Digested Synovial Fluid Samples Based on the Most Frequent Single Features

<table>
<thead>
<tr>
<th>rank</th>
<th>modification and position</th>
<th>modification mass (Da)</th>
<th>no. MS/MS events</th>
<th>modified sites of possible (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carbamidomethyl (C)</td>
<td>57.0215</td>
<td>13 201</td>
<td>100.0</td>
</tr>
<tr>
<td>2</td>
<td>Deamidated (N)</td>
<td>0.9840</td>
<td>1024</td>
<td>8.7</td>
</tr>
<tr>
<td>3</td>
<td>Acetyl at N-term</td>
<td>42.0106</td>
<td>563</td>
<td>2.9</td>
</tr>
<tr>
<td>4</td>
<td>Carbamidomethyl at N-term</td>
<td>57.0215</td>
<td>535</td>
<td>2.7</td>
</tr>
<tr>
<td>5</td>
<td>Glu → pyro-Glu at N-term</td>
<td>−18.0106</td>
<td>522</td>
<td>24.1</td>
</tr>
<tr>
<td>6</td>
<td>Deamidated (Q)</td>
<td>0.9840</td>
<td>440</td>
<td>2.9</td>
</tr>
<tr>
<td>7</td>
<td>Carbamyl (K)</td>
<td>43.0058</td>
<td>434</td>
<td>2.0</td>
</tr>
<tr>
<td>8</td>
<td>Carbamidomethyl (K)</td>
<td>57.0215</td>
<td>420</td>
<td>1.9</td>
</tr>
<tr>
<td>9</td>
<td>Carbamyl at N-term</td>
<td>43.0058</td>
<td>416</td>
<td>2.1</td>
</tr>
</tbody>
</table>

Twenty-eight and 34 unique deglycosylated peptides in the biological repeats were identified using the four criteria described in the Experimental Procedures section, compared to only two peptides in each of the untreated samples (Supporting Information Table 1). However, all deglycosylated peptides belonged to proteins that had already been identified from other peptides and thus resulted in no new protein identifications. Twenty-eight and 34 unique deglycosylated peptides are relatively low numbers for an analysis of a biological fluid known to be rich in glycoproteins, which may be due to a number of factors: First, the proteome of synovial fluid may not be particularly complex. This hypothesis is consistent with the proteomics data we observed in this study and with previously reported 2D-PAGE analyses.14,15,17 Second, N-glycosylation, which is the only glycosylation amenable to PNGase F treatment protocols, may not be highly prevalent in synovial fluid. Synovial fluid contains significant amounts of glycosaminoglycans, possibly lessening the biologic necessity for N-glycosylation.5,12,22 Third, the finite dynamic range of the MS analysis might cause abundant peptides to hinder the detection of less abundant ones. A study that exclusively focuses on O-glycosylation in synovial fluid and combines data from depleted and nondepleted samples will likely identify more extensive glycosylation than that reported here.

Identifying the Transcriptome of Healthy Pig Synovium

In order to identify synovial fluid proteins that are secreted by the synovium cells, we cannot preclude some of the keratins being contaminants. We included several keratins (10, 14, 3, and 75) in the synovial fluid proteome because keratin-depleted and nondepleted samples will likely identify more peptides are relatively low numbers for an analysis of a biological fluid. Synovial fluid contains significant amounts of glycosaminoglycans, possibly lessening the biologic necessity for N-glycosylation.5,12,22 Third, the finite dynamic range of the MS analysis might cause abundant peptides to hinder the detection of less abundant ones. A study that exclusively focuses on O-glycosylation in synovial fluid and combines data from depleted and nondepleted samples will likely identify more extensive glycosylation than that reported here.

Table 2. Summary of Overall Numbers of PSMs in the Deglycosylation Experiments

<table>
<thead>
<tr>
<th>sample</th>
<th>total no. of PSMs</th>
<th>PSM identifying deglycosylated peptides</th>
<th>deglycosylated peptide PSMs/all PSMs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ PNGase F 1</td>
<td>4827</td>
<td>91</td>
<td>1.9</td>
</tr>
<tr>
<td>+ PNGase F 2</td>
<td>4878</td>
<td>86</td>
<td>1.8</td>
</tr>
<tr>
<td>− PNGase F 1</td>
<td>4068</td>
<td>3</td>
<td>0.07</td>
</tr>
<tr>
<td>− PNGase F 2</td>
<td>4372</td>
<td>2</td>
<td>0.05</td>
</tr>
</tbody>
</table>
difference in identified proteins between the pigs likely originates from biological diversity, which is to be expected, and the finite dynamic range of the MS analysis. 53 To investigate the relative protein abundances in synovial fluid, we investigated individual synovial fluid protein abundance by calculating iBAQ values.36 The protein abundance in synovial fluid, as estimated by the 203 iBAQ quantifiable proteins, spans 5 orders of magnitude (Figure 4). As expected, albumin was the most abundant protein, and no apparent bias could be found regarding the abundances of serum-derived proteins (RPKM < 5) compared to proteins expressed by the synovium in the joint (RPKM > 5) (see below).

Estimating the Relative Contributions of Plasma, Synovium, and Cartilage to the Synovial Fluid Proteome

Synovial fluid is a transudate of blood that has additional components added by the surrounding joint tissue including synovium, cartilage, and, in the knee, ligaments and menisci. To determine whether a likely tissue source for a synovial fluid protein can be predicted, we compared protein expression with
the pig synovium transcriptome, a human articular cartilage transcriptome we previously generated, and published mouse liver\textsuperscript{32} and blood\textsuperscript{40} transcriptomes (Supporting Information Table 2). As expected, albumin in synovial fluid solely derives from blood, as its transcript is highly expressed only in liver (Figure 5).

When we compared the proteins detected in the synovial fluid with the mRNA expression levels in four possible sources, we found that the majority of proteins corresponded to highly expressed transcripts regardless of the source and that liver and synovium likely made the greatest contribution (Figure 5a). Further analysis of our data revealed that 171 proteins detected in the synovial fluid had a signal peptide and that liver, blood, and articular cartilage mRNA expression data were available for 149 proteins (Supporting Information Table 2). Of these 149 common proteins in the synovial fluid proteome, 113 most likely derived from plasma transudate since their transcript abundance is 2-fold or higher in liver or blood than that in synovium or cartilage. In contrast, 37 synovial fluid proteins have transcripts whose expression in synovium is at least 2-fold higher than that in other tissues, while 11 proteins have at least 2-fold or higher transcript expression in articular cartilage. Several proteins serve as positive controls in validation of our estimations. For example, aggrecan, type II collagen, cartilage oligomeric matrix protein, and cartilage intermediate layer protein 2 are detected in the synovial fluid, and the corresponding transcripts are expressed at greater levels in articular cartilage than in other tissues. Similarly, proteins involved in post-translational modification of type I collagen (such as procollagen C-endopeptidase enhancer) and connective tissue markers (such as fibronectin 1 and clusterin) were expressed at higher levels in synovium than in other tissues. Thus, the putative origins of a substantial number of synovial proteome constituents can be predicted from the RNA sequencing data. For those 17 proteins whose mRNAs are expressed by multiple tissues, tissue-specific changes in mRNA abundance will need to be correlated with changes in synovial protein abundance to reliably determine their most likely tissue of origin.

We have additionally detected several proteins without a signal peptide, which are likely residual fragments that remained in the synovial fluid following apoptosis. Some examples of these are hemoglobin (HBA and HBB), which are expressed at extremely higher levels in blood than the remaining tissues, and beta-actin, a ubiquitously expressed protein whose mRNA was detected at high levels (RPKM > 180) in all four tissues. Also of interest are proteins not found in healthy pig synovial fluid and expressed at low levels in healthy synovium. Increased amounts of matrix metalloproteinases, including MMP13 and ADAMTS4, have been reported in synovial fluid from humans with OA, while proteins involved in the inflammatory/immune response, including complement activation, were found in patients with RA\textsuperscript{1}. These proteins were neither detected in healthy pig synovial fluid nor were their transcripts abundant in healthy synovium. Therefore, in addition to monitoring changes in protein abundance for the healthy synovial fluid proteome, it will be of interest to compare proteomes and transcriptomes of healthy and diseased joints to identify pathways that may be integral to disease processes or mechanisms. Previous comparisons of synovial fluid obtained from patients with RA and OA identified 135 proteins at least 3-fold differentially abundant between the two groups.\textsuperscript{20}

Comparison of Synovial Fluid and Plasma

Finally, to identify similarities between synovial fluid and blood plasma, from which many of the synovial fluid components were found to originate, we characterized the human plasma proteome (Supporting Information Table 3) and identified and quantified a total of 168 human plasma proteins. The detected iBAQ values of the human plasma proteins span 5 orders of

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Analysis of identified synovial fluid proteins. (a) Identified proteins at 1\% FDR sorted by PEP. Proteins unique to the transcriptome database (red squares) were, for the most part, identified with low confidence; hence, all proteins with PEP > 1 × 10\textsuperscript{-10} (red line) were removed from further analysis. (b) Synovial fluid proteins identified using the UniProt database and the transcriptome database fulfilling the filtering criteria. Of the 267 synovial fluid proteins, 140 had RPKM > 5 in the synovial mRNA libraries, indicating they are expressed by synovium.}
\end{figure}

\begin{figure}[h!]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Protein abundance distribution in synovial fluid. Frequency histogram of the calculated iBAQ values for all proteomics quantifiable proteins. The relative proportion of proteins whose mRNA transcript abundances were less than or greater than 5 RPKM in the synovial membrane transcriptome are colored blue and yellow, respectively.}
\end{figure}
magnitude, similar to what was found for the synovial fluid, demonstrating the large degree of protein abundance variation encountered in the two body fluids. However, the actual range of protein abundances in plasma and synovial fluid is expected to be several orders of magnitude larger, due to the limited dynamic range of the MS analysis. Probing the top 10 most abundant proteins in both fluids, the most abundant protein is serum albumin, as expected (Table 3). Other shared high-abundance proteins include serotransferrin, apolipoprotein A, and several immunoglobulins. Serotransferrin and apolipoprotein A are binding and transporting proteins of iron and lipids, respectively. T h e h i g h a b u n d a n c e o f t h e t r a n s p o r t e r proteins and immunoglobulins points to the shared functions of the two body fluids as a transport media of cellular products.

The main difference between the high-abundance proteins in the plasma and synovial fluid proteomes is the high abundance of hemoglobin in the synovial fluid, which is detected with a lower abundance in plasma (rank 25). In both fluids, the protein is a likely contaminant from red blood cells prior to centrifugation. Fibrinogens, involved in blood clotting, were identified with high abundance in plasma in contrast to that in synovial fluid, as expected.

Our findings demonstrate the high degree of similarity shared by the synovial fluid proteome and the plasma proteome as well as the serum-derived nature of many synovial fluid components. The expected high degree of similarity found between human plasma and porcine synovial fluid, furthermore, indicates that porcine model systems of synovial joints are suitable for studies focusing on human joint diseases.

## CONCLUSIONS

We evaluated different proteomic strategies for analyzing healthy synovial fluid. We utilized porcine synovial fluid because this animal is used to model human joint disease. In this study, we developed a FASP-based analysis pipeline that consumes only 0.1% of the starting sample material and 10% of the MS instrument time used by, e.g., Balakrishnan; thus, our method lends itself to clinical proteomics studies on larger...
cohorts. Applying this fast and efficient workflow, we completed a normative proteomics study of synovial fluid from healthy knee joints and identified more than 250 proteins using very stringent identification criteria. Preprocessing with PNGase F to remove N-glycosylation was not found to be essential for the proteomic analysis of synovial fluid. Through an analysis of high-occurrence peptide PTMs, we found that proteomics projects using the FASP protocol for synovial fluid analysis, and possibly for other biological samples as well, should include peptide N-terminal acetylation, deamidation of asparagine and glutamine, and peptide N-terminal pyroglutamic acid in the data analysis for comprehensive peptide identification.

Going beyond the conventional LC–MS-based proteome mapping, we also classified and predicted the likely tissue origins for the majority of the detected proteins in healthy synovial fluid by cross-referencing the proteome with the RNA transcriptomes from synovium, cartilage, blood, and liver. Although many proteins derive from plasma transudate, as expected, an important fraction appears to be solely expressed by synovium or cartilage. Changes in the abundance of these latter proteins in synovial fluid or in blood may be useful biomarkers of disease onset or progression. For synovial fluid proteins whose transcripts are expressed in multiple tissues, it will be necessary to correlate tissue-specific changes in mRNA expression with changing synovial fluid protein abundance in order to determine these proteins’ principal tissues of origin. Finally, as many proteins were found to likely originate from plasma, we compared the identified porcine synovial fluid proteome to the proteome of human plasma. We found that the two body fluids share many similarities in terms of protein functions and localizations, underlining the detected plasma derived nature of many synovial fluid components.

Our methods and the resulting proteome and transcriptome data sets will be useful when comparing porcine synovial fluid in healthy and diseased states. Importantly, the high homology between pigs and humans should make these methods and data sets valuable for human studies.

ASSOCIATED CONTENT

Supporting Information

Table 1 contains all identified glycosylated peptides. Table 2 contains all identified proteins, iBAQ and RPKM values (when available), sequence coverage, and imported transcriptome information from associated mouse tissues. Table 3 contains all identified human plasma proteins. Table 4 contains additional information for all proteins identified from a single peptide. This material is available free of charge via the Internet at http://pubs.acs.org. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[^18] with the data set identifier PXD000935.

AUTHOR INFORMATION

Corresponding Author

*Phone: +1-617-919-2629. Fax: +1-617-730-0148. E-mail: hanno.steen@childrens.harvard.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This investigation was supported by National Institutes of Health under NIAIM AR054099 (M.M.M.) and AR050180 (M.L.W.) and by the Ruth L. Kirschstein National Research Service Award (F32 AR061186) (C.M.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The overseas collaboration was supported by the Knud Højgaard's Foundation Denmark, the Lundbeck Foundation Denmark, the Oticon Foundation Denmark, and the Otto Monsted Foundation Denmark. The authors would like to thank the ARCH staff, Dr. Arthur Nedder, Kathryn Mullen, Dana Bolgen, and Courtney White for their assistance and care in handling the minipigs, Saima Ahmed, Nino Esile, and Kevin Broadbelt for assistance in the proteomics laboratory, and Lau Sennels for help with the proteomics data analysis.

ABBREVIATIONS

2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; ACN, acetonitrile; GI, clusterin; COL3A1, collagen type III; CS, consensus site; DCN, decorin; FA, formic acid; FASP, filter-aided sample preparation; FDR, false discovery rate; FN1, fibronectin; HBA, alpha hemoglobin; HBB, beta hemoglobin; iBAQ, intensity-based absolute quantitation; LC, liquid chromatography; LEPT, secreted protein leptin; MALDI, matrix-assisted laser desorption/ionization; MS, mass spectrometry; OA, osteoarthritis; PEP, posterior error probability; PNGase F, peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amide; PPARG, adipocyte-derived transcription factor; PRG4, lubricin; PSM, peptide spectral matches; PTM, post-translational modification; RA, rheumatoid arthritis; RPKM, reads per kilobase of exon per million mapped reads; SDS, sodium dodecyl sulfate; TOF, time-of-flight

REFERENCES


