Effective protein extraction combined with data independent acquisition analysis reveals a comprehensive and quantifiable insight into the proteomes of articular cartilage and subchondral bone

Bundgaard, Louise; Åhrman, Emma; Malmström, Johan; Keller, Ulrich auf dem; Walters, Marie; Jacobsen, Stine

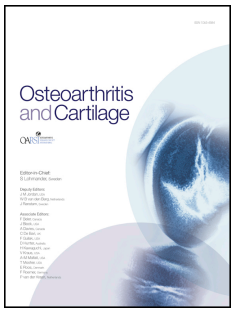
Published in:
Osteoarthritis and Cartilage

Link to article, DOI:
10.1016/j.joca.2021.09.006

Publication date:
2022

Document Version
Peer reviewed version

Journal Pre-proof



Effective protein extraction combined with data independent acquisition analysis reveals a comprehensive and quantifiable insight into the proteomes of articular cartilage and subchondral bone

Louise Bundgaard, DVM, PhD, Emma Åhrman, PhD, Johan Malmström, Professor, Ulrich auf dem Keller, Professor, Marie Walters, DVM, Stine Jacobsen, DVM, Professor

PII: S1063-4584(21)00909-2

DOI: https://doi.org/10.1016/j.joca.2021.09.006

Reference: YJOCA 4923

To appear in: *Osteoarthritis and Cartilage*

Received Date: 25 May 2021

Revised Date: 31 August 2021

Accepted Date: 13 September 2021

Please cite this article as: Bundgaard L, Åhrman E, Malmström J, Keller Uad, Walters M, Jacobsen S, Effective protein extraction combined with data independent acquisition analysis reveals a comprehensive and quantifiable insight into the proteomes of articular cartilage and subchondral bone, *Osteoarthritis and Cartilage*, https://doi.org/10.1016/j.joca.2021.09.006.

Title page

Title

Effective protein extraction combined with data independent acquisition analysis reveals a comprehensive and quantifiable insight into the proteomes of articular cartilage and subchondral bone

Authors

Louise Bundgaard, DVM, PhD; Section of Medicine and Surgery, Department of Veterinary Clinical Sciences, University of Copenhagen, 2630 Taastrup Denmark. Section for Protein Science and Biotherapeutics, DTU Bioengineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark Email: lb@sund.ku.dk

Emma Åhrman, PhD; Division of Infection Medicine Proteomics, Department of Clinical Sciences, Lund University, Lund 221 84, Sweden. Email: emma.ahrman@gmail.com

Johan Malmström, Professor; Division of Infection Medicine Proteomics, Department of Clinical Sciences, Lund University, Lund 221 84, Sweden. Email: johan.malmstrom@med.lu.se

Ulrich auf dem Keller, Professor; Section for Protein Science and Biotherapeutics, DTU Bioengineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark. Email: uadk@dtu.dk

Marie Walters, DVM; Section of Medicine and Surgery, Department of Veterinary Clinical Sciences, University of Copenhagen, 2630 Taastrup Denmark. Email: emw@sund.ku.dk

Stine Jacobsen, DVM, Professor; Section of Medicine and Surgery, Department of Veterinary Clinical Sciences, University of Copenhagen, 2630 Taastrup Denmark. Email: stj@sund.ku.dk

Corresponding author

Louise Bundgaard, DVM, PhD

University of Copenhagen

Department of Veterinary Clinical Sciences

Section of Medicine and Surgery

Agrovej 8

2630 Taastrup

Denmark

Mail: lb@sund.ku.dk

Phone: +45 20609679

Abstract

Objective

The objectives of this study was to establish a sensitive and reproducible method to map the cartilage and subchondral bone proteomes in quantitative terms, and mine the proteomes for proteins of particular interest in the pathogenesis of osteoarthritis (OA). The horse was used as a model animal.

Design

Protein was extracted from articular cartilage and subchondral bone samples from three horses in triplicate by pressure cycling technology or ultrasonication. Digested proteins were analysed by data independent acquisition based mass spectrometry. Data was processed using a pre-established spectral library as reference database (FDR 1%).

Results

We identified to our knowledge the hitherto most comprehensive quantitative cartilage (1758 proteins) and subchondral bone (1482 proteins) proteomes in all species presented to date. Both extraction methods were sensitive and reproducible and the high consistency of the identified proteomes (> 97% overlap) indicated that both methods preserved the diversity among the extracted proteins.

Proteome mining revealed a substantial number of quantifiable cartilage and bone matrix proteins and proteins involved in osteogenesis and bone remodeling, including ACAN, BGN, PRELP, FMOD, COMP, ACP5, BMP3, BMP6, BGLAP, TGFB1, IGF1, ALP, MMP3, and collagens.

A number of proteins, including COMP and TNN, were identified in different protein isoforms with potential unique biological roles.

Conclusion

We have successfully developed two sensitive and reproducible non-species specific workflows enabling a comprehensive quantitative insight into the proteomes of cartilage and subchondral bone. This facilitates the prospect of investigating the molecular events at the osteochondral unit in the pathogenesis of OA in future projects.

Keywords

Cartilage, Subchondral bone, Proteome, Osteoarthritis, Data Independent Acquisition, Horse

Running headline

Proteomes of cartilage and subchondral bone

Introduction

Osteoarthritis (OA) is the final stage of joint failure with chronic inflammation of synovial tissue, subchondral bone damage and cartilage erosion, but the temporal relationship of the pathological events in OA is still largely unknown [1]. Experimental data have shown perforations in the subchondral bone plate in the early phases of OA, followed by increased vascular invasion into the cartilage, and cartilage fibrillation [2–6]. Other studies showed simultaneous onset of OA pathology in articular cartilage and subchondral bone [7,8]. Although loss of cartilage is a prominent feature of OA, it is now commonly accepted that all layers of the osteochondral unit – the articular cartilage, subchondral bone, and the calcified layer in-between – are affected, and that a cross-talk at the bone-cartilage interface is involved in the pathogenesis [9,10]. Given that cartilage and bone pathology are strongly associated in OA initiation and progression, it is of great importance to understand the molecular changes in the osteochondral unit. A focus paper from 2016 [11] emphasized the need of high-throughput proteomics methods to study the disturbances in subchondral bone and the bone-cartilage crosstalk in OA, but also recognized the technical challenges of protein extraction from especially bone, due to the high abundance of collagens and minerals. Various homogenization, extraction, and separation protocols have been tested for this purpose, but with varying and relatively low protein yield [12–18]. Major disadvantages of the techniques used so far are the high amount of tissue needed, that some techniques are very time consuming, and/or that some cause selective or partial loss of specific groups of proteins. New homogenization and extraction methods are thus warranted. A recent refined method based on pressure cycling

technology (PCT) has proved useful for reproducible, rapid, and effective extraction of proteins from various biological samples, even with minimal sample amount [19]. Another method successfully used to disrupt small tissue samples and extract proteins is based on ultrasonication (US) [20]. These two methods displayed consistent and comparable results when processing soft tissue [20], but the efficiency when processing harsher material, such as cartilage and subchondral bone, has not been investigated.

Mass spectrometry (MS)-based proteomics enables detection and quantification of thousands of proteins across various biological samples in an unbiased fashion and circumvents the use of antibodies. The new generation of mass spectrometers and improvement of bioinformatics tools have facilitated the development of the analysis approach called data independent acquisition (DIA). Compared to the standard MS methods, where the proteins precursor ions are selected for analysis based on abundance, DIA is based on analysis of *all* precursor ions, thus giving a more comprehensive and reliable coverage of the proteome. Protein identification relies on comparing the DIA spectra to annotated sets of spectra in a spectral library and quantification is based on the extracted precursor intensity of the DIA spectra [21].

Biological events that lead to progressive joint degradation are difficult to evaluate in humans [22]. While use of smaller mammals is most common as animal models in OA research, large animal models provide more clinically relevant results [23,24], and the horse has been suggested to be the most relevant of all model animals [25,26]. Similar to humans, the horse is a long-lived, athletic species, which shows low intrinsic capacity for repair of cartilage defects [27].

In this study, we used either PCT or US in combination with DIA-MS analysis to obtain and investigate the equine articular cartilage and subchondral bone proteome. The robustness and

sensitivity of the two workflows were evaluated, and the proteomes mined for proteins of particular interest for the pathogenesis of OA. Using these workflows, we have mapped the hitherto most comprehensive quantitative cartilage and subchondral bone proteomes for horses.

[Figure 1]

Method

Articular cartilage and subchondral bone from three biological replicates (Suppl. 1), each in three technical replicates, were independently homogenized by PCT and US followed by DIA-MS analysis and data processing based on a pre-established spectral library (Fig. 1, suppl. 2).

Sampling

18 cartilage and 18 subchondral bone samples were obtained from the fetlock (metacarpophalangeal) of three research horses post-euthanasia from unrelated terminal studies (Suppl. 1). All sampling procedures were approved by the ethics and welfare committee of Department of Veterinary Clinical Sciences, University of Copenhagen, and procedures were carried out according to the Danish Act on animal experiments.

Within 1 hour post mortem, the skin was removed and the joint carefully opened with a scalpel blade. Cartilage was harvested by dissecting thin flakes from the surface of the joint by use of a

clean scalpel blade (Fig. 1, suppl. 1) and washed in PBS to remove synovial fluid (SF). After careful removal of all visible cartilage from the joint surface, subchondral bone was sampled by curetting the surface using a Volkmann bone curette (Fig. 1, suppl. 1). Samples were snap frozen in liquid nitrogen and stored at −80°C.

Sample processing

Homogenization by ultrasonication

Approximately 10 mg tissue was dissolved in 4 M guanidine hydrochloride (GuHCl) (Sigma-Aldrich) in 50 mM HEPES (pH 7.8) (Sigma-Aldrich) and homogenized by US (Bioruptor Pico, Diagenode) for 30 min (60 cycles, 30 sec ON and 30 sec OFF) at 4°C. Disulphide bonds were reduced (10 mM TCEP (Sigma-Aldrich), 30 min, 37°C, 300 rpm), alkylated (40 mM chloroacetamide (CAA) (Sigma-Aldrich) 30 min, room temperature (RT)), and 50 mM HEPES (pH 7.8) added to a final concentration of 1.2 M GuHCl. Protein concentration was measured using the Bradford assay according to the manufactures protocol (http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6835.pdf). All extracted protein was digested with Lys C (1:100 w/w) (Lysyl endopeptidase, #125-05061, Wako) (2 h, 37°C, 500 rpm), and 50 mM HEPES (pH 7.8) added to a final concentration of 0.8 M GuHCl, followed by tryptic digest (1:20 w/w) (Trypsin Gold, V5280, Promega) over night (17 h, 37°C, 500 rpm). The supernatant was recovered after centrifugation for 10 min at 13,000*g* and the digestion quenched by

addition of 10% trifluoroacetic acid (TFA) (Sigma-Aldrich) to pH 2-3. Samples were stored at -80 (Suppl. 3).

Homogenization by pressure cycling technology

The tissue samples (2.5-3 mg corresponding to 2-3 ~1 mm^3 cubes) were transferred to PCT micro-tubes, dissolved in 30 µL 4M GuHCl in 50 mM HEPES (pH 7.8), and homogenized by PCT (Barocycler 2320ETX, Pressure Biosciences Inc.) for 60 min at 4°C (60 cycles, 45 kpsi for 50s, atmospheric pressure for 10 s). Disulphide bonds were reduced and alkylated, samples diluted with HEPES and protein concentration measured as described in the section above. All extracted protein was digested with Lys C (1:100 w/w) by PCT (45 min, 33°C (45 cycles, 20 kpsi for 50s, atmospheric pressure for 10 s)) followed by addition of 50 mM HEPES (pH 7.8) to a final concentration of 0.8 M GuHCl and tryptic digest (1:20 w/w) by PCT (90 min, 33°C (90 cycles, 20 kpsi for 50s, atmospheric pressure for 10s). The supernatant was recovered and the digestion quenched as described in the section above. Samples were stored at -80 (Suppl. 4).

Desalting

The SOLAµ vacuum manifold and solid phase extraction (SPE) micro elution method (SOLAµ) (Thermo Scientific) was used to desalt the samples as follows. Columns were conditioned with 200 µL acetonitrile (ACN) (Sigma-Aldrich), equilibrated with 400 µL 0.1% TFA, and 250 µL of the sample was loaded diluted in 250 µL 0.1% TFA. The sample on column was washed with 500 µL

0.1% TFA, then 200 µL 0.1% FA (Sigma-Aldrich) and eluted with 2x 50µL 50% ACN, 0.2% FA. The samples were vacuum dried and resuspended in 50 µL HPLC-water with 2% ACN 0.2% FA and synthetic peptides (1:20 v/v) (iRT peptide kit, Biognosys AG) for retention time calibration. The peptide concentration was measured by use of the Pierce Quantitative Colorimetric Peptide Assay (Thermo Scientific) according to the manufactures protocol (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/23275_quantpeptide_color_UG.pdf).

Data independent acquisition mass spectrometry analysis

Samples (~1 µg) were separated on a PepMap RSLC C18 analytical column (75 µm x 50 cm, 2µm, 100Å, nanoViper, Thermo Scientific) using the EASY-nLC™ 1200 liquid chromatography system (Thermo Scientific) coupled in-line with a Q Exactive HF-X Hybrid Quadropole-Orbitrap mass spectrometer (Thermo Scientific). Separation was achieved by running constant flow rate of 350 nL/min in 0.1% formic acid/99.9% water and a 120 min gradient from 10% to 90% (10-30% for 90min, 30-45% for 20min, 45-90% for 1min, 90% for 10min) elution buffer (80% acetonitrile, 0.1% formic acid, 19.9% water). A DIA operated under Xcalibur 4.1.31.9 was applied to record the spectra. The full scan MS spectra (340-1400 m/z) were acquired with a resolution of 60,000 after accumulation to a target value of 3e6 and a maximum injection time of 100ms. The MS/MS data was recorded by 32 full fragmentation scans, using an isolation window adjusted to the number of precursors (Suppl. 5). The precursor ions within each

isolation window were fragmented using HCD with a resolution of 30,000 applying an AGC of 1e6 and a maximum injection time of 120ms.

Data analysis

The DIA-MS data from cartilage and subchondral bone samples were searched as individual batches based on a pre-established spectral library. The spectral library covered joint related tissue and SF from nine horses. A detailed description of sample processing and analysis for the spectral library is included in supplementary 2. The software Spectronaut (v13.11.200127.43655, Biognosys) was used to generate the spectral library from Pulsar and search the DIA-MS data with the equine proteome from uniprot (UP000002281 downloaded 01-23-20) as reference database. Trypsin was specified as the enzyme, up to two missed cleavages and a peptide length of 7-52 amino acids was allowed. False discovery rate (FDR) was controlled at 1% for both precursor and protein group levels. Other parameters were kept at default settings. The data was median normalized in Spectronaut and further processed in R (v4.0.2). The spectral library represented 44,922 unique peptides covering 3147 protein groups with high confidence (1% FDR) (Suppl. 2).

Venn diagrams were made in Venny v2.1, Metascape was used for GO term cluster analysis on biological processes [28], and additional graphics were created in GraphPad Prism v8.4.3. The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE[29] partner repository with the dataset identifier PXD025882 and is publicly available.

[Figure 2]

Results

Using our sample preparation and MS workflows, we identified almost 2000 (1953) distinct proteins from cartilage and subchondral bone, 1232 of which were present in samples from both tissues independent of homogenization method (Fig. 2A).

Sensitivity and reproducibility of the workflows

Cartilage

The average amount of protein extracted per mg cartilage was similar for both processing methods (PCT: 12.2 µg/mg; US: 12.1 µg/mg). The proteome recorded from cartilage comprised 1758 different proteins, whereby PCT yielded a mean number of 1468 proteins and US 1525 protein identifications on average (Suppl. 6). For each horse, more than 77% of the identified proteins were present in all three technical replicates. The median CV for protein quantification for technical replicates processed by PCT was 18%, 21%, and 23% for H1, H2, and H3, respectively, and for samples processed by US slightly lower with respective CVs of 12%, 16%, and 12%. Reproducibility between biological replicates was high, with 89% (1527 out of 1724) protein identifications in cartilage samples from all three horses upon processing with PCT (Fig. 2B) and 86% (1482 out of 1728) in US processed specimens (Fig. 2C). High qualitative

reproducibility was matched with quantitative robustness as indicated by CVs between biological replicates of 20% for samples processed by PCT and 21% for samples processed by US (Fig, 2D). Notably, when comparing the identified global proteomes for samples processed by PCT and US, 98% were overlapping. Only 2% (PCT: 19 proteins, US: 21 proteins) of the identified proteins were unique to each of the two preparation methods (Fig. 2A).

Subchondral bone

The average protein concentration per mg tissue was 8.5 µg/mg for samples processed by PCT and 10.9 µg/mg for samples processed by US. Similar to cartilage, our workflow yielded high coverage of the subchondral bone proteome with 1482 high confidence protein identifications combined from both sample processing methods (Fig. 2A). Thereby, the mean number of proteins identified was 1264 for samples processed by PCT and 1218 for samples processed by US (Suppl. 6). More than 78% of the identified proteins were present in all three technical replicates for each horse and could be quantified with median CVs of 22%, 29%, and 24% for H1, H2, and H3, respectively, when processed using PCT, and with respective CVs of 22%, 18%, and 30% upon US sample preparation. Both for samples processed by PCT and US, a percentage of 88% (PCT: 1287 proteins, US: 1265 proteins) of the identified proteome were identified in all three biological replicates (Fig. 2E-F). The CV for biological replicates was 29% and 21% for PCT- and US-processed samples, respectively (Fig. 2D). 1429 proteins were identified both in samples processed by PCT and US, accounting for 97% and 99% of the identified global proteomes, respectively (Fig. 2A).

Protein biological process

To increase confidence in observations even further, we only included proteins identified in at least three samples in the analyses of the proteomes. This reduced the number of proteins from 1953 to 1938.

A biological process enrichment analysis based on GO classification showed that for proteins identified in cartilage only (471 proteins) the top 5 clusters were RNA splicing (GO:0000377), leukocyte degranulation (GO:0043299), regulation of mRNA metabolic process (GO:1903311), ribonucleoprotein complex subunit organization (GO:0071826) and establishment of protein localization to organelle (GO:0071826) (Fig. 3A). For proteins identified in subchondral bone only (195 proteins), the top 5 clusters were ossification (GO:0001503), response to wounding (GO:0009611), skeletal system development (GO:0001501), tissue remodelling (GO:0048771), and negative regulation of hydrolase activity (GO:0010466) (Fig. 3B, 3C).

[Figure 3]

Further mining of the obtained cartilage and subchondral bone proteomes revealed a high abundance of extracellular matrix proteins of major relevance in cartilage and subchondral bone. Figure 4 shows a heat map of a merge of the top 50 most abundant proteins identified in all of the four different types of samples. Biglycan (BGN), fibromodulin (FMOD), prolargin (PRELP), and aggrecan (ACAN) were among the top 5 most abundant proteins in both cartilage

and subchondral bone processed with US or PCT. Various other proteins involved in extracellular matrix organization were identified among the top 50 proteins such as hyaluronan and proteoglycan link protein 1 (HAPLN1), decorin (DCN), cartilage intermediate layer protein (CILP) and CILP2, lumican (LUM), osteomodulin (OMD), fibrillin 1 (FBN1), COL2A1, COL6A1, and COL9A1. Proteins specific to top 50 in cartilage were thrombospondin (THBS) 1, THBS4, vitrin (VIT), and mimecan (OGN), whereas osteopontin (SSP1), SPARC, and COL1A1 were specific to top 50 in subchondral bone.

[Figure 4]

Protein level evidence of expression of different splice variants

When aligning the identified proteins, all sequences were unique, but 62 of the proteins were identified with gene name in duplicates, 3 in triplicates, 8 in quadruplicates, and 5 in quintuplicates (Suppl. 7). This may reflect either potential point mutations in the protein sequence or different splice variants of the gene. A substantial part of the proteins in the equine reference database from Uniprot is derived from transcripts nominated by genomic and transcriptomic technologies and contain alternative splice forms. Proteomics data can be used to obtain protein-level evidence of expression of the transcripts and different splice variants, an approach called proteogenomics [30]. As examples, our data provided protein level validation of sequence variants of cartilage oligomeric matrix protein (COMP) and tenascin N (TNN).

COMP was identified with two different protein sequences, accession F6U3D3 and A0A3Q2H402, in samples from both cartilage and subchondral bone (Fig. 5A). More than 50 different peptides were annotated to accession F6U3D3. Three peptides were annotated to A0A3Q2H402, of which two exon-exon junction peptides were unique to this protein sequence (Fig. 5A). At least one of these unique peptides was identified in all samples from cartilage, but only in five out of the 18 samples from subchondral bone.

Accession A0A3Q2LJB3 and A0A3Q2KYM7 with the common gene name TNN was identified in samples both from cartilage and subchondral bone (Fig. 5B). The exon-exon peptide APTEIDSPQNLVTDR unique to accession A0A3Q2LJB3 span an alternative splice site compared to accession A0A3Q2KYM7. This peptide was identified in all samples where the protein was present (21 out of 36 samples). Accession A0A3Q2KYM7 was present in all but one sample from cartilage.

These sequence variants represent different protein isoforms and some may have unique biological roles.

[Figure 5]

Discussion

We have established two novel reproducible and sensitive workflows for quantitative studies of protein expression in cartilage and subchondral bone, based on either US or PCT for tissue

261 homogenization coupled with DIA-MS analysis. Using these workflows, we have mapped the
262 hitherto most extensive quantitative proteomes for equine cartilage and subchondral bone.
263 There was a high consistency of the identified proteomes using the two workflows (> 97%
264 overlap) for both cartilage and subchondral bone, indicating that both extraction methods
265 preserve the diversity among the extracted proteins. For the cartilage samples the median
266 variation across the technical replicates (12-23%) was comparable to the results from another
267 study, where kidney biopsies were processed employing a similar workflow [31]. The slightly
268 higher variation for samples from subchondral bone (18-30%) was expected because it is a
269 harder tissue and therefore more difficult to homogenize. The hard nature of the tissue may
270 also explain why the median variation is higher than the down to 6% CV found for technical
271 replicate PCT and US processing of spleen samples [20]. However, it is worth to notice that in that
272 study only proteins quantifiable in samples from both extraction methods were included when
273 calculating the CV. Pre-homogenization grinding of the tissue might decrease the variation but
274 will increase the time of processing.
275 The inclusion of three technical replicates ensured a reliable coverage of the global proteome in
276 both cartilage and subchondral bone. The between-horse variation (20-29%) was markedly
277 lower than the up to 45% inter-patient variation found in the comparable study on kidney
278 biopsies [31]. Overall, this method is regarded very reproducible and with a total of 1758 unique
279 proteins identified in cartilage and 1482 unique proteins identified in subchondral bone with
280 high confidence, the results reveals the most comprehensive quantitative cartilage and
281 subchondral bone proteomes in all species presented to date [12–17].

Proteins in cartilage and subchondral bone that are embedded in the solid material, insoluble, or extensively cross-linked are challenging to detect by MS. In future studies, the insight into the proteome could be expanded even further by including alternative proteases such as LysargiNase and the endoproteinases GluC and AspN to identify proteins not amenable to digestion by endoproteinase LysC and trypsin [32].

The spectral library described herein is the first published library covering joint tissue and SF samples from the horse. This high-quality dataset is a resource for the research community for bioinformatical mining, or for future DIA experiments on joint tissues and fluids. Even though this is a comprehensive spectral library, the proteome of joint-related structures is not exhaustively covered. In the future, the depth of the spectral library proteome can be further extended by including data from e.g. pre-fractionation of samples prior to LC separation [33]. The advantage of DIA is that as the spectral library is expanded with more peptide information the acquired data can be re-searched and maybe contribute with even more information. As expected, many proteins identified only in SB are involved in osteogenesis and bone remodelling, e.g. tartrate-resistant acid phosphatase type 5 (ACP5), bone morphogenetic protein (BMP) 3 and 6, osteocalcin (BGLAP), and cathepsin K (CTSK), and the top 1 and 3 biological processes assigned in the GO enrichment analysis were *ossification* and *skeleton system development* [34,35]. Proteins known to be of major importance in bone tissue were identified in subchondral bone only, this further adds to the reliability and validity of the sampling and analysis methods used in this study and emphasize that these are promising workflows for getting a better insight into the molecular events in subchondral bone.

Histological studies have shown that subchondral bone is vascularized and venous plexuses

304 exist on the border between the subchondral bone and the deep cartilage layer [36]. Activation of
305 blood coagulation post mortem or by tearing of these small vessels upon sampling may explain
306 the identification of coagulation factor VII, IX, and X, kininogen-1 (KNG1), vitamin K-dependent
307 protein C (PROC), vitamin K-dependent protein Z (PROZ) in SB only, and that *response to*
308 *wounding* is the top 2 biological process found in the GO enrichment analysis.
309 The 5 most enriched GO terms in cartilage were all related to cellular processes, and this is
310 most likely because cartilage has a substantial number of chondrocytes compared to the more
311 cell poor subchondral bone [10]. The proteoglycan ACAN and the small leucine-rich proteoglycans
312 BGN, PRELP, and FMOD are all major components of the extracellular matrix in cartilage [10,37]
313 and were also found in high abundance in cartilage. These proteins were also found in high
314 abundance in subchondral bone and their presence in bone is in accordance with findings in
315 previous studies [12,37].
316 Mining of the cartilage and subchondral bone proteomes revealed a number of proteins with
317 mutual gene names but different sequence variants. These protein isoforms might have
318 different biological functions.
319 COMP is an extracellular matrix glycoprotein expressed in a wide variety of tissues including
320 cartilage and subchondral bone [38,39] and represents a promising marker for joint tissue
321 degradation [40]. Interestingly, A0A3Q2H402 was identified in all samples from cartilage, but only
322 a few samples from subchondral bone, suggesting that A0A3Q2H402 could be a more dominant
323 COMP isoform in cartilage. However, to our knowledge, no research studies have described
324 isoform variants of COMP in any species, but Uniprot lists different isoforms in humans and

mice and mutant COMP variants specific for certain diseases exits [38]. Further research is needed to determine if these COMP isoforms have distinct tissue specificity.

TNN was another protein identified in two isoforms. TNN is an extracellular matrix protein involved in osteogenesis, and a study in mice showed that TNN was dysregulated in OA [41]. TNN was identified with a longer and a shorter sequence in both cartilage and subchondral bone comparable to the two splice variants described in mice [42]. In mice, the biological function of the two transcripts differed and it could be of high interest to investigate if the same applies to the horse and maybe have an impact on the pathogenesis for OA.

For many of the proteins with two or more transcripts the distinction were based on only one unique peptide, but according to the guidelines from the Human Proteome Organization protein quantification should rely on two unique peptides [43]. However, we found that we missed out potential important findings, e.g. some of the potential protein isoforms, by including the more strict setting of two unique peptides per protein in the search criteria. The unique peptides identified for all the different proteins are useful to build parallel reaction monitoring studies for further investigation and validation of the different transcripts, specific degradation products and the biological function or relevance for specific diseases.

A review from 2020 [44] summarized the results from 11 different OA studies over the last decade on gene expression profiling of various tissues, such as cartilage, subchondral bone, and synovium from human OA and mouse OA models. They listed 87 differentially expressed genes mainly involved in matrix metabolism, bone remodeling, and inflammation pathways in OA tissue compared to normal tissue, including ALPL, OGN, IGF1, TGF-β1, TGFBI, POSTN, MMP3, MMP13, ACP5, ASPN, IHH and multiple collagens. We identified the corresponding proteins to

36 of the genes and close protein family members to 19 of the genes. The prospect of investigating the interrelation of these proteins together with the other identified cartilage and subchondral bone related proteins in one single analysis emphasizes that PCT or US homogenization combined with DIA analysis is a very promising workflow to get a valuable and deep insight into the molecular events at the osteochondral unit.

In conclusion, we have successfully developed two workflows based on either PCT or US for tissue homogenization and protein extraction coupled with DIA-MS analysis. The workflows enabled to our knowledge the hitherto most comprehensive quantitative insights into the equine proteomes of cartilage and subchondral bone, and facilitate the prospect of investigating the molecular events at the osteochondral unit in the pathogenesis of OA and other cartilage and bone related diseases in future projects in all species including humans. Both workflows were very reproducible and the high overlap of more than 97 percentage between the two workflows for both cartilage and subchondral bone indicates that the proteomes cover the true biological composition.

The proteomes and the spectral library will be open access resources for the research community for further mining of proteins of interest, and as support for development of methods for targeted analysis of specific proteins or potential protein isoforms.

Acknowledgements

Not applicable

Author contributions

LB conceived the study, sampled most of the samples, prepared all samples for analysis, interpreted the data, and drafted the manuscript. EA provided expert assistance with sample preparation for MS analysis, MS analysis, data interpretation, and revised the manuscript critically. JM conceived the study, provided expert assistance with design of the project, MS analysis, data interpretation, and revised the manuscript critically. UadK conceived the study, provided expert assistance with design of the project, MS analysis, data interpretation, and revised the manuscript critically. MW assisted with design of the project and method development, sampling, and revised the manuscript critically. SJ conceived the study, provided expert assistance with the project design and method development, sampling, data interpretation, and manuscript preparation and revision. All authors read and approved the final version of the manuscript.

Role of the funding source

Funding was generously provided by The Danish Council for Independent Research (grant number DFF-7017-00066) and Gerda and Aage Haensch's foundation to cover the project expenses. JM acknowledges support by The Swedish Research council (Grant no.

2019-00206). UadK acknowledges support by a Novo Nordisk Foundation Young Investigator Award (NNF16OC0020670).

Competing interest statement

The authors declare to have no conflict of interests.

References

1. Kapoor M, Martel-Pelletier J, Lajeunesse D, Pelletier JP, Fahmi H. Role of proinflammatory cytokines in the pathophysiology of osteoarthritis. *Nat Rev Rheumatol*. 2011;7(1):33-42. doi:10.1038/nrrheum.2010.196
2. Botter SM, van Osch GJVM, Clockaerts S, Waarsing JH, Weinans H, van Leeuwen JPTM. Osteoarthritis induction leads to early and temporal subchondral plate porosity in the tibial plateau of mice: an in vivo microfocal computed tomography study. *Arthritis Rheum*. 2011;63(9):2690-2699. doi:10.1002/art.30307
3. Liu C, Liu C, Si L, Shen H, Wang Q, Yao W. Relationship between subchondral bone microstructure and articular cartilage in the osteoarthritic knee using 3T MRI. *J Magn Reson Imaging*. 2018;48(3):669-679. doi:10.1002/jmri.25982
4. Hayami T, Pickarski M, Zhuo Y, Wesolowski GA, Rodan GA, Duong LT. Characterization of

articular cartilage and subchondral bone changes in the rat anterior cruciate ligament transection and meniscectomized models of osteoarthritis. *Bone*. 2006;38(2):234-243. doi:10.1016/j.bone.2005.08.007

5. Walsh DA, Mcwilliams DF, Turley MJ, Dixon MR, Fransès RE, Mapp PI, et al. Angiogenesis and nerve growth factor at the osteochondral junction in rheumatoid arthritis and osteoarthritis. *Rheumatology*. 2010;49(10):1852-1861. doi:10.1093/rheumatology/keq188

6. Suri S, Gill SE, De Camin SM, Wilson D, McWilliams DF, Walsh DA. Neurovascular invasion at the osteochondral junction and in osteophytes in osteoarthritis. *Ann Rheum Dis*. 2007;66(11):1423-1428. doi:10.1136/ard.2006.063354

7. Fang H, Huang L, Welch I, Norley C, Holdsworth DW, Beier F, et al. Early Changes of Articular Cartilage and Subchondral Bone in The DMM Mouse Model of Osteoarthritis. *Sci Rep*. 2018;8(1):2855. doi:10.1038/s41598-018-21184-5

8. Thomsen JS, Straarup TS, Danielsen CC, Oxlund H, Brüel A. Relationship between articular cartilage damage and subchondral bone properties and meniscal ossification in the Dunkin Hartley guinea pig model of osteoarthritis. *Scand J Rheumatol*. 2011;40(5):391-399. doi:10.3109/03009742.2011.571218

9. Pan J, Wang B, Li W, Zhou X, Scherr T, Yang Y, et al. Elevated cross-talk between subchondral bone and cartilage in osteoarthritic joints. *Bone*. 2012;51(2):212-217. doi:10.1016/j.bone.2011.11.030

10. Goldring SR, Goldring MB. Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk. *Nat Rev Rheumatol*. 2016;12(11):632-

644. doi:10.1038/nrrheum.2016.148

11. Fellows CR, Matta C, Mobasheri A. Applying Proteomics to Study Crosstalk at the Cartilage-Subchondral Bone Interface in Osteoarthritis: Current Status and Future Directions. *EBioMedicine*. 2016;11:2-4. doi:10.1016/j.ebiom.2016.08.047

12. Jiang X, Ye M, Jiang X, Liu G, Feng S, Cui L, et al. Method development of efficient protein extraction in bone tissue for proteome analysis. *J Proteome Res*. 2007;6(6):2287-2294. doi:10.1021/pr070056t

13. Pastorelli R, Carpi D, Airoldi L, Chiabrando C, Bagnati R, Fanelli R, et al. Proteome analysis for the identification of in vivo estrogen-regulated proteins in bone. *Proteomics*. 2005;5(18):4936-4945. doi:10.1002/pmic.200401325

14. Schreiweis MA, Butler JP, Kulkarni NH, Knierman MD, Higgs RE, Halladay DL, et al. A proteomic analysis of adult rat bone reveals the presence of cartilage/chondrocyte markers. *J Cell Biochem*. 2007;101(2):466-476. doi:10.1002/jcb.21196

15. Desjardin C, Balliau T, Valot B, Zivy M, Wimel L, Guérin G, et al. A method for proteomic analysis of equine subchondral bone and epiphyseal cartilage. *Proteomics*. 2012;12(11):1870-1874. doi:10.1002/pmic.201100366

16. Guo D, Tan W, Wang F, Lv Z, Hu J, Lv T, et al. Proteomic analysis of human articular cartilage: Identification of differentially expressed proteins in knee osteoarthritis. *Jt Bone Spine*. 2008;75(4):439-444. doi:10.1016/j.jbspin.2007.12.003

17. Folkesson E, Turkiewicz A, Englund M, Önnerfjord P. Differential protein expression in human knee articular cartilage and medial meniscus using two different proteomic methods: A pilot analysis. *BMC Musculoskelet Disord*. 2018;19(1):416.

doi:10.1186/s12891-018-2346-6

18. Vincourt J-B, Lionneton F, Kratassiouk G, Guillemin F, Netter P, Mainard D, et al. Establishment of a reliable method for direct proteome characterization of human articular cartilage. *Mol Cell Proteomics*. 2006;5(10):1984-1995. doi:10.1074/mcp.T600007-MCP200

19. Shao S, Guo T, Gross V, Lazarev A, Koh CC, Gillessen S, et al. Reproducible Tissue Homogenization and Protein Extraction for Quantitative Proteomics Using MicroPestle-Assisted Pressure-Cycling Technology. *J Proteome Res*. 2016;15(6):1821-1829. doi:10.1021/acs.jproteome.5b01136

20. Kuras M, Betancourt LH, Rezeli M, Rodriguez J, Szasz M, Zhou Q, et al. Assessing Automated Sample Preparation Technologies for High-Throughput Proteomics of Frozen Well Characterized Tissues from Swedish Biobanks. *J Proteome Res*. 2019;18(1):548-556. doi:10.1021/acs.jproteome.8b00792

21. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016;537(7620):347-355. doi:10.1038/nature19949

22. Anderson DD, Chubinskaya S, Guilak F, Martin JA, Oegema TR, Olson SA, et al. Post-traumatic osteoarthritis: improved understanding and opportunities for early intervention. *J Orthop Res*. 2011;29(6):802-809. doi:10.1002/jor.21359

23. Gregory MH, Capito N, Kuroki K, Stoker AM, Cook JL, Sherman SL. A Review of Translational Animal Models for Knee Osteoarthritis. *Arthritis*. 2012;2012:1-14. doi:10.1155/2012/764621

24. Cook JL, Hung CT, Kuroki K, Stoker AM, Cook CR, Pfeiffer FM, et al. Animal models of

cartilage repair. *Bone Joint Res*. 2014;3(4):89-94. doi:10.1302/2046-3758.34.2000238

25. Ahern BJ, Parvizi J, Boston R, Schaer TP. Preclinical animal models in single site cartilage defect testing: a systematic review. *Osteoarthr Cartil*. 2009;17(6):705-713. doi:10.1016/j.joca.2008.11.008

26. Chu CR, Szczodry M, Bruno S. Animal models for cartilage regeneration and repair. *Tissue Eng - Part B Rev*. 2010;16(1):105-115. doi:10.1089/ten.teb.2009.0452

27. Convery FR, Akeson WH, Keown GH. The repair of large osteochondral defects. An experimental study in horses. *Clin Orthop Relat Res*. 1972;82:253-262. doi:10.1097/00003086-197201000-00033

28. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-09234-6

29. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res*. 2019;47(D1):D442-D450. doi:10.1093/nar/gky1106

30. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. *Nat Methods*. 2014;11(11):1114-1125. doi:10.1038/NMETH.3144

31. Guo T, Kouvonen P, Koh CC, Gillet LC, Wolski WE, Röst HL, et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat Med*. 2015;21(4):407-413. doi:10.1038/nm.3807

32. Bell PA, Solis N, Kizhakkedathu JN, Matthew I, Overall CM. Proteomic and N-Terminomic TAILS Analyses of Human Alveolar Bone Proteins: Improved Protein Extraction

Methodology and LysargiNase Digestion Strategies Increase Proteome Coverage and Missing Protein Identification. *J Proteome Res*. 2019;18(12):4167-4179. doi:10.1021/acs.jproteome.9b00445

33. Parker SJ, Venkatraman V, Van Eyk JE. Effect of peptide assay library size and composition in targeted data-independent acquisition-MS analyses. *Proteomics*. 2016;16(15-16):2221-2237. doi:10.1002/pmic.201600007

34. Han Y, You X, Xing W, Zhang Z, Zou W. Paracrine and endocrine actions of bone - The functions of secretory proteins from osteoblasts, osteocytes, and osteoclasts. *Bone Res*. 2018;6(1). doi:10.1038/s41413-018-0019-6

35. Charles JF, Aliprantis AO. Osteoclasts: More than "bone eaters." *Trends Mol Med*. 2014;20(8):449-459. doi:10.1016/j.molmed.2014.06.001

36. Imhof H, Sulzbacher I, Grampp S, Czerny C, Youssefzadeh S, Kainberger F. Subchondral bone and cartilage disease: A rediscovered functional unit. *Invest Radiol*. 2000;35(10):581-588. doi:10.1097/00004424-200010000-00004

37. Yang CH, Culshaw GJ, Liu MM, Lu CC, French AT, Clements DN, et al. Canine tissue-specific expression of multiple small leucine rich proteoglycans. *Vet J*. 2012;193(2):374-380. doi:10.1016/j.tvjl.2012.01.018

38. Posey KL, Coustry F, Hecht JT. Cartilage oligomeric matrix protein: COMPopathies and beyond. *Matrix Biol*. 2018;71-72:161-173. doi:10.1016/j.matbio.2018.02.023

39. Di Cesare PE, Fang C, Leslie MP, Tulli H, Perris R, Carlson CS. Expression of cartilage oligomeric matrix protein (COMP) by embryonic and adult osteoblasts. *J Orthop Res*. 2000;18(5):713-720. doi:10.1002/jor.1100180506

40. Neidhart M, Hauser N, Paulsson M, Dicesare PE, Michel BA, Häuselmann HJ. Small fragments of cartilage oligomeric matrix protein in synovial fluid and serum as markers for cartilage degradation. *Br J Rheumatol*. 1997;36(11):1151-1160. doi:10.1093/rheumatology/36.11.1151

41. Gardiner MD, Vincent TL, Driscoll C, Burleigh A, Bou-Gharios G, Saklatvala J, et al. Transcriptional analysis of micro-dissected articular cartilage in post-traumatic murine osteoarthritis. *Osteoarthr Cartil*. 2015;23(4):616-628. doi:10.1016/j.joca.2014.12.014

42. Neidhardt J, Fehr S, Kutsche M, Löhler J, Schachner M. Tenascin-N: Characterization of a novel member of the tenascin family that mediates neurite repulsion from hippocampal explants. *Mol Cell Neurosci*. 2003;23(2):193-209. doi:10.1016/S1044-7431(03)00012-5

43. Deutsch EW, Lane L, Overall CM, Bandeira N, Baker MS, Pineau C, et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *J Proteome Res*. 2019;18(12):4108-4116. doi:10.1021/acs.jproteome.9b00542

44. Liu W, Jiao Y, Tian C, Hasty K, Song L, Kelly DM, et al. Gene Expression Profiling Studies Using Microarray in Osteoarthritis: Genes in Common and Different Conditions. *Arch Immunol Ther Exp (Warsz)*. 2020;68(5):28. doi:10.1007/s00005-020-00592-4

Figure legends

Figure 1

A) Articular cartilage was harvested by dissecting thin flakes from the surface of the joint by use of a clean scalpel blade. After careful removal of all visible cartilage from the joint surface, subchondral bone was sampled by scraping the surface using a Volkmann bone curette (see suppl. 1 for magnified pictures of the tissue sampling techniques). B) Samples were homogenized by pressure cycling technology or ultrasonication and the extracted proteins digested, before C) analysis by data independent acquisition (DIA) based mass spectrometry and data analysis. *Created with BioRender.com.*

Figure 2

A) Venn diagram showing the relation between the global proteomes identified in cartilage (CAR) and subchondral bone (SB) processed by pressure cycling technology (PCT) or ultrasonication (US). B-C) CAR samples presented in Venn diagrams showing the relation of the proteins quantified in the three biological replicates (horse 1,2,3). D) The violin plot shows the distribution of CV values for the quantified proteins in the biological replicates. The line indicates the median CV, the dashed lines indicates the quartiles. E-F) SB samples presented in Venn diagrams showing the relation of the proteins quantified in the three biological replicates (horse 1,2,3).

Figure 3

563 The column bar graphs shows A) the top 5 GO clusters after biological process enrichment
564 analysis of proteins identified only in A) cartilage samples, B) subchondral bone samples. C)
565 Proteins (represented by gene name) identified in subchondral bone mapping to the top 5 GO
566 clusters. Blue: protein in GO term, White: protein not in GO term.
567
568 Figure 4
569 The heatmap combines the log2 transformed intensities of the 50 most abundant proteins
570 (represented by gene name) in each of the four sample types and corresponding intensities
571 found across the four types of samples. Cartilage (CAR), subchondral bone (SB), pressure cycling
572 technology (PCT), ultrasonication (US).
573
574 Figure 5
575 A) Alignment of part of the protein sequences for accession A0A3Q2H402 and F6U3D3 with the
576 common gene name COMP in the Uniprot database, and the peptides identified in this region.
577 The protein-coding exons are included to show sites for alternative splicing and exon-exon
578 peptides. Peptides identified are marked in red. B) Alignment of the protein sequences for
579 accession A0A3Q2KYM7 and A0A3Q2LJB3 with the common gene name TNN in the Uniprot
580 database, and the peptides identified in this region. The protein-coding exons are included to
581 show sites for alternative splicing and exon-exon peptides. Peptides identified are marked in
582 red. *Created with BioRender.com.*
583
584 Supplementary 1

Samples included in the experimental study including information about, type of processing (pressure cycling technology (PCT), ultrasonication (US)), denaturing agent (guanidine hydrochloride (GuHCl)), desalting method (details for SOLAμ), MS analysis approach (settings for DIA are described in the paper). Magnified pictures of the tissue sampling techniques.

Supplementary 2

Description of the samples and methods used for sampling, processing and analysis of the samples included in the spectral library.

Supplementary 3

Protocol for homogenization and digestion of tissue samples using ultrasonication

Supplementary 4

Protocol for homogenization and digestion of tissue samples using pressure cycling technology

Supplementary 5

Isolation windows used for DIA-MS analysis

Supplementary 6

Venn diagrams showing the relation of the proteins quantified in each of the three technical replicates for all three horses (H1, H2, H3) for cartilage and subchondral bone samples processed with either pressure cycling technology (PCT) or ultrasonication (US).

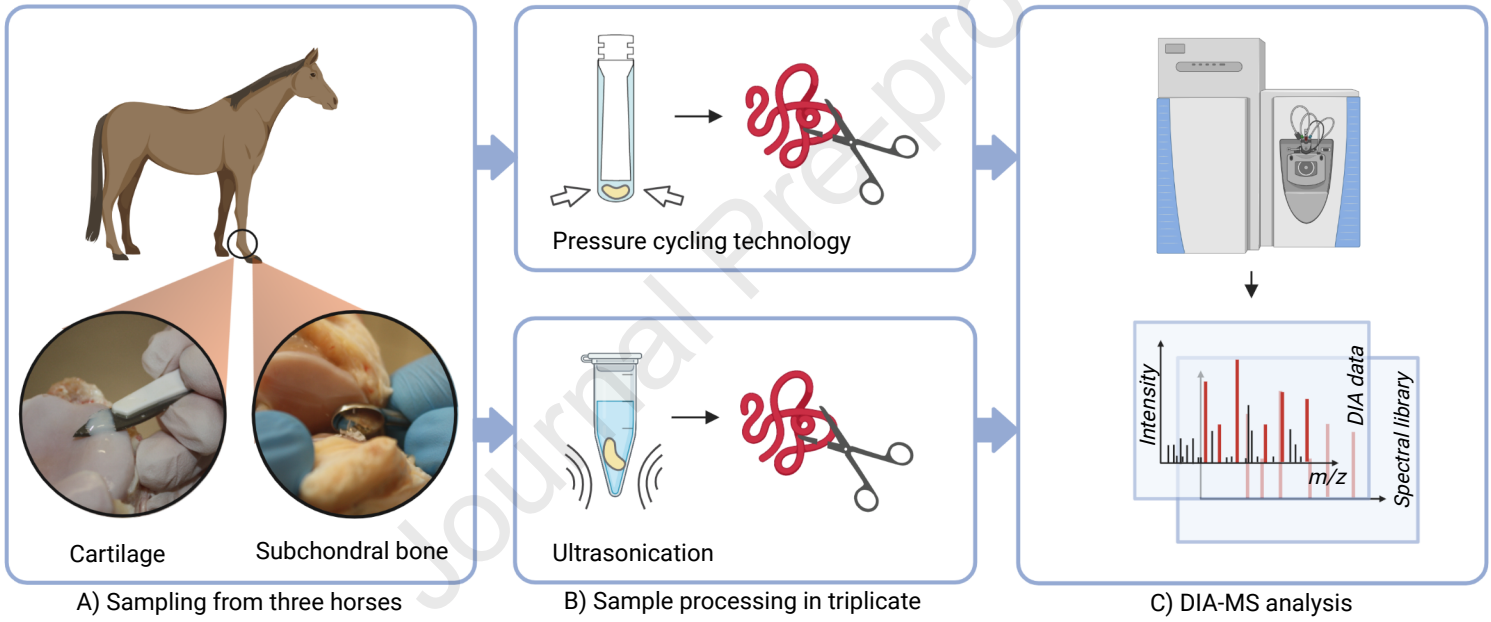
607

608 Supplementary 7

609 All protein sequences with gene name in duplicate, triplicate, quadruplicate, and quintuplicate.

610

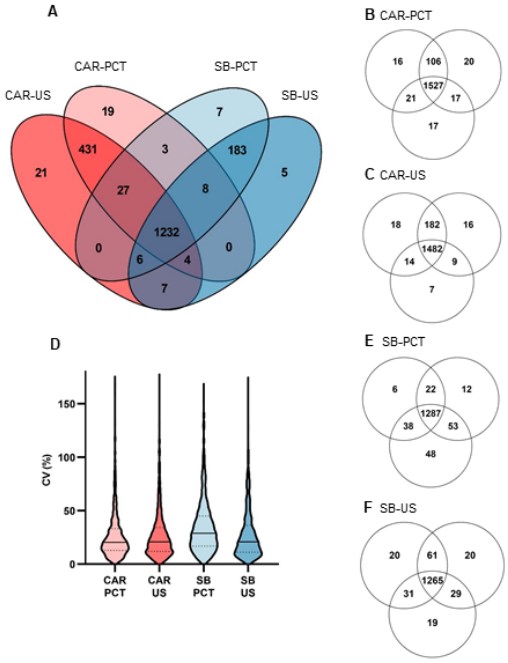
611

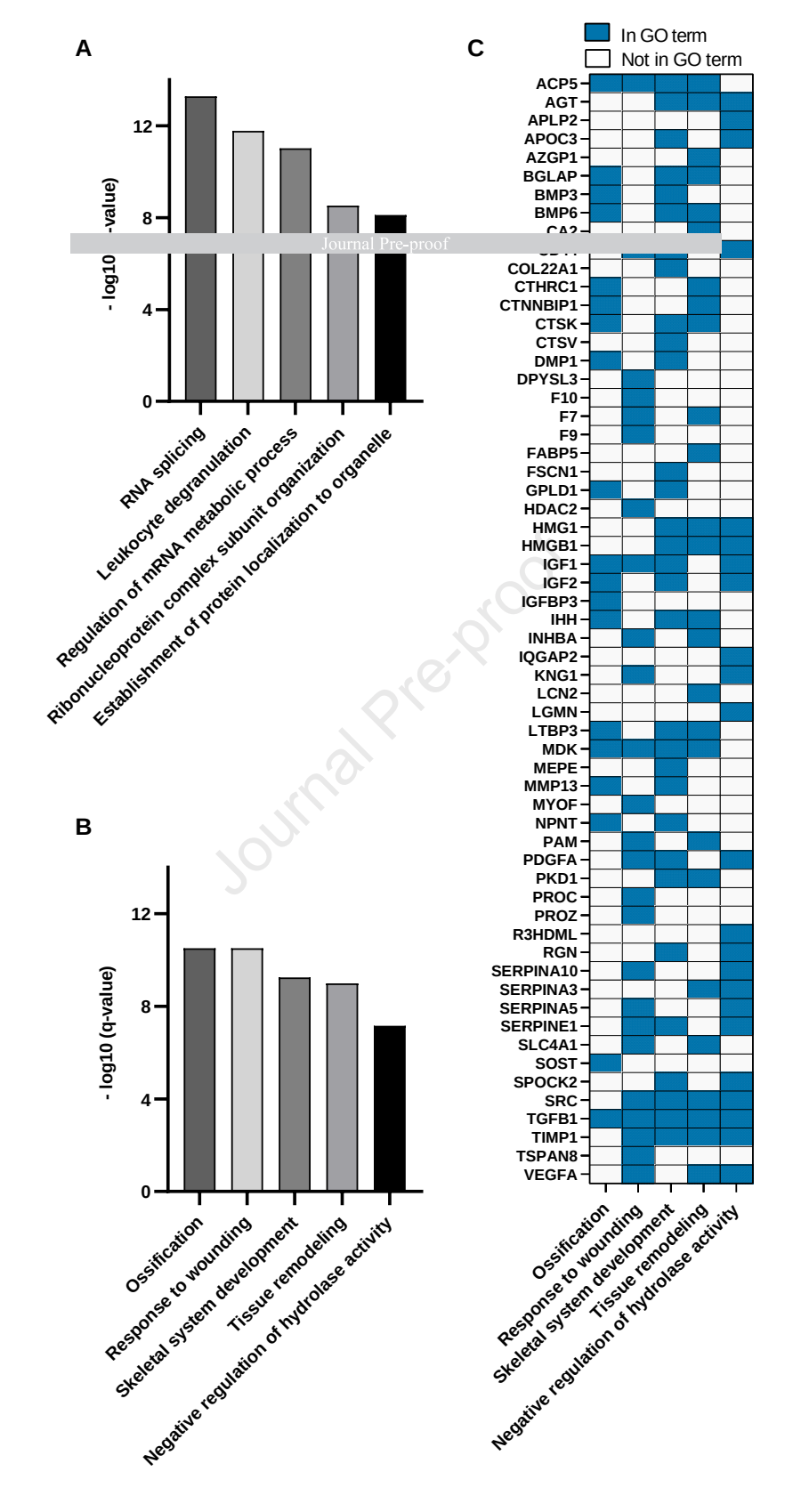


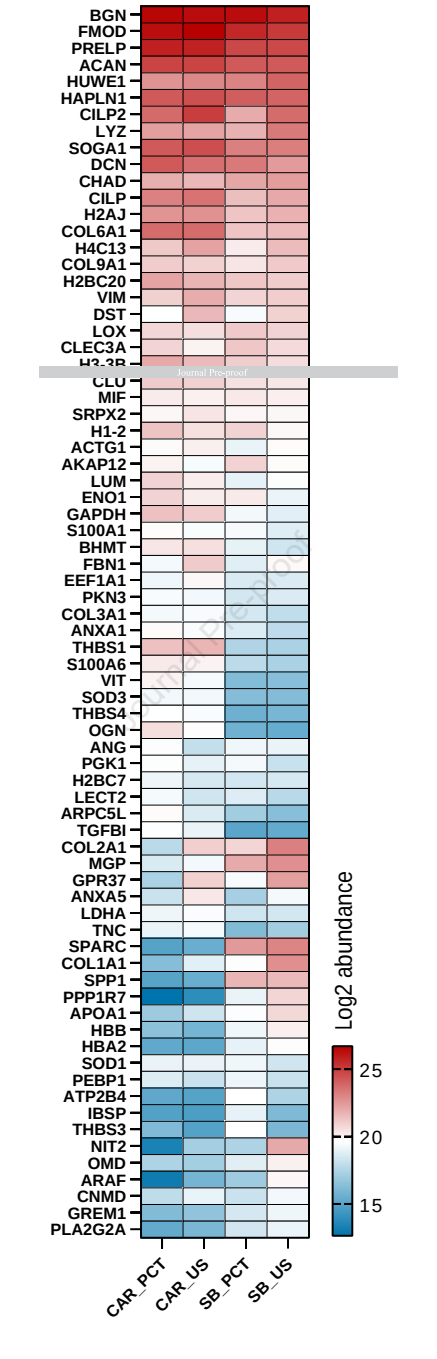
A) Sampling from three horses

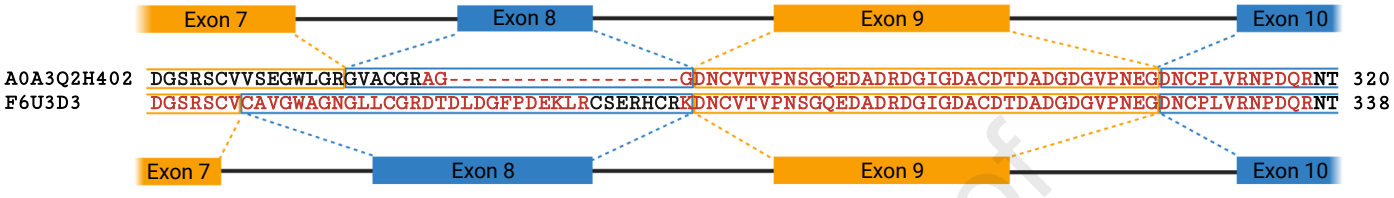
B) Sample processing in triplicate

C) DIA-MS analysis





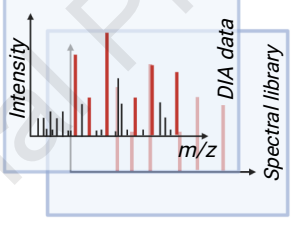

BGN
FMOD
PRELP
ACAN
HUWE1
HAPLN1
CILP2
LYZ
SOGA1
DCN
CHAD
CILP
H2AJ
COL6A1
H4C13
COL9A1
H2BC20
VIM
DST
LOX
CLEC3A
H3-3B
CLU
MIF
SRPX2
H1-2
ACTG1
AKAP12
LUM
ENO1
GAPDH
S100A1
BHMT
FBN1
EEF1A1
PKN3
COL3A1
ANXA1
THBS1
S100A6
VIT
SOD3
THBS4
OGN
ANG
PGK1
H2BC7
LECT2
ARPC5L
TGFBI
COL2A1
MGP
GPR37
ANXA5
LDHA
TNC
SPARC
COL1A1
SPP1
PPP1R7
APOA1
HBB
HBA2
SOD1
PEBP1
ATP2B4
IBSP
THBS3
NIT2
OMD
ARAF
CNMD
GREM1
PLA2G2A
CAR_PCT
CAR_US
SB_PCT
SB_US
Log2 abundance
25
20
15
Journal Preproof



COMP

.....
DNCVTVPNSGQEDADRDGIGDACDTDADGDGVPNEGDNCPLVRNPDQR
AGGDNCVTVPNSGQEDADRDGIGDACDTDADGDGVPNEGDNCPLVRNPDQR
AGGDNCVTVPNSGQEDADRDGIGDACDTDADGDGVPNEGDNCPLVR
KDNCVTVPNSGQEDADRDGIGDACDTDADGDGVPNEGDNCPLVR
SCVCAVGWAGNGLLCGRDTDLDGFPDEKLR
SCVCAVGWAGNGLLCGRDTDLDGFPDEK
DGIGDACDTDADGDGVPNEGDNCPLVR
DTDLDGFPDEK
SCVCAVGWAGNGLLCGR
DNCVTVPNSGQEDADR
DGSRSCVCAVGWAGNGLLCGRDTDLDGFPDEK
DTDLDGFPDEKLR
.....

A)



B)

.....
EQSSTVLTGLRPGVEYTVHVWAEK
NLVTDLVTENTATVSWDPVQAVIDR
APTDIDSPQNLVTNHVTENTAAVSWDPVQAVIDR
APTEIDSPQNLVTNQVTENTATVSWDPVQADIDR
APTEIDSPQNLVTDR
.....

TNN