Proteomic profiling of osteoarthritis. A computational approach to biomarker discovery.

Rydén, Martin

2023

Document Version:
Publisher's PDF, also known as Version of record

Proteomic profiling of osteoarthritis

A computational approach to biomarker discovery

MARTIN RYDÉN
DEPARTMENT OF CLINICAL SCIENCES | FACULTY OF MEDICINE | LUND UNIVERSITY

Proteomic profiling of osteoarthritis:
A computational approach to biomarker discovery

Proteomic profiling of osteoarthritis

A computational approach to biomarker discovery

Martin Rydén

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University to be publicly defended on 19th of October at 09.00 in Segerfalk Hall, BMC A10, Lund

Faculty opponent
Mohit Kapoor

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation **Date of issue:** 2023-09-27

Author: Martin Rydén **Sponsoring organization:**

Title and subtitle: Proteomic profiling of osteoarthritis: a computational approach to biomarker discovery

Abstract:

Understanding the molecular mechanisms of osteoarthritis (OA) is critical for early diagnosis and effective treatment. OA is a leading cause of disability and poses an increasing burden on healthcare systems, particularly with an aging global population. Despite the potential of proteomics to elucidate the complex biology underlying OA, its application remains limited due to challenges including gaps in computational tools and insights into the early stages of the disease. This thesis addresses these limitations through approaches that combines inventive computational strategies with biological exploration. The thesis presents a comprehensive computational framework that aims to facilitate proteomics analyses. In **paper I**, we developed *ProteoMill*, a user-friendly, web-based platform designed to make proteomics data analysis and biological interpretation available to a broader scientific community. In **paper II**, we introduced, *proteasy*, a specialized computational tool aimed at identifying proteolytic events. Using this tool, we performed a peptidomic analysis to identify key proteolytic enzymes involved in the degradation of proteins that contribute to OA progression in human synovial fluid (SF). This study presented a broad array of differentially abundant endogenously cleaved peptides, and their potential cleaving actor. We demonstrated that the proteolytic activity of the predicted proteases extends beyond the extracellular matrix (ECM) of the surrounding tissues, and can also affect factors such as chylomicron assembly, potentially leading to hampered homeostasis. In **paper III**, we established a human meniscus *ex vivo* model, that enabled us to perform controlled studies on cytokine-mediated effects on meniscal tissues. Our analyses highlight an increase in catabolic processes in response to some of the cytokine treatments while IL1ß had a limited catabolic effect. Finally, in **paper IV** we utilized the SOMAscan assay, an aptamer-based proteomics platform that is capable of measuring 7,000 proteins. This allowed us to get an unprecedented look into early-stage OA. Gaussian Graphical Models (GGMs) were further utilized to elucidate complex protein interactions, revealing new insights into disrupted joint homeostasis in OA. Through this, we identified novel proteins and sub-networks implicated in the early stages of the disease.

Key words: Bioinformatics, proteomics, knee osteoarthritis

Classification system and/or index terms (if any) Supplementary bibliographical information

Language: English **ISSN and key title**: 1652-8220

ISBN: 978-91-8021-461-2

Recipient's notes **Number of pages**: 76

Price Security classification

Signature  Date 2023-09-06

Proteomic profiling of osteoarthritis

A computational approach to biomarker discovery

Martin Rydén

LUND
UNIVERSITY

Cover by Victoria Udy

Figures 1, 2, 5, and 6 were created with Biorender.com

Faculty of Medicine

Department of Clinical Sciences

ISBN 978-91-8021-461-2

ISSN 1652-8220

Lund University, Faculty of Medicine Doctoral Dissertation Series 2023:120

Printed in Sweden by Media-Tryck, Lund University

Lund 2023

To my family

Table of Contents

1 Thesis at a glance

Osteoarthritis (OA) is more than just a medical condition; it's a life-altering disease that affects millions globally, causing pain, reduced mobility, and a decrease in quality of life. The economic toll is substantial, with escalating healthcare costs due to frequent medical visits, long-term medication, and surgical interventions like joint replacements. These costs are compounded by lost productivity and the emotional toll on patients and their families.

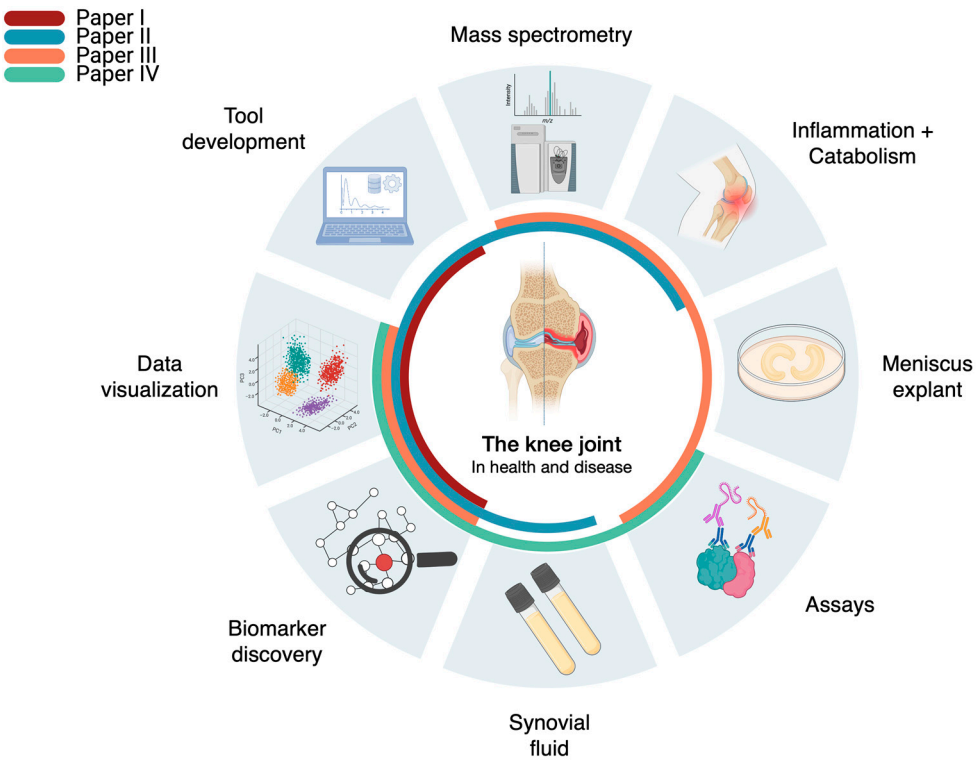

Figure 1. A visual synopsis of key themes and papers of this thesis.

In this context, this thesis focuses on addressing the challenge of understanding OA through the development and application of proteomic analysis methodologies (Figure 1). The work introduces two computational tools: the *ProteoMill* portal, an

integrated platform for omics data analysis, and *proteasy*, a specialized tool for mapping proteolytic events. These tools aim to facilitate more efficient and accurate analyses, addressing critical gaps in the field.

The works of this thesis also include a series of studies that delve into the molecular mechanisms of OA. In paper II we perform a peptidomic analysis of synovial fluid (SF) from OA patients, identifying key actors involved in enzymatic degradation that may contribute to the disease's progression. In paper III we establish a human meniscus *ex vivo* model to explore the effects of cytokine treatments on meniscal tissues, revealing increased catabolic processes that may contribute to OA development. Paper IV employs a comprehensive proteomic platform to analyze SF in the early stage of OA, identifying a range of proteins and sub-networks that are implicated in the early pathogenesis of the disease. This paper also introduces Gaussian Graphical Models (GGMs) to analyze protein interactions, offering novel insights into disrupted joint homeostasis in OA.

Collectively, the thesis combines a robust combination of practical tools and novel insights that enriches our understanding of OA, and provides a foundation for future research in the field.

2 Populärvetenskaplig sammanfattning

Knäartros är en av de vanligaste formerna av ledsjukdom och drabbar varje år miljontals människor världen över. Det är en degenerativ sjukdom som oftast drabbar äldre, och kännetecknas av nedbrytning av brosk i leden, vilket leder till smärta, svullnad och minskad rörlighet. Sjukdomen är inte bara en medicinsk utmaning utan innebär också en omfattande socioekonomisk belastning. Den leder ofta till långvarig arbetsfrånvaro och höga vårdkostnader, vilket påverkar både den enskilda individen och samhället i stort.

Trots sin omfattning är de underliggande molekylära mekanismerna som driver utvecklingen av artros delvis okända, vilket också gör att effektiva behandlingsalternativ saknas. För att förstå mer om knäartros har jag i mitt doktorandprojekt använt mig av tekniker inom området proteomik. Detta är studien av alla proteiner som produceras vid en viss tidpunkt, och masspektrometri-baserad proteomik är en teknik som tillåter oss att identifiera och kvantifiera tusentals proteiner samtidigt.

Bioinformatik är viktig komponent i detta arbete. Det är en disciplin som kombinerar biologi, dataanalys och statistik för att tolka stora mängder biologiska data. I mitt första delarbete har jag utvecklat ett webb-baserat verktyg som är särskilt applicerbart vid proteomikstudier i explorativ fas. Detta verktyg gör det möjligt för forskare att snabbt och intuitivt analysera komplexa biologiska system, utan att kräva programmeringserfarenhet.

I de efterföljande delarbetena har jag fokuserat på att använda dessa tekniker för att utforska de molekylära mekanismerna vid knäartros. Genom att jämföra proteomikdata från individer med och utan knäartros har jag identifierat proteiner och peptider som är över- eller underrepresenterade hos de drabbade individerna. Dessa molekyler kan vara inblandade i sjukdomens utveckling och kan potentiellt vara mål för nya behandlingsstrategier.

Sammanfattningsvis kan mitt arbete bidra till en djupare förståelse för de molekylära mekanismerna bakom knäartros. Detta kan i sin tur leda till utveckling av effektivare behandlingsmetoder och bidra till att förbättra livskvaliteten för de miljontals människor som lever med denna sjukdom.

3 List of Papers

3.1 Included in thesis

Paper I

Rydén, M., Englund, M. & Ali, N. ProteoMill: Efficient network-based functional analysis portal for proteomics data. Bioinformatics btab373- (2021) doi:10.1093/bioinformatics/btab373.

Paper II

Rydén, M. *et al*. Identification and quantification of degradome components in human synovial fluid reveals an increased proteolytic activity in knee osteoarthritis patients vs controls. PROTEOMICS e2300040 (2023) doi:10.1002/pmic.202300040.

Paper III

Rydén, M. *et al*. A human meniscus explant model for studying early events in osteoarthritis development by proteomics. J. Orthop. Res. (2023) doi:10.1002/jor.25633.

Paper IV

Rydén, M., Sjögren A. *et al.* Exploring the early molecular pathogenesis of osteoarthritis using differential network analysis of human synovial fluid. Manuscript submitted to Molecular and Cellular Proteomics.

3.2 Not included in thesis

Original papers

Folkesson E, Turkiewicz A, Ali N, **Rydén M**, Hughes HV, Tjörnstrand J, et al. Proteomic comparison of osteoarthritic and reference human menisci using data-independent acquisition mass spectrometry. Osteoarthr Cartilage. 2020;28: 1092–1101. doi:10.1016/j.joca.2020.05.001

Folkesson E, Turkiewicz A, **Rydén M**, Hughes HV, Ali N, Tjörnstrand J, et al. Proteomic characterization of the normal human medial meniscus body using data-independent acquisition mass spectrometry. J Orthop Res. 2020. doi:10.1002/jor.24602

Black RM, Wang Y, Struglics A, Lorenzo P, Tillgren V, **Rydén M**, et al. Proteomic analysis reveals dexamethasone rescues matrix breakdown but not anabolic dysregulation in a cartilage injury model. Osteoarthr Cartil Open. 2020;2: 100099. doi:10.1016/j.ocarto.2020.100099

Ortenlöf N, Vallius S, Karlsson H, Ekström C, Kristiansson A, Holmqvist B, Göransson O, Vaváková M, **Rydén M**, Carey G, Barton N, Ley D, Gram M. Characterization of choroid plexus in the preterm rabbit pup following subcutaneous administration of recombinant human IGF-1/IGFBP-3. Fluids Barriers CNS. 2023;20: 59. doi:10.1186/s12987-023-00460-1

Book chapter

Rydén M, Önnerfjord P. Electromechanobiology of Cartilage and Osteoarthritis, A Tribute to Alan Grodzinsky on his 75th Birthday. Adv Exp Med Biol. 2023;1402: 57–68. doi:10.1007/978-3-031-25588-5_4

4 Abbreviations

A1CF	APOBEC1 complementation factor
A2M	Alpha-2-macroglobulin
ACAN	Aggrecan core protein
ACBD6	Acyl-CoA-binding domain-containing protein 6
ACL	Anterior cruciate ligament
ACTB	Actin
ACTN2	Alpha-actinin-2
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
ADH1C	Alcohol dehydrogenase 1C
AHSG	Alpha-2-HS-glycoprotein
ALB	Albumin
APOA1	Apolipoprotein A-I
APOB	Apolipoprotein B
AWS	Amazon Web Services
BIC	Bayesian Information Criterion
C3	Complement C3
CFB	Complement factor B
CLUAP1	Clusterin-associated protein 1
COL1A1	Collagen alpha-1(I) chain
COL2A1	Collagen alpha-1(II) chain
COMP	Cartilage oligomeric matrix protein
CRP	C-reactive protein
CRTAC1	Cartilage acidic protein 1
CXCL7	Platelet basic protein
DCN	Decorin
DDA	Data-dependent acquisition
DHX8	ATP-dependent RNA helicase DHX8
DIA	Data-independent acquisition
DNAJB2	DnaJ homolog subfamily B member 2
ECM	Extracellular matrix
ELMO1	Engulfment and cell motility protein 1
FBLN5	Fibulin-5
FGA	Fibrinogen alpha chain
FN1	Fibronectin
GGM	Gaussian Graphical Model

GPD1	Glycerol-3-phosphate dehydrogenase 1-like protein
H1-10	Histone H1.10
HIBCH	3-hydroxyisobutyryl-CoA hydrolase, mitochondrial
HRG	Histidine-rich glycoprotein
IL	Interleukin
ITIH1	Inter-alpha-trypsin inhibitor
ITIH2	Inter-alpha-trypsin inhibitor
JAK-STAT	Janus kinase-signal transducer and activator of transcription
KNG1	Kininogen-1
LC-MS	Liquid chromatography–mass spectrometry
m/z	Mass-to-charge ratio
MAPK	Mitogen-activated protein kinase
MAR	Missing at Random
MCAR	Missing Completely at Random
MMP	Matrix metalloproteinase
MMP1	Interstitial collagenase
MNAR	Missing Not at Random
MVK	Mevalonate kinase
MYOC	Myocilin
NF-kB	Nuclear factor-kappa B
NMT1	Glycylpeptide N-tetradecanoyltransferase 1
OA	Osteoarthritis
OATP	Organic anion transporting polypeptide
OSM	Oncostatin M
PADI4	Protein-arginine deiminase type-4
PCA	Principal Component Analysis
PEBP1	Phosphatidylethanolamine-binding protein 1
PGS2	Decorin
PHF3	PHD finger protein 3
PLG	Plasminogen
PPBP	Platelet basic protein
PTM	Post-translational modification
PTPN11	Self-ligand receptor of the signaling lymphocytic activation molecule family member 5
SAA1	Serum amyloid A-1 protein
SERPINA3	Alpha-1-antichymotrypsin
SF	Synovial fluid
sIL6R	Soluble IL6 receptor
SLCO5A1	Solute carrier organic anion transporter family member 5A1
SOCS3	Suppressor of cytokine signaling 3
SOD3	Extracellular superoxide dismutase
SWATH	Sequential windowed acquisition of all theoretical fragment ion spectra

TAILS	Terminal amine isotopic labeling of substrates
TCN2	Transcobalamin-2
TNF	Tissue necrosis factor
TPM3	Tropomyosin alpha-3 chain

5 Introduction

5.1 Bioinformatics

5.1.1 Bridging the gaps

Bioinformatics stands at a crossroads of biology, computer science, statistics, and engineering, and serves as a bridge between these disciplines. Bioinformaticians can be in a sense be thought of as *translators*, working to establish common data standards and formats, converting raw biological data into meaningful insights, and communicating concepts between researchers from different fields.

A more traditional definition of bioinformatics is as the discipline dedicated to leveraging computational resources to interpret biological data[1]. This broad definition leaves room for multiple interpretations and ambitions for the field, but also underscores its multifaceted nature. In most definitions, bioinformatics also involves the problems of designing and sharing software, storing, and organizing data in databases, and focus on conducting reproducible experiments[2–4].

5.1.2 Reproducible research

Reproducible research is the principle that scientific results should be independently verifiable by other researchers using the same data and methods[5]. Reproducibility is important for several reasons: it increases the trust in scientific findings, it enables the verification and validation of results, it facilitates the comparison of results between studies, it encourages reuse and extension of existing data, and it promotes transparency and accountability of one's research[4,6].

Bioinformaticians play an important role in ensuring the reproducibility of research by developing and implementing standardized protocols, workflows, and data management practices, and they create robust computational tools and pipelines that can be used by other researchers to reproduce analyses and validate results[7–10]. Ensuring that data is annotated consistently necessitates standardized vocabularies for describing biological entities (such as genes or proteins) and their relationships. Here, ontologies play an important role, as formal representations of the concepts and relationships in a domain of knowledge[11]. For example, the Gene Ontology (GO)[12] provides a controlled vocabulary for describing the molecular function,

biological process and cellular component of gene products across different species. The Proteomics Standards Initiative (PSI)[13] develops ontologies for various aspects of proteomics, such as protein identification, quantification, interaction and modification. Ontologies help improve reproducibility by providing a common vocabulary and knowledge integration for bioinformatics data and methods[14]. This way, ontologies can help standardize the terminology, format, metadata, and the annotation of bioinformatics data across different platforms and databases.

Technologies like containerization and hashing enhance reproducibility by providing consistent computational environments and validating data identity[15]. Efficient data storage solutions preserve and share data, aiding in bioinformatics data management. Containers provide consistent and controlled environments for running bioinformatics software tools and workflows across different platforms and systems. Moreover, containers can help to simplify the installation, configuration, deployment, execution, and sharing of bioinformatics software tools and workflows.[16] Hashing can help to validate the storage, retrieval, comparison, and sharing of bioinformatics data and software[17]. One example of a hash function commonly used in bioinformatics is MD5[18], which is a hash function that generates a 128-bit hash from any input data. In ProteoMill (paper I)[19], MD5 is used to keep track of project settings, and allowing users to generate a "reproducibility token".

5.1.3 Exploratory analysis of the proteome

5.1.3.1 Functional analysis of omics data

Appending "-*omics*" to a biological term implies the comprehensive assessment of a set of biological components. Similarly, the "-*ome*" suffix represents the complete set of that biological entity[20]. In functional analyses of omics data, we aim to understand the full picture: the complete set of genes or gene products, expanding the scope from studying individual genes, transcripts, or proteins to a system-wide approach[21]. In proteomics, functional analysis is often used to describe the molecular function, biological process and cellular component of sets of proteins[12], and to characterize their involvement in the pathways and molecular processes that are associated with the pathogenesis of the disease (Figure 2)[22,23].

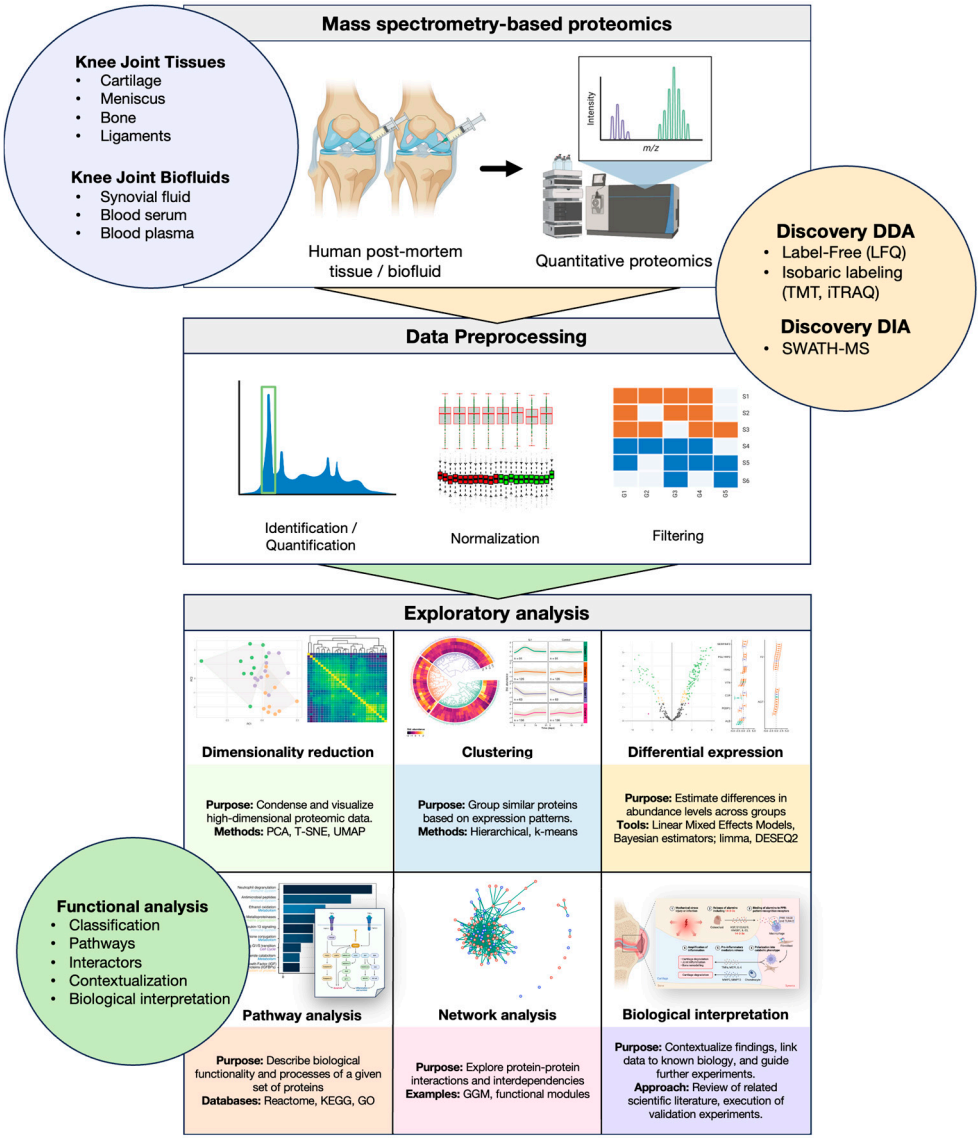

Figure 2. From knee joint sample to interpretation of exploratory analysis results.

5.1.3.2 Normalization, scaling, and batch effect

Normalization is a process of adjusting the peptide abundance values in proteomics data to reduce the technical variation and to make the samples more comparable[24]. The benefit of improving the sensitivity and accuracy of downstream analysis should be weighed against the risk of introducing new biases or removing biologically relevant variation. The idea of normalization, in other words, is to

ensure that the observed variation is due to the biological differences and not because of the experimental procedure or technical artifacts.

Standardization is the process of transforming data to a common scale. This is useful when the data from different experiments or instruments have different units.[25] One commonly used method for standardization is the Z-score[26]. It is calculated by taking the difference between a value and the mean of all values, and then dividing by the standard deviation of all values, as expressed in the following formula:

$$Z = \frac{x - \mu}{\sigma}$$

For example, if one instrument measures protein abundance in arbitrary units and another in intensity counts, standardization can transform these values, so they are on a common scale. The goal is essentially to ensure that the data from various sources or experiments can be combined or compared without introducing biases.

Batch effects are systematic non-biological differences between or groups of samples in an experiment. These could be introduced during sample collection, preparation, or data acquisition, for instance if samples are processed on different days or by different instruments.[27]

5.1.3.3 Dimensionality reduction

Dimensionality reduction techniques are methods that reduce the number of variables or features in a high-dimensional data set, while preserving the essential information. These techniques are useful for handling problems when analyzing proteomics data, such as noise, redundancy, multicollinearity, and computational complexity.[28] One commonly used method is Principal Component Analysis (PCA). PCA transforms the original variables into a set of linearly uncorrelated variables called principal components, which capture the maximum variance of the data. PCA can be used to visualize the structure and variability of omics data, identify outliers and clusters, and reduce the dimensionality for downstream analysis.[29]

5.1.3.4 Clustering

Cluster analysis is a data exploration technique that aims to group data based on their similarity[30]. It can for example be performed on proteins, samples, or time variables[31]. Within these clusters, proteins can then be functionally annotated through the principle of "guilt by association" - which assumes that proteins in the same cluster are likely to share functional characteristics[32–34]. Cluster analysis can help us to discover patterns, trends, and relationships in the data, and to evaluate the effects of different factors or interventions. Most clustering algorithms used in analysis of high-throughput biological data are distance-based, having the advantages of being relatively simple to implement, but may be sensitive to noise and outliers[35–37].

Hierarchical clustering is a method that builds a hierarchy of clusters by either successively merging or dividing groups of data points. In the agglomerative approach, the algorithm starts with each data point as a separate cluster and then merges the closest clusters into larger ones. Alternatively, in the divisive approach, it starts with one large cluster and divides it based on some criterion. The process continues until only a single cluster remains or some other stopping criterion is met. One of the advantages of hierarchical clustering is that it doesn't require the number of clusters to be specified *a priori*. Another advantage is that hierarchical clustering can be visualized as a dendrogram.[38–42]

k-means clustering is a partitioning method that divides a dataset into k distinct, non-overlapping subsets, or clusters. The algorithm starts by choosing k initial centroids, either randomly or based on some criterion. Each data point is then assigned to the nearest centroid, becoming a member of that cluster. The centroids are updated as the mean of all the data points in each cluster, and the process is repeated until the centroids no longer change.[30,43–46]

Hierarchical clustering and k-means commonly use Euclidean distance or other simple metrics that do not account for temporal alignment[47–51]. In time series clustering, our aim is to identify groups of proteins with similar expression patterns over time[49]. Time series clustering assigns data points, protein abundances, to groups or clusters based on patterns and trends in the data. This may help us to identify groups of proteins that have similar characteristics and to evaluate an effect over time.

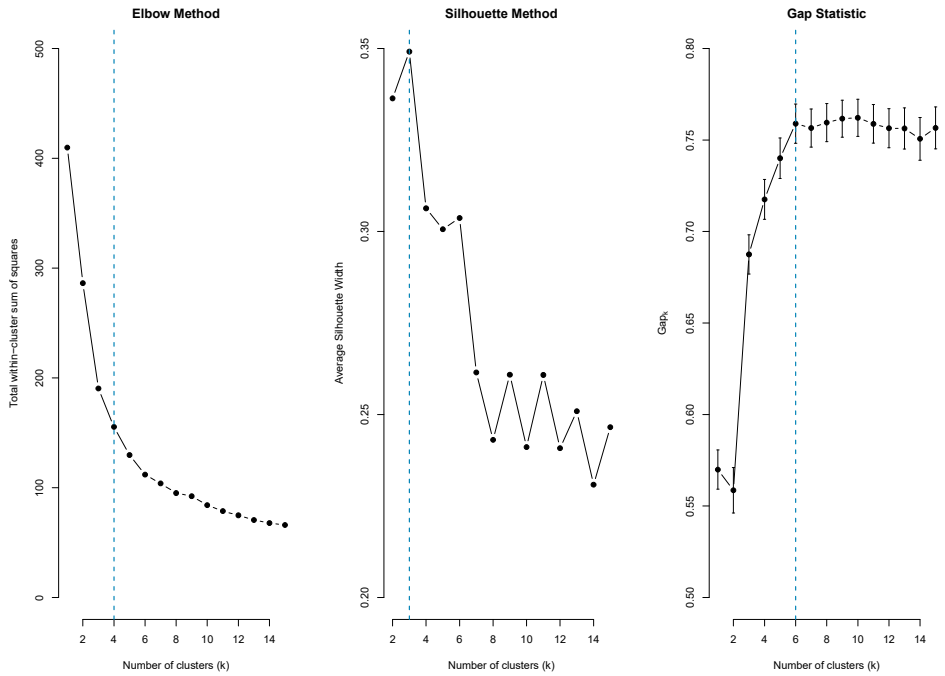

Figure 3. Comparison of three methods for estimating the optimal number of clusters (k) in cytokine-treated (IL1) versus control meniscus explants over multiple timepoints. The methods include the Elbow method (k=4), Silhouette method (k=3), and Gap statistic (k=6). The vertical blue dashed line in each subplot indicates the optimal k value as per the respective method.

Simply put, the objective of clustering is to form clusters wherein proteins have maximum similarity with other proteins within the cluster, but minimum similarity with proteins in other clusters. Based on these similarities, we must find a way to decide how many clusters the proteins should be divided into. There is no standard approach to this problem, and finding the appropriate number of clusters to be generated is subjective[52]. One commonly used method is the "elbow method", which involves plotting the relationship between the number of clusters and Within Sum of Squares (WSS) or the total variance explained by the clusters (Figure 3)[53]. The optimal number of clusters is the value of *k* at the "elbow" point of the plot, where the change in WSS begins to level off. Another is to use statistical methods such as "gap statistic" to estimate the number of clusters by comparing intra-cluster variation for different numbers of clusters, selecting aggregations that are the least uniformly distributed (Figure 3)[54]. While these methods can offer a standardized approach to finding the, these methods are heuristics and do not guarantee us finding the optimum. It is therefore a good idea to try multiple methods and then comparing the results. Perhaps the most useful approach is by visual methods like generating and visually inspect a dendrogram using hierarchical clustering (Figure 4), to see if there is a clear partitioning of branches.

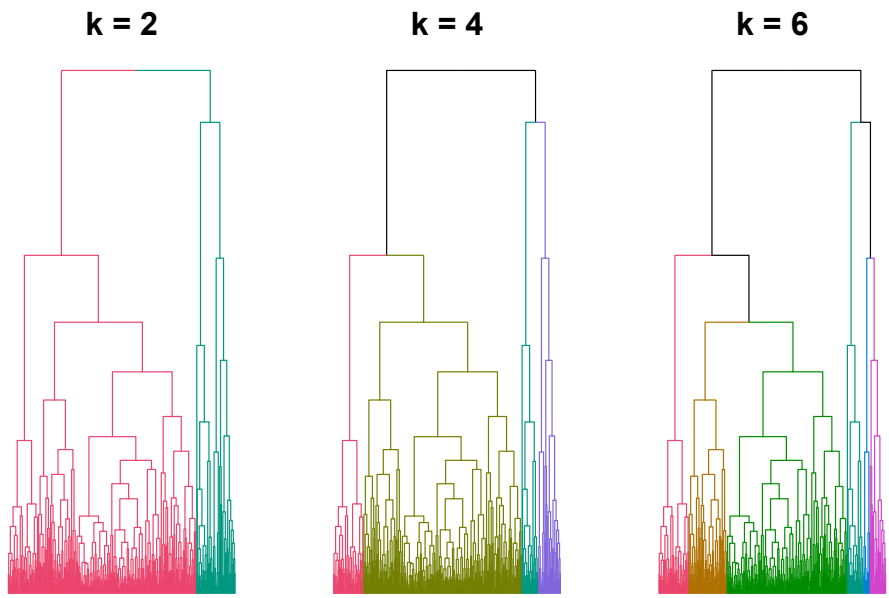

Figure 4: Dendrograms representing hierarchical clustering of proteins in meniscus explants subjected to IL1 and control treatments over multiple timepoints. The dendrogram is color-coded to indicate clusters formed at different numbers of clusters (k = 2, k = 4, k = 6). This offers a visual approach to assess the quality of the clustering results.

5.1.3.5 Missing data

In proteomics studies, it is common to encounter missing values, i.e. instances where a protein's abundance is not recorded in a given sample[55]. While statistical methods exist for handling missing data, their validity and reliability often depend on understanding the source of missingness. In many practical scenarios involving liquid chromatography–mass spectrometry (LC-MS) data, this source is not determinable, making traditional methods questionable and necessitating a careful consideration of alternative approaches.

Missing data in statistical analyses are commonly categorized as Missing Completely at Random (MCAR), Missing at Random (MAR), or Missing Not at Random (MNAR)[56]. The strategy for handling missing data often depends on the missingness mechanism. However, when the source of missingness is unknown, there is a risk of introducing bias by imputation[57,58]. It is important to remember that imputation cannot add new information to the data[59]. As stated by Dempster and Rubin (1983)[60]:

> *"The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to the real and imputed data have substantial bias."*

In the context of statistical modelling, data is considered informative when it contributes to the precision or validity of the model estimates. Data with extensive missing values can be less informative and may introduce ambiguity in the interpretation of the model. For LC-MS data, where the source of missingness is often unclear, it becomes challenging to determine when data should be considered informative enough for inclusion in the model. Simply removing proteins with missing values can result in the loss of potentially meaningful information. On the other hand, including proteins with frequent missing values might introduce noise or bias, thereby reducing the overall informativeness of the data for the intended analyses.

5.1.3.6 Differential abundance analysis

Proteomics experiments often rely on estimating a protein's abundance in one condition relative to another to understand biological processes or disease conditions[61]. The complexity of this analysis often arises from the need to account for multiple sources of variability, such as biological replicates and technical noise.

Linear mixed effects models have become increasingly useful in this context, as they can model both fixed effects, which are consistent across all observations, and random effects, which are specific to particular subgroups or conditions[62,63]. This flexibility allows for more accurate estimation of protein or peptide abundance changes across different conditions or time points[61,64–66].

The use of linear mixed effects models is especially beneficial when the data structure is hierarchical or when observations are correlated. These models provide a way to partition the observed variance into its different components, thereby offering a nuanced understanding of the factors contributing to differential abundance.

The R package limma has been widely adopted for differential abundance analysis in proteomics[67,68]. It was originally designed for microarray data and uses an empirical Bayes approach to shrink the sample variance towards a pooled estimate, increasing the stability of the inference when the number of samples is small[68]. While limma is not inherently a mixed effects model, it has gained popularity for use in proteomics data analysis due to its robustness and computational efficiency.

5.1.3.7 Pathways

Biological pathways are series of interactions among molecules in a cell that lead to a certain product or a change in the cell. Pathways are involved in various biological processes, such as metabolism, gene regulation and signal transduction, and analysis of pathways is important in understanding the mechanisms of diseases.[69–71]

Pathway analysis is a technique to identify pathways that are altered in a given proteomics dataset, such as proteins that are differentially expressed between two conditions. Pathway analysis can help interpret the biological meaning and implications of proteomics data, as well as discover novel pathways or interactions that are relevant to the experimental context.[72,73]

There are different methods and tools for performing pathway analysis of proteomics data. Some common methods include over-representation analysis, gene set enrichment analysis and topology-based analysis[74]. These methods use different statistical tests and assumptions to evaluate the enrichment of pathways in a proteomics dataset. Pathway analysis tools use different databases and formats of pathway gene sets, as well as different visualization and reporting features.[75,76]

Gene set analysis methods are easy to implement and widely used, but they ignore the structure and interactions of the pathways, which may limit their biological relevance and accuracy. Topology-based analysis methods incorporate the pathway topology or network information into the analysis, and may use mathematical models to score and rank the pathways based on their impact or deregulation in the proteomics dataset[23]. Topology-based analysis methods have the potential to better capture the complexity and dynamics of biological systems, but they also face several challenges, such as data integration, model selection, and result validation. Another challenge is the lack of consensus in the definition of a pathway[77].

Some pathways that have been identified to be associated with OA using functional analysis includes the MAPK[78,79], NF-kB[79–81], transforming growth factor-β (TGFβ)[82,83], and Wnt[83,84] signaling pathways. MAPK induces chondrocyte differentiation, synthesis of matrix metalloproteases (MMPs) and production of pro-inflammatory cytokines[85].

5.1.3.8 Protein-protein interaction networks

Protein complexes are molecular machines by which many biological functions are conducted[86]. They are formed when two or more proteins bind together by protein–protein interactions (PPIs) to perform a function in a specific biological context[87,88]. PPIs play an important role in many cellular processes such as signal transduction, gene regulation, metabolic pathways, and disease mechanisms[89]. We study PPIs to understand the molecular functions of proteins, their interactions with other proteins and molecules, and their involvement in complex biological systems[90,91].

There are many different databases that store PPI data derived from experimental and computational methods[92,93]. Some of the popular databases for PPIs are STRING[94], BioGRID[95], IntAct[96], MINT[97]. These databases allow for searching, browsing, visualizing, analyzing, and downloading PPI data.

PPI networks are graphical representations of PPI data, where nodes represent proteins and edges represent interactions. To highlight certain proteins of particular interest, we can study the topological properties of the PPI network, such as degree distribution, clustering coefficient, betweenness centrality, and modularity.[92,98–100]

These approaches can also be used to identify functional modules or complexes of proteins that work together to perform a specific biological function or process. PPI networks can also be used to identify key proteins or "hubs" that have high connectivity or centrality in the network[101,102]. Hub proteins are often essential for the network integrity and functionality and may be involved in regulating cellular dynamics and responses, and tends to carry important biological information[103–105].

5.1.3.9 Correlation, co-expression, and conditional dependence

Correlation measures the strength and direction of a linear relationship between two variables, such as the expression levels of two proteins[106]. The most commonly used metric is Pearson's correlation coefficient, which ranges from -1 to 1. A value of 1 indicates a perfect positive correlation, meaning that the variables increase or decrease together. A value of -1 indicates a perfect negative correlation, meaning that one variable increases as the other decreases. A value of 0 indicates no linear correlation between the variables. However, Pearson's correlation assumes that the variables have a normal distribution and a linear relationship, which may not be true for omics data[*]. Therefore, robust correlation methods are sometimes preferable. One example is Spearman correlation, which ranks the values of the variables before calculating the correlation. This method is less sensitive to outliers and does not require a linear relationship.

Co-expression refers to the phenomenon where two or more proteins have similar expression patterns across different conditions or samples. Co-expression can imply a functional relationship between the proteins, such as being part of the same biological pathway, participating in the same cellular process, or being regulated by the same set of transcription factors[107–109]. Co-expression analysis can help identify novel gene functions, gene-gene interactions, and gene set topology.

Conditional dependence is a concept that extends beyond correlation and co-expression. It captures the dependence between two variables given the information

[*] Mass spectrometry-based proteomics data, like microarray-based transcriptomics data, tends to be approximately log-normally distributed[278]. RNA-seq data are discrete counts of reads that can be modelled by the Poisson distribution or its extensions, such as the negative binomial distribution or the generalized Poisson distribution[279,280].

from other variables. For example, two proteins may be correlated or co-expressed in general, but not when another protein is present or absent. This can indicate a more direct or specific interaction or association between the proteins.

One way to infer conditional dependencies from proteomic data is to use Gaussian Graphical Models (GGMs)[110]. They represent the variables as nodes and the conditional dependencies as edges in a network. While correlation matrices capture all pairwise linear associations, GGMs only retain the edges that indicate conditional dependencies among variables, given all other variables in the network. This makes the network sparser and more informative for inferring direct interactions or associations.

5.1.3.10 Visualization and interpretation

Proteomics studies produce massive, high-dimensional datasets that represent a snapshot of the cellular environment at a given time or condition. Visualizing and interpreting this data is both an art and a science, where the aim to extract meaningful information requires careful navigation. It has, for instance, been shown that nonsense data could yield "significant" biological predictions for canonical pathways, by chance[111]. There are a multitude of tools and pipelines for visualizing and interpreting omics data. A general assumption should be that all tools return their own outputs, which represent their frame of reference, but lack the perspective of other tools[112].

Transformation of data from numerical matrices into more tangible representations lets us comprehend and interpret our findings. While there can be countless variations of visual representations of omics data, some important principals for data visualization are to have simple figures with detailed captions, to include measures of uncertainty, and to include meaningful geometry and colors[113]. Visual representations, such as network graphs and heatmaps, can derive narratives about biological processes, molecular functions, and cellular dynamics. This, in turn, guides us to further investigations, and sparks new scientific inquiries.

5.2 Knee osteoarthritis

Knee osteoarthritis (OA) is a chronic joint disease that affects the whole joint, including the articular cartilage, subchondral bone, synovium, menisci, and ligaments[114]. It is characterized by joint pain, stiffness, swelling, crepitus, reduced range of motion, and functional impairment[115]. Knee OA is also associated with structural changes in the joint, such as cartilage loss, bone remodeling, osteophytes, subchondral cysts, and meniscal damage (Figure 5), which can be detected by radiographic or magnetic resonance imaging (MRI) techniques[116]. However, there is often a poor correlation between the clinical symptoms and the radiographic findings of knee OA[115,117]. The diagnosis of knee OA is based on a combination of history, physical examination, and imaging tests, depending on the context and purpose of the assessment[115]. The etiology of knee OA is multifactorial and involves both genetic and environmental factors that interact to initiate and progress the disease[114].

5.2.1 Prevalence and socio-economic burden

OA is the most common form of arthritis and one of the leading causes of chronic pain and disability worldwide[118,119]. According to the Global Burden of Disease Study, the number of prevalent cases of OA affected 528 million people in 2019, more than doubling from 248 million cases in 1990 and making it among the most prevalent diseases globally[120,121]. The prevalence of OA varies by geographical region, age, sex, and joint site. The highest prevalence of OA is found in Western Europe, North America, and Australasia, while the lowest prevalence is found in sub-Saharan Africa and South Asia[122]. The prevalence of OA increases with age, reaching a peak in the 55-64 years age group for both sexes[122]. Women have a higher prevalence of OA than men, especially after the age of 50 years[123,124]. The most commonly affected joints by OA are the knee, hip, and hand, with knee OA being the most prevalent and disabling form of OA[118,121,122]. In Sweden, the prevalence of OA was reported to be 26.6% among adults aged 45 years and older between, with a higher prevalence among women (30.5%) than men (22.4%)[125].

OA imposes a significant socio-economic burden on individuals, health systems, and society. The direct costs of OA include medical expenses for diagnosis, treatment, and management of the disease, such as consultations, medications, surgery, rehabilitation, and assistive devices[126]. The indirect costs of OA include productivity losses due to sick leave, early retirement, or disability[119].

5.2.2 Risk factors and symptoms

OA is a multifactorial disease that results from the interaction of genetic and environmental factors that affect the joint structure and function[114]. The most important person-level risk factors for OA are older age, female sex, obesity, genetics, and systemic diseases, such as diabetes, gout, or hemophilia [114,115,123,127–129].

Older age is associated with cumulative exposure to mechanical stress, oxidative damage, cellular senescence, and reduced regenerative capacity of the joint tissues[114,130]. Obesity is a major modifiable risk factor for OA, especially in the knee, as it increases the mechanical load on the joint and induces metabolic and inflammatory changes that affect the joint homeostasis[114,129]. However, OA in joints not typically exposed to mechanical loading, such as hand-OA, is more prevalent in obese individuals[131,132], suggesting OA is not only mechanically induced through obesity, but is also influenced by systemic factors (and could therefore be affected by metabolic diseases such as diabetes)[133]. Genetic factors can influence the risk of OA and may affect risk factors such as obesity and bone mass (although these mechanisms have not been well studied)[134]. While hip arthroplasty is strongly influenced by genetic factors, knee arthroplasty depends more on modifiable factors like BMI[135]. Several genes have been identified that influence the susceptibility and severity of OA, such as those involved in cartilage metabolism, bone formation, inflammation, and pain perception[129]. Systemic diseases can predispose to OA by altering the joint metabolism, immunity, or vascularization[114].

The most common joint-level risk factors for OA are joint injury, repetitive joint use, and joint malalignment[114]. Joint injury, which may account for 12% of all OA cases[136] can result from trauma or inflammation that disrupts the integrity and stability of the joint structures[114]. Joint malalignment can increase the stress concentration on certain areas of the joint and accelerate the cartilage wear and tear[114].

The main symptoms of OA are pain, stiffness, swelling, crepitus (i.e. the popping, clicking, or crackling sound and sensation that occurs in the joint), reduced range of motion, and functional impairment[115]. Pain is usually intermittent and worsened by activity or mechanical loading[123]. Functional impairment can affect the mobility, independence, and quality of life of people with OA. OA can also cause psychological distress, such as depression or anxiety, due to chronic pain and disability[115].

5.2.3 Anatomy of the knee joint

The knee joint is the largest and one of the most complex joints in the human body[137]. It consists of three bones: the femur, the tibia and the patella, which are connected by ligaments, tendons and muscles[137]. The joint surfaces are covered by articular cartilage, which has friction-reducing and load-bearing functions, offering

structural support and acts as a cushion between the bones to absorb shock[137,138]. Between the femur and tibia, there are two menisci, which are fibrocartilaginous discs that also act as shock absorbers and stabilizers of the joint[137,139]. The knee joint also contains a synovial membrane, which produces synovial fluid (SF) that lubricates and nourishes the joint tissues[137].

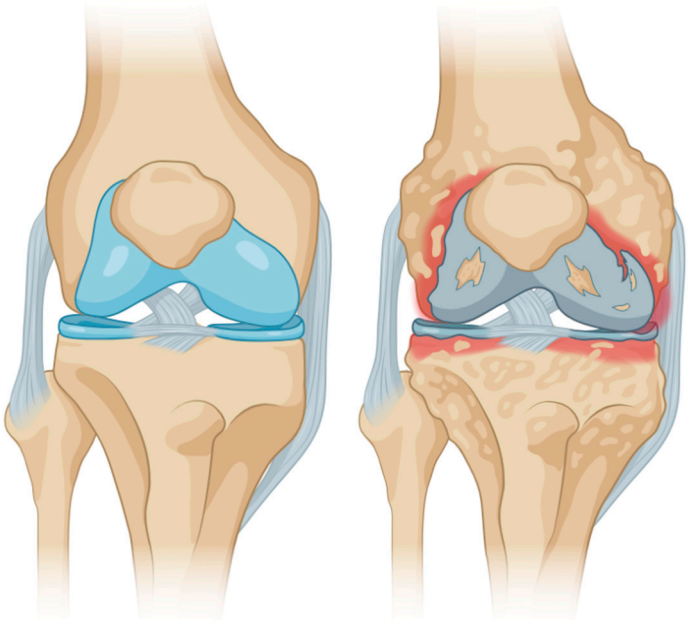

Figure 5. Illustration of a knee joint in health (left) and OA (right).

5.2.4 Pathogenesis

OA is a complex and heterogeneous disease that involves multiple biological processes and pathways that affect the whole joint[140]. The current understanding of OA pathogenesis is that it is initiated by a combination of mechanical and molecular factors that trigger a cascade of events that lead to joint damage and dysfunction[141]. The main events that occur during OA pathogenesis are inflammation, cartilage degradation, bone remodeling, synovial changes, and meniscal damage[140,142,143].

Inflammation is a key feature of OA that contributes to both the initiation and progression of the disease[136,144]. Inflammation can be triggered by mechanical stress, cartilage breakdown products, cytokines, chemokines, adipokines, or danger-associated molecular patterns (DAMPs) that activate the innate immune system and induce the production of pro-inflammatory mediators, such as interleukin-1 beta

(IL-1β), or tumor necrosis factor alpha (TNF-α).[144–146]. These mediators can further amplify the inflammatory response and cause pain, swelling, and joint damage[144].

Cartilage degradation is a hallmark of OA that results from an imbalance between the anabolic and catabolic activities of the chondrocytes, the cells that maintain the cartilage matrix[127,147]. Cartilage degradation can be caused by mechanical stress, inflammation, oxidative stress, or enzymatic cleavage by MMPs or aggrecanases[127,148]. Cartilage degradation leads to the loss of proteoglycans and collagen fibers that provide the cartilage with its resilience and strength[127]. Cartilage degradation also exposes the subchondral bone to increased mechanical load and stimulates bone remodeling[127].

Bone remodeling is another important process in OA pathogenesis that involves the formation and resorption of bone tissue by osteoblasts and osteoclasts, respectively[149,150]. Bone remodeling can be influenced by mechanical stress, inflammation, growth factors, hormones, or cytokines[149,151]. Bone remodeling can result in subchondral sclerosis, osteophyte formation, subchondral cysts, or bone marrow lesions (BMLs) that affect the joint structure and function[149]. Subchondral sclerosis is the thickening and hardening of the subchondral bone plate that reduces its shock-absorbing capacity and increases the stress on the cartilage[149]. Osteophytes are bony outgrowths at the joint margins that can limit the joint range of motion and cause pain or nerve compression[149]. Subchondral cysts are fluid-filled cavities in the subchondral bone that can cause pain or fracture[149]. BMLs are areas of increased signal intensity on MRI that reflect bone edema or microdamage that can be associated with pain or cartilage loss[149].

Synovial changes are also common in OA and involve the inflammation, hypertrophy, fibrosis, or angiogenesis of the synovial membrane that lines the joint cavity[144]. Synovial changes can be induced by mechanical stress, cartilage fragments, cytokines, chemokines, or growth factors that stimulate the synovial cells to produce pro-inflammatory mediators, such as IL-1β, TNF-α, IL-6, or prostaglandin E2 (PGE2)[144]. These mediators can cause synovitis, SF accumulation, or synovial hyperplasia that contribute to joint pain, stiffness, or damage[144].

Another common condition affecting the knee joint is meniscal injury, which can be classified into traumatic or degenerative types[152,153]. Traumatic tears occur when a normal meniscus is subjected to an abnormal force, usually during sports or accidents that involve twisting or pivoting of the knee[152,154,155]. Traumatic tears are more common in younger individuals and often involve other structures of the knee, such as ligaments or cartilage[152,154,155]. Degenerative lesions occur when a weakened meniscus is subjected to a normal or near-normal force, usually due to aging or underlying OA[152,154–156]. Degenerative lesions are more common in older individuals and often affect both knees[152,154–156]. Meniscal injuries can cause symptoms such as pain, locking, clicking or catching of the knee, and can increase the risk of developing or progressing OA[152,154–157].

5.3 Proteomics

Proteins are the main functional components of biological systems, as they carry out most of the biochemical reactions and interactions that underlie cellular processes and functions[158]. To understand the molecular mechanisms and pathways that govern life, we need to measure and characterize the proteins that are present in a given biological sample, such as a tissue, a cell, or a biofluid. This is the aim of proteomics, the large-scale study of the protein content and composition of a sample[159].

5.3.1 From genes to proteins

Proteins are encoded by genes, which are segments of deoxyribonucleic acid (DNA) that store the genetic information of an organism (Figure 6). DNA is organized into chromosomes in the nucleus of eukaryotic cells[160], and consists of a long chain of four types of nucleotides: adenine (A), thymine (T), cytosine (C), and guanine (G). The sequence of nucleotides in a gene determines the sequence of amino acids in a protein, according to the genetic code[161]. Amino acids are the building blocks of proteins, and there are 20 different types of amino acids that can be combined in various ways to form different proteins.

The process of protein expression involves two main steps - transcription and translation (Figure 6). Transcription involves the synthesis of RNA from DNA by RNA polymerases. Different types of RNA, like mRNA, tRNA, rRNA, and ncRNA, perform different functions in the cell[162]. mRNA transports genetic information from DNA to the ribosomes, where translation occurs. Ribosomes facilitate peptide bond formation between amino acids, while tRNAs recognize specific mRNA codons and add the corresponding amino acids to the growing peptide chain. As the ribosome moves along the mRNA, it matches each codon with its complementary anticodon on tRNA, and adds the amino acid carried by tRNA to the growing peptide chain[160].

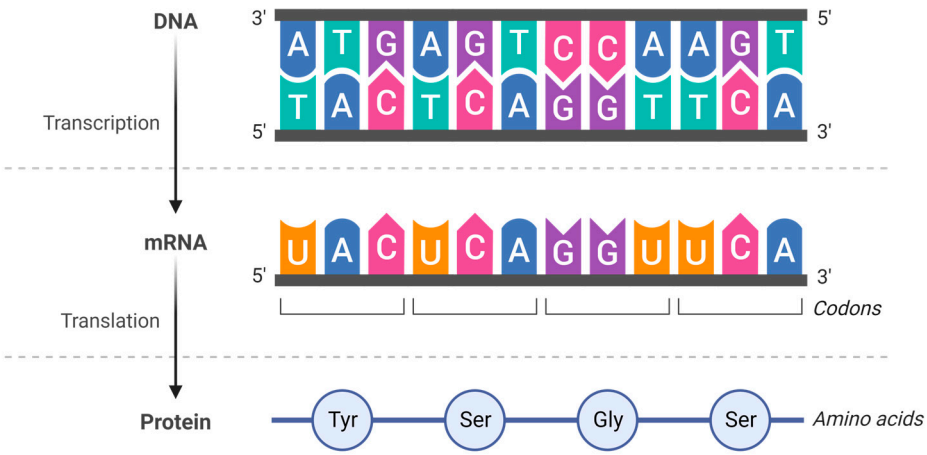

Figure 6. Central dogma of molecular biology.

5.3.2 Beyond translation

The peptide chain produced by translation is not yet a functional protein. It needs to undergo several modifications and interactions before it can perform its biological role. First, the peptide chain needs to fold into a specific three-dimensional shape that determines its function. The folding process is assisted by molecular chaperones, which are proteins that help other proteins to fold correctly and prevent aggregation[163]. Second, the peptide chain may need to be cleaved or spliced to remove or rearrange some segments. For example, many proteins have a signal peptide at their N-terminal end that directs them to their proper cellular location, such as the plasma membrane or extracellular environment. This signal peptide is usually removed after translation by signal peptidases[164]. Third, the peptide chain may need to be modified by adding or removing chemical groups or molecules[165]. These modifications are called post-translational modifications (PTMs), and they can affect the activity, localization, or interactions of proteins[166]. There are hundreds of different types of PTMs that have been identified in proteomics studies, such as phosphorylation, acetylation, methylation, glycosylation, ubiquitination, and proteolysis[165]. Fourth, the peptide chain may need to interact with other peptide chains or molecules to form complexes or assemblies[167]. Many proteins function as part of larger structures that consist of multiple subunits or domains that can be either identical or different.

5.3.3 Degradomics

The ECM is constantly remodeled by proteases, which are enzymes that cleave peptide bonds in proteins. In OA, dysregulated proteolysis destabilizes the structural integrity of the ECM and generates pro-catabolic fragments.

Degradomics is an emerging sub-discipline within proteomics that seeks to decode the interplay between proteases and their substrates, and understanding the catalytic mechanisms of proteases.

In peptidomics*, enrichment protocols designed to enhance the detectability of specific classes of peptides, whether they are biologically active fragments like hormones and cytokines or biomarker-type peptides indicative of disease states. Commonly used enrichment protocols for such studies include ultracentrifugation, filtration, precipitation or other type of small peptide fragment enrichment[168]. The enrichment techniques are typically designed to target peptides with specific physiochemical properties, and are not equally suitable for all peptide classes. These protocols enhance the detection of smaller peptides, generally enabling detection of up to a thousand peptides.

One such method is the terminal amine isotopic labeling of substrates (TAILS), which allows the selective capture and identification of N-terminal peptides. TAILS can reveal the endogenous peptidome and the proteolytic events in a given sample. For example, TAILS has been applied to investigate the degradome of knee OA cartilage[169].

In paper II, we used a protocol that eliminates enrichment selection of a specific subgroup of peptides. By studying the fragments/peptides of the of the proteins that remain in the pellet after ethanol precipitation, we were able to identify a larger number of peptides that was not cleaved by trypsin.

5.3.4 Challenges in proteomics

The complexity and diversity of proteins in biological samples pose challenges for proteomics. Proteins vary greatly in size, structure, charge, hydrophobicity, abundance, and modifications, and they can interact with each other and with other molecules in dynamic and context-dependent ways[170]. For example, in cartilage tissue, different types of collagens and proteoglycans have different molecular weights, structures, functions, and binding affinities[171]. Collagens and proteoglycans can also interact with each other and with other ECM components or

* Peptidomics and degradomics are closely related fields within proteomics. Degradomics focuses on the system-wide study of proteases, substrates, and their resulting cleavage fragments. It aims to understand how proteolysis contributes to physiological processes and diseases. Peptidomics focuses more broadly on the identification and characterization of peptides that may or may not be the result of proteolysis.

signaling molecules in various ways that affect the mechanical and biological properties of the tissue[172]. Protein content and composition of a sample can change over time and in response to different stimuli or conditions, such as development, differentiation, disease, or treatment[170]. For example, in cartilage tissue, the expression and modification of collagens and proteoglycans can change during endochondral ossification or in response to mechanical loading or inflammation[127]. Therefore, proteomics requires advanced techniques and methods to separate, identify, quantify, and characterize proteins in a comprehensive and accurate manner. One of the most powerful and widely used techniques for proteomics is mass spectrometry (MS), which can measure the mass-to-charge ratio (m/z) of ions in a gas phase[173]. However, MS also generates large and complex datasets that require sophisticated data analysis, computational methods, software, and/or accessible analysis platforms and tools for researchers in proteomics. Data analysis involves several steps, such as preprocessing, peak detection, alignment, normalization, feature selection, identification, quantification, annotation, and interpretation[174]. Computational methods in proteomics span a range of algorithms, from statistical techniques that assess data quality to machine learning models that enable analysis of mass spectrometry results.[175]. Accessible analysis platforms and tools are the online or cloud-based resources that offer data storage, processing, sharing, visualization, and integration for proteomics research[176,177]. In the next sections, I will describe the fundamental concepts of MS and how it can be applied to different types of proteomics workflows.

5.3.5 Mass spectrometry

Mass spectrometry is an analytical technique that measures the m/z of ions in a sample. MS can provide information about the identity, quantity, structure, and interactions of molecules, such as proteins, peptides, metabolites, etc. MS is widely used in proteomics, as it can identify and quantify thousands of proteins and PTMs in a single experiment[170]. Bottom-up proteomics is a common approach used in proteomic research that involves breaking down proteins into smaller fragments called peptides, usually through enzymatic digestion, before they are analyzed[178].

5.3.5.1 Protein extraction

Proteins are isolated from a biological sample, such as a cell, tissue, or fluid. Protein extraction can be performed using various methods, such as mechanical disruption, chemical lysis, or enzymatic digestion. Protein extraction can also be combined with fractionation or enrichment techniques, such as gel electrophoresis, chromatography, or immunoprecipitation, to reduce sample complexity and increase protein coverage[179].

5.3.5.2 Protein digestion

This is the process of breaking down proteins into smaller fragments, called peptides, using enzymes, such as trypsin, chymotrypsin, or Lys-C. Protein digestion can increase the solubility and ionization efficiency of proteins, as well as to generate peptides with specific termini that can be recognized by MS[180].

5.3.5.3 Liquid chromatography (LC)

LC works by separating peptides based on their physicochemical properties, such as hydrophobicity, charge, or size. LC can improve the sensitivity and resolution of MS, as well as to reduce ion suppression and co-elution effects. LC can be performed using various modes and columns, such as reversed-phase (RP), ion-exchange (IE), or hydrophilic interaction (HILIC) chromatography[181].

5.3.5.4 Ionization and fragmentation

Peptides are then converted into gas-phase ions and generating smaller fragments from them. Ionization and fragmentation can be performed using different methods, such as electrospray ionization (ESI), matrix-assisted laser desorption/ionization (MALDI), collision-induced dissociation (CID), electron transfer dissociation (ETD), and electron-capture dissociation (ECD).[182]

5.3.5.5 Acquisition methods

Depending on the type of information and analysis required, different acquisition modes of mass spectrometry can be used. Data-dependent acquisition (DDA) is also known as shotgun proteomics, because it randomly selects the most abundant precursor ions (peptides) in each scan and fragments them to generate MS/MS spectra[183]. This mode is useful for identifying and quantifying as many peptides and proteins as possible in a complex sample, but it may miss low-abundant or co-eluting peptides[184]. DDA data can be analyzed by matching the MS/MS spectra to a database of known peptide sequences or by using spectral libraries or de novo sequencing methods. Data-independent acquisition (DIA) is also known as sequential windowed acquisition of all theoretical fragment ion spectra (SWATH), depending on the instrument configuration. This mode aims to acquire MS/MS spectra for all possible precursor ions in a predefined mass range, regardless of their abundance or detection[185]. This mode is useful for generating comprehensive and consistent data sets that can be queried for any peptide of interest, but it requires high-resolution and high-speed instruments and sophisticated data analysis methods. DIA data can be analyzed by searching the raw or demultiplexed MS/MS spectra against a database or a spectral library, or by extracting the fragment ion chromatograms using targeted extraction methods[186].

5.3.5.6 Computational analysis

The spectra generated by MS are rich with information. Interpreting the spectra generated by MS is necessary to extract meaningful information from them which can eventually lead to biological insights. Computational analysis can involve various tasks and tools, such as spectrum deconvolution, peak picking, peptide identification, protein inference, quantification, PTM analysis, protein-protein interaction analysis, etc.[158].

5.3.6 Biomarkers

Biomarkers are measurable indicators of a biological state or condition, such as a disease, a response to a treatment, or a risk factor. Biomarkers can be derived from various sources, such as genes, transcripts, proteins, metabolites, etc. Biomarkers can have applications in biomedical research and clinical practice, such as diagnosis, prognosis, prediction, monitoring, and stratification.[187]

Proteomics is a powerful tool for biomarker discovery and validation, as it can identify and quantify proteins (and PTMs) that are associated with a biological state or condition. Proteomics can also provide information about the function and interactions of biomarkers, as well as their potential mechanisms of action. Proteomics can be applied to different types of biological samples relevant for biomarker research, such as serum, plasma, urine, saliva, and SF.[188].

Proteomics has been used to identify and validate biomarkers for knee OA, which can provide insights into the molecular mechanisms and clinical outcomes of the disease. Some well-studied biomarkers for knee OA are:

Collagen Type I (COL1A1)

Collagen type I is a major structural component of the meniscus in the knee joint, among other tissues[189]. The levels of COL1A1 fragments in SF can be an indicator of meniscal degradation as well as overall joint tissue remodeling.

Collagen Type II (COL2A1)

Collagen type II is the main form of collagen found in cartilage and is essential for cartilage integrity. In OA, cartilage breaks down, leading to the release of collagen type II fragments into the SF and bloodstream. These fragments serve as specific markers for cartilage degradation and can help in the assessment of disease severity and progression.

Aggrecan (ACAN)

ACAN is a large proteoglycan that provides the cartilage with its load-bearing properties[190]. Degradation of ACAN is one of the early events in OA development.

Monitoring the levels of aggrecan fragments in SF and serum could provide information on disease activity and progression[191].

Cartilage oligomeric matrix protein (COMP)

This is a structural protein that is involved in the assembly and maintenance of cartilage matrix[192]. COMP is released into the SF and serum as a result of cartilage degradation. COMP levels are elevated in patients with knee OA compared to healthy controls or patients with other joint diseases[193,194].

Matrix metalloproteinase 3 (MMP-3)

This is an enzyme that degrades various components of cartilage matrix, such as *COL2A*, ACAN, and proteoglycans. MMP-3 is produced by chondrocytes, synoviocytes, and inflammatory cells in response to mechanical stress or cytokines. MMP-3 levels are increased in the SF and serum of patients with knee OA compared to healthy controls or patients with other joint diseases. MMP-3 levels are also associated with disease activity, inflammation, and cartilage loss[195].

C-reactive protein (CRP)

CRP is an acute-phase protein that is produced by the liver in response to inflammation or infection. CRP is a marker of systemic inflammation and increased CRP levels have been demonstrated as a disease progression predictor of knee OA[196,197].

Interleukin 6 (IL-6)

IL-6 is produced by chondrocytes, synoviocytes, macrophages, and adipocytes in response to mechanical stress or cytokines. IL-6 levels are increased in the SF and serum of patients with knee OA compared to healthy controls or patients with other joint diseases[198]. IL-6 levels are also linked to disease activity, inflammation, cartilage degradation, bone remodeling, and metabolic disorders[199].

Tumor necrosis factor alpha (TNF-α)

TNF-α is produced by chondrocytes, synoviocytes, macrophages, and adipocytes in response to mechanical stress or cytokines. TNF-α levels are elevated in the SF and serum of patients with knee OA compared to healthy controls or patients with other joint diseases. TNF-α levels are also connected to inflammation, cartilage degradation, bone remodeling, and metabolic disorders[200–203].

Circulating Cartilage Acidic Protein 1 (CRTAC1)

Two recent large-scale studies have identified CRTAC1 as a predictor of OA progression[204,205]. Styrkarsdottir *et al.* identified the potential novel biomarker and found associations with both advanced OA, and future total joint replacement. CRTAC1 is a glycosylated ECM protein primarily produced in chondrocytes[206].

5.3.7 Role of proteomics in OA research

The pathogenesis of OA is affected by biological processes such as inflammation, matrix degradation, oxidative stress, cell death, and angiogenesis. To understand the molecular basis of OA and to discover biomarkers for its diagnosis, prognosis, and treatment response, it is important to analyze proteins from different sources that reflect the disease state and progression.

One of the most commonly used tissues for proteomic analysis in OA is articular cartilage, which is the main target of OA damage. Articular cartilage consists of chondrocytes embedded in ECM rich in collagen and proteoglycans. As previously mentioned in this thesis, the degradation of this matrix by proteases such as MMPs and ADAMTS is a hallmark of OA. Proteomic analysis of articular cartilage can reveal the changes in the composition and turnover of the matrix components, as well as the activation of signaling pathways and inflammatory mediators in chondrocytes. Articular cartilage sampling can be performed by biopsy or arthroscopy, but both methods are invasive and may cause damage to the tissue. Articular cartilage sample preparation for proteomic analysis involves several steps, such as homogenization, solubilization, reduction, alkylation, digestion, and peptide extraction.[207]

Another important tissue for proteomic analysis in OA is subchondral bone, which lies beneath the articular cartilage and provides mechanical and nutritional support to it. Subchondral bone undergoes structural and metabolic alterations in OA, such as sclerosis, cyst formation, and osteophyte growth. These changes affect the biomechanical properties of the bone and its interaction with cartilage, leading to increased mechanical stress and cytokine production. Proteomic analysis of subchondral bone can reveal the changes in bone remodeling processes, such as osteoclast and osteoblast activity, mineralization, angiogenesis, and innervation. Subchondral bone sampling can be performed by biopsy or arthroscopy, but both methods are invasive and may cause complications such as bleeding or infection. Subchondral bone sample preparation for proteomic analysis involves similar steps as articular cartilage, but with additional challenges due to the presence of calcium phosphate minerals that interfere with protein extraction and digestion.[208]

The synovial membrane is a soft tissue that lines the joint cavity and produces SF, a viscous liquid that lubricates and nourishes the joint surfaces. The synovial membrane is involved in the inflammatory response in OA, characterized by synovial hyperplasia, fibrosis, vascularization, and infiltration of immune cells. The SF reflects the changes occurring in the synovium and other joint tissues, as it contains proteins derived from cartilage degradation, synovial inflammation, and blood leakage. Synovial membrane sampling can be performed by biopsy or arthroscopy and sample preparation for proteomic analysis involves similar steps as articular cartilage and subchondral bone.[209]

The meniscus is a fibrocartilaginous structure that acts as a shock absorber and stabilizer between the femur and tibia. The meniscus is prone to injury and degeneration in OA, resulting in tears, calcifications, and loss of function. The meniscus also contributes to the production of SF and participates in joint homeostasis. Proteomic analysis of the meniscus can reveal the changes in extracellular matrix (ECM) organization, biomineralization processes, and crosstalk with other joint tissues. Meniscus sampling can be performed by biopsy or arthroscopy, but both methods are invasive and may cause damage to the tissue or affect its biomechanical properties. Meniscus sample preparation for proteomic analysis involves similar steps as articular cartilage, subchondral bone, and synovial membrane.[210,211]

The cruciate ligaments are collagenous connective tissues that connect the femur and tibia in the intercondylar area of the knee. They provide stability and balance to the joint during movement. Injury to the anterior cruciate ligament (ACL) is a common cause of post-traumatic OA, regardless of surgical reconstruction. The cruciate ligaments are also affected by degenerative processes in OA, such as collagen degradation, fibrosis, and vascularization. Proteomic analysis of the cruciate ligaments can reveal the changes in ECM organization, development of joint tissues, and biomechanical properties. Cruciate ligament sampling can be performed by biopsy or arthroscopy, but both methods are invasive and may cause complications such as bleeding or infection. Cruciate ligament sample preparation for proteomic analysis involves similar steps as articular cartilage, subchondral bone, synovial membrane, and meniscus.[212]

Biofluids are important sources of proteins for proteomic analysis in OA. Blood (plasma or serum) is easily obtained by minimally invasive procedures[213]. Another common source is SF, which is in direct contact with many of the tissues and cells of the knee joint, and thus may contain specific biomarkers that are not detectable in blood[214]. However, SF sampling is more invasive than blood sampling and requires clarification by centrifugation to remove contaminating cells and hyaluronic acid before proteomic analysis[215,216]. SF also has a high dynamic range and inter-individual variability, which require adequate sample preparation and normalization[217]. The secretome, i.e. the proteins secreted by a tissue, is another potential source for biomarker discovery because secreted proteins are released into the extracellular space and may be detectable in body fluids[218]. The secretome can be derived from cartilage, chondrocyte, or meniscus explants that are stimulated by various factors related to OA pathogenesis, such as cytokines (paper III), mechanical stress, or oxidative stress[219–221]. The secretome can provide information about the protein processing, modification, interaction, and function that occur in response to these stimuli.

6 Aims

Understanding the disease mechanisms that drive OA poses a considerable challenge that calls for a comprehensive and multifaceted approach to data analysis. Addressing this challenge, proteomics serves as a suitable approach, as it provides a “snapshot” of the proteome at a particular timepoint.

This thesis focuses on the development and application of proteomic data analysis methods to explore the OA proteome. **The overarching aim of this thesis is to develop and apply proteomic analytical methods to systematically explore proteolytic activity, protein abundance, and protein interactions across different stages of the disease.**

To achieve this aim, the following four key objectives were identified:

I. Develop a comprehensive and user-friendly bioinformatics tool to streamline and enhance the analysis of proteomics data in OA research.

II. Explore the proteolytic activity in synovial fluid of OA patients, highlighting proteolytic enzymes' role in the disease and providing a deeper understanding of the degradome's role in OA progression.

III. Establish an *ex vivo* model to elucidate early events in OA development, particularly focusing on the role of cytokine-mediated proteomic alterations in the meniscus tissue, thereby providing a functional understanding of the disease's pathogenesis.

IV. Perform a comparative analysis of protein dependencies in SF among different stages of OA to identify key proteins and subnetworks that play critical roles in disease progression, thereby contributing to the identification of potential biomarkers for OA.

7 Results and Discussion

7.1 Paper I - A platform for exploratory omics analysis

A common requirement for performing downstream proteomic analyses is having knowledge of using a programming or scripting language, and having the computational experience to convert between data formats. ProteoMill offers codeless analysis of protein expression datasets and is designed to be an easy to use alternative to script-based analyses for life science researchers. The tool is a web-based application that integrates data-upload, identifier conversion, quality control, differential expression, and network-based functional analysis into a single fast, interactive easy to use workflow.

7.1.1 Core functionality and architecture

The ProteoMill portal is developed using R Shiny[222]. It is hosted using Amazon Web Services (AWS), due to its capibility to adjust the server capacity depending on user activity. The AWS server instance used is EC2 of type t3.medium. The application was deployed to the server using R Shiny Server.

A common problem with omics analysis tools is outdated annotation data[223]. In the current software, annotation data sources are automatically maintained and kept up to date. Annotation is available for multiple model organisms (Figure 7).

7.1.2 User interface

The user interface consists of a left-hand menu through which the user navigates through the analysis workflow. The data-upload section lets the user upload the main dataset and optional annotation data with a single click. Throughout the analysis, a task menu is dynamically updated with feature documentation and hints to progress. Additional documentation available at https://bookdown.org/martin_ryden/proteomill_documentation/.

7.1.3 Reproducibility

Proteomill supports the use *reproducibility tokens* as a simple way to load the settings and database versions used in a previous experiment (Figure 7). When a user wants to save the settings used in an analysis, so that they or other researchers (who have access to the input file used) can reproduce the experiment, they may generate a unique key that allows to recover the current state and settings of the tool and its dependencies.

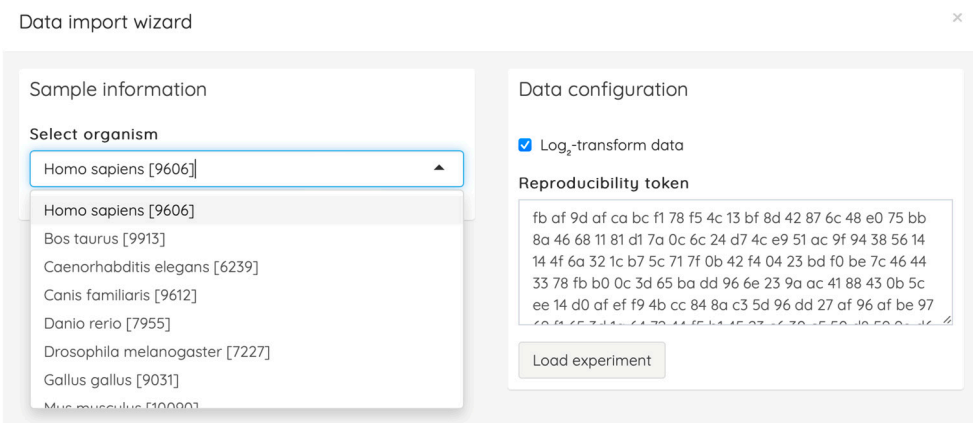

Figure 7. The initial step of data import in Proteomill, showing organisms for which annotation data is available, and the reproducibility token field which allows users to recover settings from previous analyses.

7.1.4 Differential expression analysis

Two R-packages, limma and DESeq2 are implemented for differential expression analysis. Differential expression analysis is conducted by specifying two contrasts and choosing a paired or non-paired design.

7.1.5 Visualization and functional analysis

The development of ProteoMill was driven by the need to make proteomics data analysis accessible to a wider scientific community. ProteoMill offers multiple possibilities for reducing data complexity. The implemented PCA plots (Figure 8) and pathway-based categorizations (Figure 9) are effective for distilling large datasets into more manageable and interpretable formats.

The sets of up- and downregulated proteins generated from differential expression are used to explore which pathways the proteins are over-represented in. The results are displayed as two separate tables for up- and down-regulation, both containing

the over-represented pathway name, its corresponding top-level pathway in the hierarchy, and a description of the relevance of the pathway.

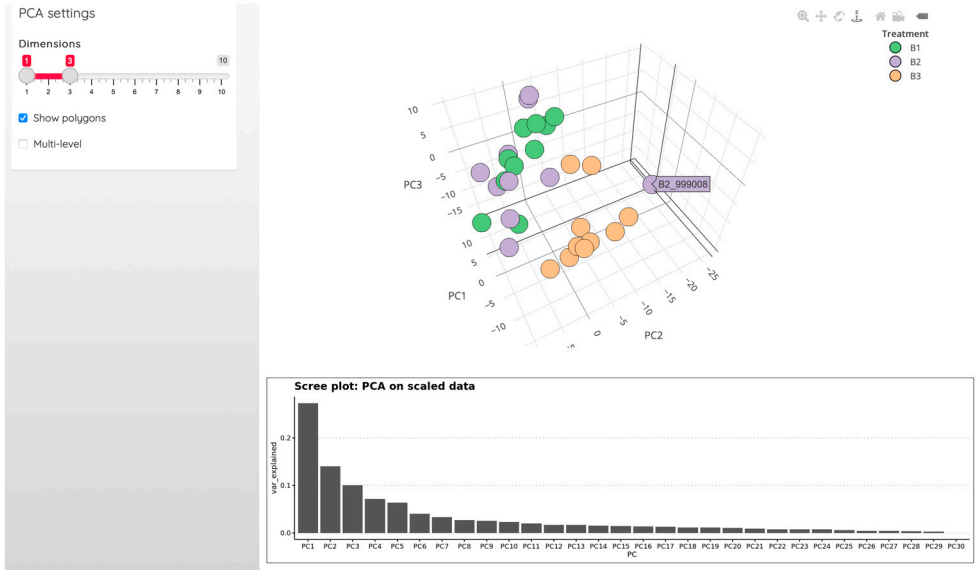

Figure 8. Principal component analysis (PCA), visualized as a three-dimensional plot (top), and scree plot, revealing the percentage of the total variance in the data that is accounted for by each principal component (bottom).

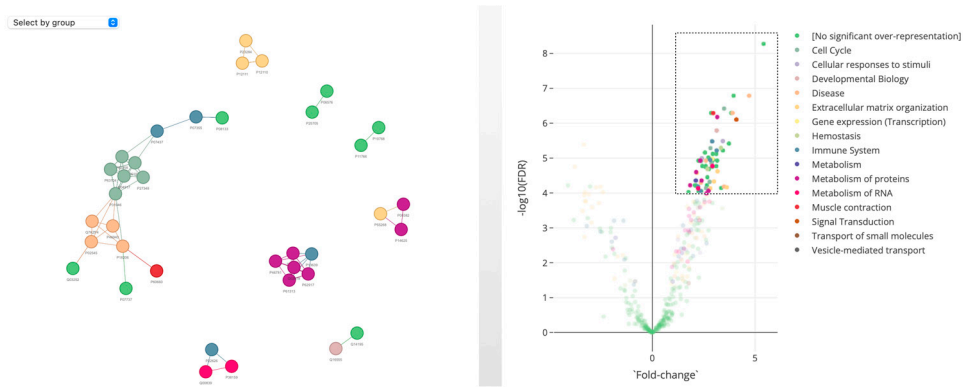

Figure 9. Visualization of a protein-protein interaction network (left) and volcano plot (right). Both plots are annotated with pathway data from Reactome. A selection tool is used to inspect a smaller group of proteins in the volcano plot. This selection is mirrored in the network plot.

7.1.6 Case study

We demonstrated the software's capabilities on a previously published dataset that contained 638 proteins. Out of 638 initial proteins in the original dataset, 12 proteins had obsolete identifiers which were substituted to their updated equivalents. Data inspection revealed that the B3 samples clustered together, separate from B1 and B2. The results are in line with the findings by Folkesson *et al.* characterizing the same dataset[224].

7.2 Paper II - Analysis of the SF degradome

Degradomics is the study of the breakdown products of proteins, such as peptides, that result from the activity of proteases, which are enzymes that cleave proteins[225]. This field can provide insights into the molecular mechanisms and biomarkers of various diseases. SF is a lubricating fluid that fills the joint cavities and reflects the metabolic and inflammatory changes in the joint tissues. In this study, we performed a peptidomic analysis of SF from knee OA patients and healthy controls, using LC-MS and DIA.

We identified and quantified endogenously cleaved peptides that were differentially abundant between the two groups, and mapped them to potential cleaving proteases using the MEROPS database[226] and an R-package we developed called proteasy[227]. We also performed pathway and network analysis to explore the functional implications of the differentially cleaved peptides and proteases.

The results of this study reveal the complexity and diversity of the proteolytic activity in SF of knee OA patients versus controls. The identification and quantification of endogenously cleaved peptides from 69 host proteins suggest that OA is associated with increased degradation of not only ECM components, but also plasma proteins involved in various biological functions. The pathway analysis of the host proteins of differentially abundant peptides indicated that OA protein activity affects processes such as immune system, transport of small molecules, and hemostasis.

7.2.1 An R-package for retrieving cleavage data

Protease cleavages affect many vital physiological processes, and dysregulation of proteolytic activity is associated with a variety of pathological conditions. *proteasy* is an R-package available on Bioconductor, and allows for batch identification of possible proteases for a set of substrates (protein IDs and peptide sequences). This tool may be useful in peptide-centric analyses of endogenously cleaved peptides.

This package utilizes data derived from the MEROPS database[226], a manually curated knowledgebase with information about proteolytic enzymes, their inhibitors, and substrates.

7.2.1.1 Using proteasy

A fast way to find which possible proteases, if any, are annotated as cleaving actors for a substrate is by using the function *searchSubstrate*. Using the parameter *summarize* will return only a vector of reviewed proteases (if true), a table with details about each cleaving event (if false). A corresponding function, *searchProtease*, exists to find which (if any) substrates a protease cleaves. The function *findProtease* automatically maps the peptide sequences to the full-length

protein sequence and obtains the start- and end-positions for the peptide. Then, the positions are searched against the MEROPs database and matches are returned. The function *browseProtease* takes a UniProt or MEROPS ID and opens the MEROPS summary page which corresponds to that ID in a web browser.

7.2.2 Main results and discussion

One of the main findings of this study is the increased abundance of cleaved peptides from HRG, AHSG, and APOA1 in OA SF. These are plasma proteins that have multiple roles in inflammation, bone metabolism, and lipid transport, respectively[228–231]. The increased cleavage of these proteins may reflect a loss of their regulatory functions or a response to inflammation in OA. For example, HRG has been shown to modulate the complement system, coagulation cascade, and cytokine production[229,232]. AHSG is known to inhibit hydroxyapatite formation and bone mineralization[233]. APOA1 is the major component of HDL and plays a role in reverse cholesterol transport[234]. The degradation of these proteins may affect the homeostasis of the joint and contribute to OA pathogenesis.

We also found an increased abundance of cleaved peptides from C3, FN1, and KNG1 in OA SF. These are proteins that are involved in complement activation, ECM remodeling, and regulation of serum peptides, respectively[235–238]. The cleavage of these proteins may indicate an enhanced inflammatory and catabolic activity in OA. For example, C3 is a key mediator of the complement system, which can induce cartilage damage, synovitis, and pain in OA[239,240]. FN1 is a glycoprotein that interacts with various ECM components and cell surface receptors, and modulates cell adhesion, migration, and differentiation[241,242]. KNG1 is a precursor of bradykinin, a potent vasodilator and pro-inflammatory mediator that can stimulate nociceptors and increase vascular permeability[243,244]. The degradation of these proteins may affect the structure and function of the joint tissues and may exacerbate OA symptoms.

We identified 192 potential proteases that could cleave the differentially abundant peptides in our dataset, 11 of which were also detected in SF. The main protease families that accounted for most of the proteolytic events were metalloproteinases, cathepsins, and caspases. These proteases have been implicated in OA pathogenesis, as they can degrade the ECM of cartilage and other joint tissues, as well as modulate inflammatory and apoptotic pathways. We also found other proteases, such as chymase, elastase, granzyme, and plasminogen activator, that could cleave plasma proteins involved in hemostasis, coagulation, and fibrinolysis. These proteases may affect the vascularization and fibrin deposition in OA[245,246].

We performed pathway analysis of the host proteins of differentially abundant peptides and found 65 pathways with three or more differentially abundant proteins in each set. These pathways were mainly related to immune system (particularly

complement activation), transport of small molecules (especially lipoproteins), and hemostasis. These results suggest that OA protein activity affects not only the ECM of the joint tissues, but also the homeostasis of plasma proteins. The complement system is a key mediator of inflammation and tissue damage in OA. The lipoprotein metabolism is altered in OA and may contribute to lipid accumulation and oxidative stress in the joint. The hemostasis system is involved in angiogenesis and fibrin deposition in OA.

7.3 Paper III - An *ex vivo* explant model of the human meniscus

The main objective of this study was to establish a human meniscus *ex vivo* model to study the meniscal response to cytokine treatment using a proteomics approach. We found that different cytokine treatments had effects on the release of proteins involved in inflammation, ECM degradation, and catabolism. We also observed zonal differences in the protein release profiles between the inner and outer regions of the meniscus.

The most prominent effect of interleukin-1 beta (IL1) treatment was the upregulation of several inflammatory mediators, such as interleukin-6 (IL6), interleukin-8 (CXCL8), growth-regulated alpha protein (CXCL1), C-X-C motif chemokine 6 (CXCL6), and neutrophil gelatinase-associated lipocalin (LCN2). These proteins are known to play important roles in the recruitment and activation of immune cells, such as neutrophils and macrophages, which can further amplify the inflammatory response and contribute to tissue damage. IL1 also increased the release of MMPs, such as MMP1, MMP2, MMP3, and MMP9, which are involved in the degradation of ECM components. However, we did not observe a strong increase in the release of ECM proteins in response to IL1 treatment, suggesting that IL1 alone may not be sufficient to induce a strong catabolic effect in healthy human menisci explants.

In contrast, we found that treatment with oncostatin M (OSM) + tumor necrosis factor alpha (TNF) or TNF + IL6 + soluble IL6 receptor (sIL6R) resulted in a marked increase in the release of ECM proteins (such as collagens, integrins, prolargin, tenascin, and versican). These proteins are essential for maintaining the structural integrity and biomechanical properties of the meniscus. The increased release of ECM proteins indicates that these cytokine combinations can induce a more pronounced catabolic effect than IL1 alone. This is consistent with previous studies showing that OSM and TNF can synergistically stimulate cartilage degradation and MMP expression *in vitro* and *in vivo*. Similarly, TNF and IL6 have been shown to cooperate in inducing cartilage catabolism and inflammation. The mechanisms underlying the synergistic effects of these cytokines may involve the activation of multiple signaling pathways, such as nuclear factor-kappa B (NF-kB), Janus kinase-signal transducer and activator of transcription (JAK-STAT), and mitogen-activated protein kinase (MAPK), which can regulate the expression of genes involved in inflammation and ECM metabolism.

7.3.1 Zonal differences

We also observed zonal differences in the protein release profiles between the inner and outer regions of the meniscus (Figure 10). The inner region represents the avascular zone, while the outer region is vascularized. The inner region is more cartilage-like and has a higher content of proteoglycans, while the outer region is more fibrous and has a higher content of type I collagen. These differences may affect the susceptibility and response of the meniscus to cytokine stimulation. We found that some cytokine treatments had a stronger effect on the inner region than on the outer region, such as TNF + IL6 + sIL6R, which induced a higher release of ECM proteins and proteases in the inner region. Conversely, some cytokine treatments had a stronger effect on the outer region than on the inner region, such as OSM + TNF, which induced a higher release of inflammatory mediators and protease inhibitors in the outer region. These results suggest that different regions of the meniscus may have different regulatory mechanisms and feedback loops that modulate the cytokine-induced effects.

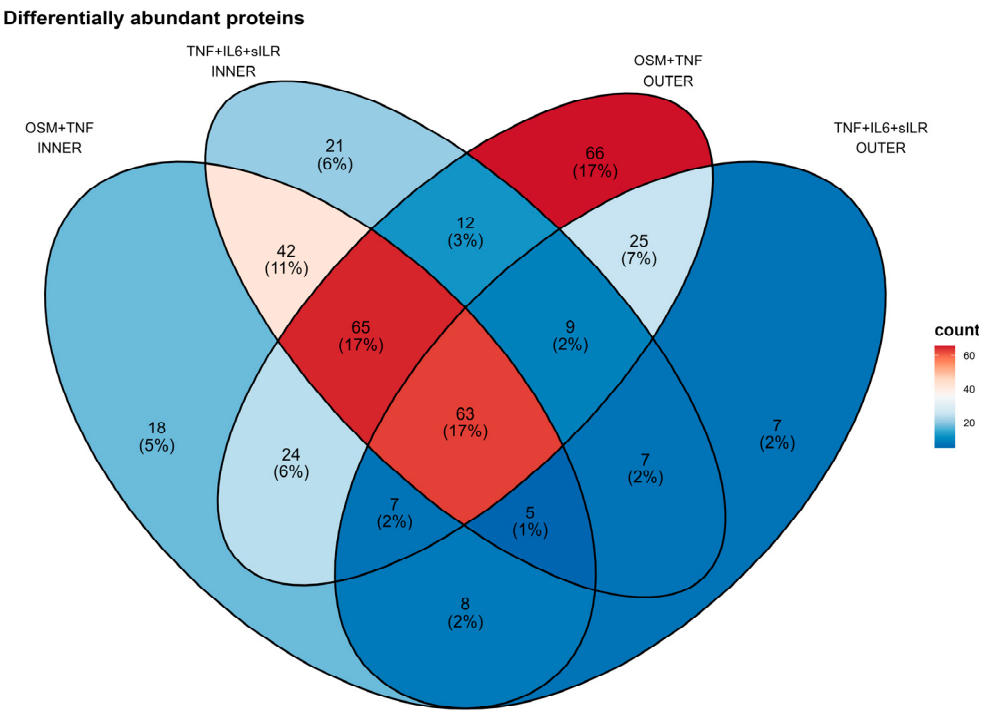

Figure 10. Venn diagram illustrating the number of differentially abundant proteins estimated in treatments oncostatin M (OSM)+ tumor necrosis factor (TNF) and TNF+IL6+ soluble IL6 receptor (sIL6R). The figure represents comparisons of treatment versus controls in the MEX-3 experiment.

7.3.2 Time-series clustering

We studied the time-dependent effect of treatments using time-series cluster analysis and found a reoccurring pattern of peak release on Day 9 in multiple treatments (Figure 11-12). This effect was not found to the same extent in the control group.

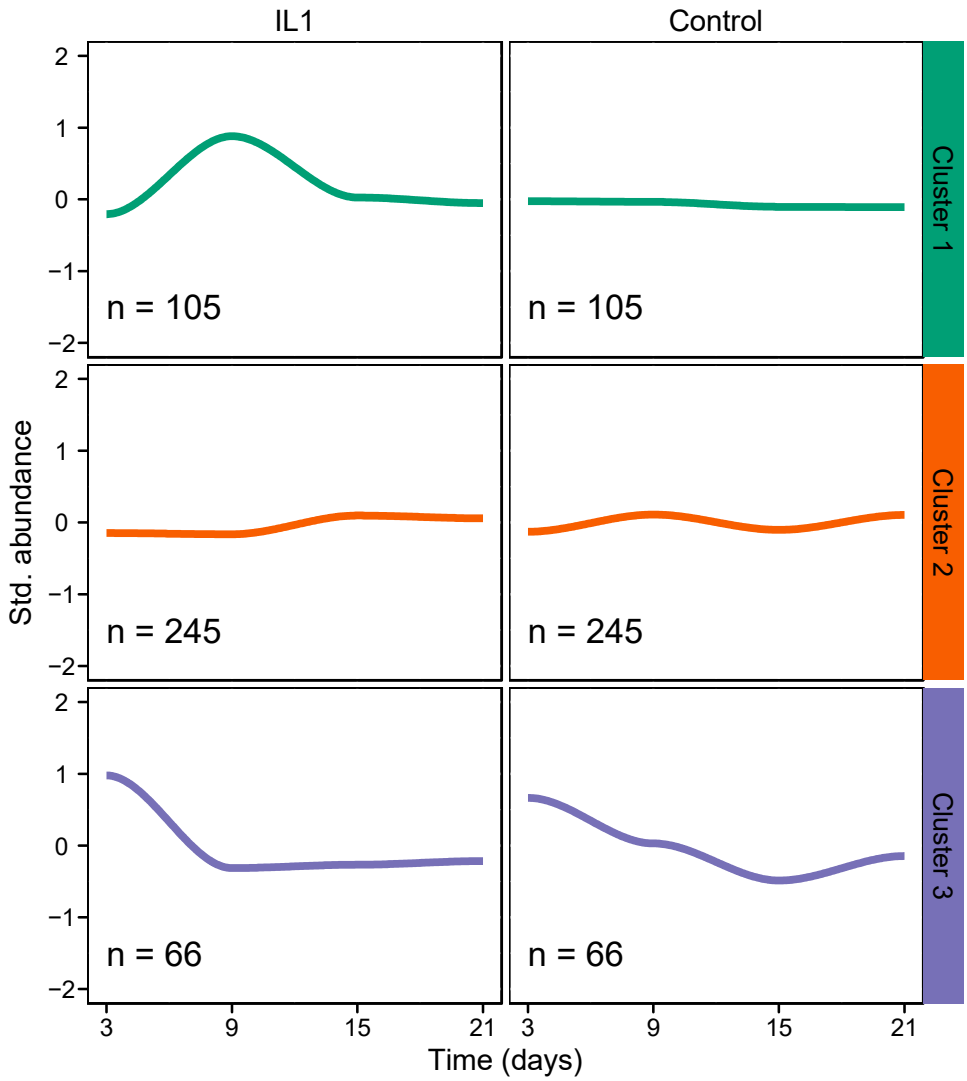

Figure 11. Release profiles for upregulated proteins by cluster for treatments interleukin-1 (IL1) and the respective release patterns for control group. The highlighted line indicates the mean standardized abundance for proteins within a cluster.

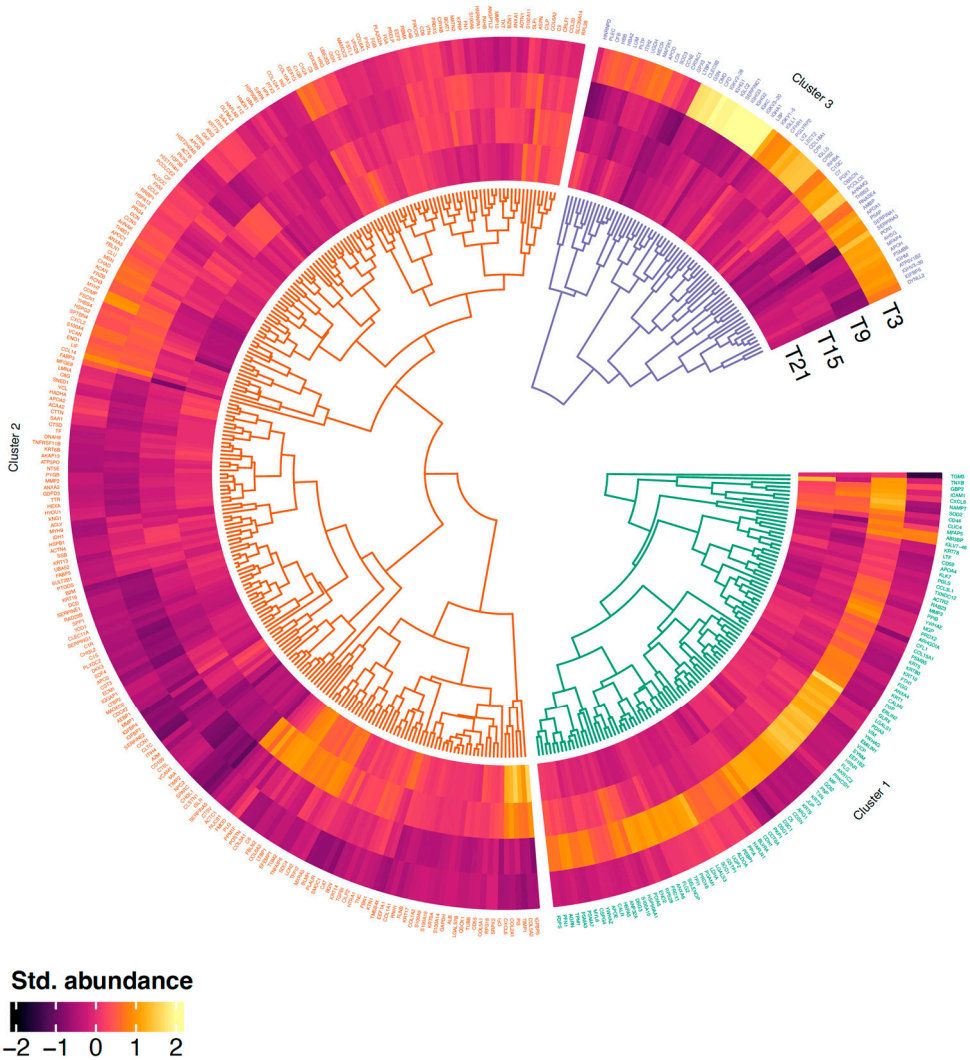

Figure 12. Circular heatmap illustrating the standardized abundance of temporal inflammatory markers (represented on the radial axis) over four timepoints (represented as concentric rings). The markers are grouped based on their temporal expression clusters (identified through time-series analysis), with each cluster represented by a distinct color. The innermost ring represents a hierarchical tree, demonstrating the proximity of markers within each cluster.

In summary, we have established a human meniscus *ex vivo* model to study the early events in OA development by proteomics. We have demonstrated that different cytokine treatments can induce different effects on the release of proteins involved in inflammation, ECM degradation, and catabolism. We have also revealed zonal differences in the protein release profiles between the inner and outer regions of the

meniscus. Our findings provide new insights into the molecular mechanisms involved in OA initiation and progression and may help identify potential biomarkers and therapeutic targets for OA.

7.4 Paper IV - The interactome of early OA

We performed a differential abundance analysis of SF proteins from healthy, mildly degenerated, and late-stage OA knees using the SOMAscan assay, which enabled us to measure over 6000 proteins in a single assay. We found 583 proteins that were upregulated in the late-stage OA group compared to the controls, including MMP1, MMP13, and IL6, which are known to be involved in cartilage degradation and inflammation in OA[247–249]. We also found several proteins that were downregulated in the late-stage OA group, such as NQO1, and ALDH2, which are involved in antioxidant defence and metabolism[250–252]. These results are consistent with previous studies that reported oxidative stress and metabolic dysregulation in OA joints[253–255]. Interestingly, we also identified some proteins that have not been previously associated with OA, such as ADH1C and GPD1, which are involved in alcohol metabolism and glycerol metabolism, respectively[256,257]. These proteins may represent novel biomarkers or therapeutic targets for OA.

To explore the protein interactions and networks in SF from healthy and mildly degenerated knees, we applied GGMs to a subset of 760 proteins that had the largest absolute fold changes between these two groups. We identified 102 proteins that were involved in GGMs networks, which revealed the conditional dependencies of proteins given all other proteins. We observed that the network complexity was reduced in the mild degeneration group compared to the controls, suggesting a disruption of joint homeostasis and protein interplay in early OA. We also identified several proteins and subnetworks that had unique edges in either group, indicating differential connectivity.

Among the proteins with unique edges in the healthy group, we found SLCO5A1, which is a member of the OATP family of solute transporters[258]. SLCO5A1 had the highest betweenness centrality score in this group, indicating its importance for the flow of information through the network. SLCO5A1 has not been previously linked to OA, but it may play a role in transporting substances that affect joint function and health, such as chemokines and metabolites. The downregulation of SLCO5A1 in the mild degeneration group may reflect impaired transport and signalling within the joint. Further studies are needed to elucidate the role of SLCO5A1 and other OATPs in OA.

Among the proteins with unique edges in the mild degeneration group, we found four potential bottleneck proteins with high betweenness centrality scores: HIBCH, DHX8, A1CF, and PHF3. These proteins had an influence on the network structure

and function by mediating information between other proteins. HIBCH is involved in valine catabolism, DHX8 is involved in splicing events, A1CF is involved in RNA editing, and PHF3 is involved in transcription and RNA processing[259–265]. These proteins may reflect altered gene expression and regulation in early OA pathogenesis. To our knowledge, these proteins have not been previously reported to be associated with OA and warrant further investigation.

We also detected several ECM proteins that were differentially abundant or connected between the healthy and mild degeneration groups. ECM proteins are expected to be prominent in OA due to their involvement in tissue degradation and remodelling. For example, we found ACAN, FN1, HAPLN1, and VWF to be downregulated in the mild degeneration group compared to the controls. These proteins are important for maintaining the structure and function of cartilage and meniscus[266–271]. We also found FBLN5 to have unique edges in the healthy group. FBLN5 is an ECM glycoprotein that modulates cell adhesion, migration, and differentiation[272–274]. FBLN5 may play a role in regulating joint homeostasis or inflammation in OA.

In summary, we have performed a comprehensive proteomic analysis of SF from healthy and early-stage OA knees using the SOMAscan assay and GGMs. We have identified several proteins and networks that are differentially abundant or connected between these two groups, suggesting potential mechanisms for early OA pathogenesis. Our study provides new insights into the complex and dynamic nature of protein interactions in OA joints and may pave the way for better understanding and treatment of this disease.

7.5 Limitations

This thesis covers a wide exploration of the molecular mechanisms underlying OA, and while each paper contributes a unique angle, there are important limitations that must be acknowledged.

The methodologies involved in paper I have a strong focus on P-value estimation and thresholding. P-values are based on the assumption that the null hypothesis is true, which is often implausible in real world scenarios[*]. A more informative approach to express uncertainty is to use confidence intervals, which giving us a range within which the true value is likely to lie, thereby providing more context to the point estimate.[275,276]

Paper II and paper III dealt with the issue of missing data in mass spectrometry analyses. While our approach attempted to maximize the use of available data (particularly the "qualitative analysis" in paper II), we recognize that this might introduce an arbitrary bias in the peptide/protein ranking.

The sample sizes, particularly in paper III, are notably small. In paper IV, we made use of regularization and stability selection procedures to minimize this limitation.

Mechanical loading is an important element in understanding the biomechanical aspects of OA[277]. While this was outside the scope of paper III, we acknowledge that the lack of this element was a limitation of the model.

As discussed in paper II, the representation of biological pathways through differential peptides/proteins is both abstract and subject to interpretation, potentially not capturing the complexity of underlying biology. In an attempt to circumvent these limitations in paper IV, as well as to address known biases in pathway databases toward more commonly studied diseases at the expense of less-explored areas like musculoskeletal diseases such as OA, we adopted a completely data-driven approach for this study.

[*] The null hypothesis that there is *no difference* between two treatments or *no effect* of an intervention is rarely true in a strict sense. There are usually many factors that influence the outcome of interest, and for two treatments or interventions to be exactly identical in is unlikely; there is always some difference or effect, even if it is very small or negligible.

8 Concluding remarks

Osteoarthritis remains a leading cause of disability and reduced quality of life worldwide. The disease affects an aging population, which becomes a growing challenge for our healthcare systems. Current therapies can only provide symptomatic relief, and there is no available cure. Furthermore, there are no reliable methods for early diagnosis and targeted treatment of OA, which underscores the need for a comprehensive understanding of its underlying mechanisms.

Despite the promise of proteomics for unravelling the complex biological processes that drive OA, its application to OA research is still limited. Gaps in the availability of comprehensive bioinformatics tools, insights into early OA events and stage-specific proteomic changes hamper the progress of OA research and prevent the development of effective diagnostic tools and treatments.

To address these gaps, the present thesis aimed to make contributions to the field of computational proteomics in OA research. It developed a comprehensive computational framework tailored for OA, explored the understudied area of proteolytic enzymes, created a novel *ex vivo* model to study early OA, and identified potential biomarkers and therapeutic targets through comparative analyses.

8.1.1 Contributions to the Field

The work of this thesis introduces valuable tools for proteomics data analysis, provides new insights into the molecular mechanisms of OA, and offers methodological advancements that can serve as a foundation for future studies in the field.

The first paper developed a comprehensive and user-friendly bioinformatics platform specifically designed for the streamlined analysis of proteomics data in OA research, filling a critical gap in the existing computational tools. In the the second paper, we presented a novel peptidomic analysis of SF in OA patients, identifying key proteases and their substrates. In the third paper, we establish a human meniscus *ex vivo* model to explore cytokine-induced catabolic processes. Lastly, in paper fourth paper, we used GGMs to identify proteins and subnetworks involved in early OA pathogenesis, including several proteins not previously associated with the disease.

Collectively, the thesis offers both practical tools and novel insights that advance the field.

8.1.2 Future Perspectives

The works in this thesis have provided new knowledge about the proteome and suggested potential biomarkers which will require further validation. The identification of key enzymes involved in the proteolytic degradation of SF proteins opens an opportunity for therapeutic intervention. Designing targeted inhibitors against these enzymes could be evaluated as potential treatments for OA. A modified implementation of the *ex vivo* model developed in this thesis can serve as a preliminary platform for testing the efficacy of these inhibitors. Further studies will be needed to assess the safety and specificity of these targets before advancing to clinical trials.

Multi-omics data integration, involving genomics, transcriptomics, and metabolomics, together with proteomics data, could offer a holistic view of OA disease mechanisms and reveal novel targets. The ProteoMill portal, with its capabilities for analysing omics data, could be further improved to integrate multi-omics datasets. Such an integrative approach would provide a more comprehensive view, potentially allowing us to reveal novel interactions and regulatory networks that may have been overlooked in these and other studies focused on single omics. Future versions of ProteoMill could also integrate more robust uncertainty measures such as confidence intervals and Bayesian methods, to provide a fuller picture of the data's variance and bring more nuance to the analytical findings.

The introduction of GGMs in paper IV highlights the potential of computational models in understanding complex protein interactions. Future research could explore machine learning techniques such as deep learning models for identifying non-linear relationships in proteomics data.

8.1.3 Conclusions

This thesis lays the foundation for a multi-faceted approach to understanding OA through computational proteomics. It provides both functional insights and practical tools that future research may build upon.

9 Acknowledgements

I am grateful for having had the opportunity to work with incredibly knowledgeable, engaged, and kind colleagues and collaborators at the Clinical Epidemiology Unit and at BMC in Lund. Your inspiration and guidance have meant everything during my PhD journey, and I know they will continue to influence me long after.

First and foremost, I want to extend my deepest gratitude to my main supervisor, *Martin Englund*, for bringing me on board on this journey, and for your trust in me to join your group. Your support has made all the difference.

My co-supervisor *Patrik Önnerfjord*, your mentorship has been invaluable. You have consistently made the time to guide me (be it in research or on the golf course), providing thoughtful feedback that has shaped my work.

To my co-supervisor *Aleksandra Turkiewicz*, your expertise has been nothing short of enlightening. Your advice has made me a better researcher and thinker, and I am grateful for your ability to cut through the noise to provide feedback and encouragement.

To my co-supervisor *Neserin Ali*, the insightful advice, methodological help, and ideas you have provided since my first day at Wigerthuset has been inspirational. Thank you for laying down the tracks on which I could explore and express my own research ideas.

I also wish to thank my colleagues at the Molecular Skeletal Biology group at BMC, *Anders* and *Karin*, and *André,* for your helpful efforts and valuable insights that has genuinely enriched my thesis.

Velocity, we have walked on parallel paths from Indiana to Wigerthuset, and now onto entrepreneurial adventures. Your mind embodies true interdisciplinarity (yes, it's a word!) and I look forward to the masterpiece your next chapter will synthesize, both in art and life. Speaking of new chapters, to the additional members of our non-fiction book club, *Amanda* and *Andrea*, you are amazing researchers and lunch mates. May we never find the bottom of our reading list.

To *Lotte*, thank you for always being so helpful and making us appreciate our workplace as if it was a second home. And to all past and present fika enthusiasts at Clin. Epi.; *Ali*, *Annette*, *Barbara*, *Clara*, *Ebba*, *Elin, Emma*, *Karin*, *Sara, and Terese*, to name a few - thank you for your wonderful company.

Aaron, friend and fellow bioinformagician, thank you for having my back and helping me win at pub-quizzes.

KB, *Svante*, *Victoria*, *Gustav* – My friends, thank you for your constant support (albeit not for the throne).

Christian, *Niklas*, *Mattias* –  ,  ,  .

Emma, thank you for being a constant source of reassurance and insight, and for offering a unique perspective on things. Our haikus about hedgehogs are moments I continually reflect upon with fondness.

Tack till mina föräldrar, *Eva* och *Sven* som har byggt en värld full av möjligheter och stöttat mig så jag kunnat utforska världen på mitt eget sätt. Tack till lillebror *Henrik* och mormor *Titti*, som ständigt påminner mig om vad som verkligen är viktigt.

10 References

1. Lewis, J. & Bartlett, A. Inscribing a discipline: tensions in the field of bioinformatics. *N. Genet. Soc.* **32**, 243–263 (2013).
2. Altman, R. B. & Dugan, J. M. Defining Bioinformatics and Structural Bioinformatics. in *Structural Bioinformatics* 1–14 (2003). doi:10.1002/0471721204.ch1.
3. Mulder, N. *et al.* The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Comput. Biol.* **14**, e1005772 (2018).
4. KIM, Y.-M., Poline, J.-B. & Dumas, G. Experimenting with Reproducibility: a case study of Robustness in Bioinformatics. *GigaScience* **7**, giy077- (2018).
5. Peng, R. D. Reproducible Research in Computational Science. *Science* **334**, 1226–1227 (2011).
6. Goodman, S. N., Fanelli, D. & Ioannidis, J. P. A. What does research reproducibility mean? *Sci. Transl. Med.* **8**, 341ps12 (2016).
7. Roy, S. *et al.* Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J. Mol. Diagn.* **20**, 4–27 (2018).
8. Marx, V. To share is to be a scientist. *Nat. Methods* **20**, 984–989 (2023).
9. Deutsch, E. W. *et al.* Proteomics Standards Initiative at Twenty Years: Current Activities and Future Work. *J. Proteome Res.* **22**, 287–301 (2023).
10. Wratten, L., Wilm, A. & Göke, J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat. Methods* **18**, 1161–1168 (2021).
11. Hoehndorf, R., Schofield, P. N. & Gkoutos, G. V. The role of ontologies in biological and biomedical research: a functional perspective. *Brief. Bioinform.* **16**, 1069–1080 (2015).
12. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
13. Martens, L., Orchard, S., Apweiler, R. & Hermjakob, H. Human Proteome Organization Proteomics Standards Initiative. *Mol. Cell. Proteom.* **6**, 1666–1667 (2007).
14. Hoehndorf, R. *et al.* Integrating systems biology models and biomedical ontologies. *BMC Syst. Biol.* **5**, 124 (2011).

15. Gruening, B. *et al.* Recommendations for the packaging and containerizing of bioinformatics software. *F1000Research* **7**, 742 (2018).
16. Tommaso, P. D. *et al.* The impact of Docker containers on the performance of genomic pipelines. *PeerJ* **3**, e1273 (2015).
17. Mohamadi, H., Chu, J., Vandervalk, B. P. & Birol, I. ntHash: recursive nucleotide hashing. *Bioinformatics* **32**, 3492–3494 (2016).
18. Rivest, R. *The MD5 message-digest algorithm*.
19. Rydén, M., Englund, M. & Ali, N. ProteoMill: Efficient network-based functional analysis portal for proteomics data. *Bioinformatics* btab373- (2021) doi:10.1093/bioinformatics/btab373.
20. Vailati-Riboni, M., Palombo, V. & Loor, J. J. Periparturient Diseases of Dairy Cows. 1–7 (2017) doi:10.1007/978-3-319-43033-1_1.
21. Hieter, P. & Boguski, M. Functional Genomics: It's All How You Read It. *Science* **278**, 601–602 (1997).
22. García-Campos, M. A., Espinal-Enríquez, J. & Hernández-Lemus, E. Pathway Analysis: State of the Art. *Front. Physiol.* **6**, 383 (2015).
23. Mitrea, C. *et al.* Methods and approaches in the topology-based analysis of biological pathways. *Front. Physiol.* **4**, 278 (2013).
24. Wang, M. *et al.* RobNorm: model-based robust normalization method for labeled quantitative mass spectrometry proteomics data. *Bioinformatics* **37**, 815–821 (2020).
25. Mertens, B. J. A. Transformation, normalization and batch effect in the analysis of mass spectrometry data for omics studies. (2016).
26. Cheadle, C., Vawter, M. P., Freed, W. J. & Becker, K. G. Analysis of Microarray Data Using Z Score Transformation. *J. Mol. Diagn.* **5**, 73–81 (2003).
27. Čuklina, J. *et al.* Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol. Syst. Biol.* **17**, e10240 (2021).
28. Huang, X., Wu, L. & Ye, Y. A Review on Dimensionality Reduction Techniques. *Int. J. Pattern Recognit. Artif. Intell.* **33**, 1950017 (2019).
29. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
30. Jaeger, A. & Banks, D. Cluster analysis: A modern statistical review. *Wiley Interdiscip. Rev.: Comput. Stat.* **15**, (2023).
31. Oyelade, J. *et al.* Clustering Algorithms: Their Application to Gene Expression Data. *Bioinform. Biol. Insights* **10**, BBI.S38316 (2016).
32. Walker, M. G., Volkmuth, W., Sprinzak, E., Hodgson, D. & Klingler, T. Prediction of Gene Function by Genome-Scale Expression Analysis: Prostate Cancer-Associated Genes. *Genome Res.* **9**, 1198–1203 (1999).
33. Lee, H. K., Hsu, A. K., Sajdak, J., Qin, J. & Pavlidis, P. Coexpression Analysis of Human Genes Across Many Microarray Data Sets. *Genome Res.* **14**, 1085–1094 (2004).
34. Nies, H. W. *et al.* A Review of Computational Methods for Clustering Genes with Similar Biological Functions. *Processes* **7**, 550 (2019).

35. Pirim, H., Ekşioğlu, B., Perkins, A. D. & Yüceer, Ç. Clustering of high throughput gene expression data. *Comput. Oper. Res.* **39**, 3046–3061 (2012).

36. Sloutsky, R., Jimenez, N., Swamidass, S. J. & Naegle, K. M. Accounting for noise when clustering biological data. *Brief. Bioinform.* **14**, 423–436 (2013).

37. Jiang, B., Pei, J., Tao, Y. & Lin, X. Clustering Uncertain Data Based on Probability Distribution Similarity. *IEEE Trans. Knowl. Data Eng.* **25**, 751–763 (2013).

38. Murtagh, F. & Contreras, P. Algorithms for hierarchical clustering: an overview. *WIREs Data Min. Knowl. Discov.* **2**, 86–97 (2011).

39. Day, W. H. E. & Edelsbrunner, H. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* **1**, 7–24 (1984).

40. Chormunge, S. & Jena, S. Evaluation of Clustering Algorithms for High Dimensional Data Based on Distance Functions. 78 (2014) doi:10.1145/2677855.2677933.

41. Cunningham, K. M. & Ogilvie, J. C. Evaluation of Hierarchical Grouping Techniques: A Preliminary Study. *Comput. J.* **15**, 209–213 (1972).

42. Boley, D. Principal Direction Divisive Partitioning. *Data Min. Knowl. Discov.* **2**, 325–344 (1998).

43. Huang, W., Peng, Y., Ge, Y. & Kong, W. A new Kmeans clustering model and its generalization achieved by joint spectral embedding and rotation. *PeerJ Comput. Sci.* **7**, e450 (2021).

44. Mirzal, A. Statistical Analysis of Microarray Data Clustering using NMF, Spectral Clustering, Kmeans, and GMM. *IEEEACM Trans. Comput. Biol. Bioinform.* **19**, 1173–1192 (2020).

45. Borgelt, C. & Yarikova, O. Initializing k-means Clustering. *Proc. 9th Int. Conf. Data Sci., Technol. Appl.* 260–267 (2020) doi:10.5220/0009872702600267.

46. Chen, X. K-Mean Clustering. *Int. J. Comput. Appl.* **177**, 24–27 (2020).

47. Vichi, M., Cavicchia, C. & Groenen, P. J. F. Hierarchical Means Clustering. *J. Classif.* **39**, 553–577 (2022).

48. Rodriguez, M. Z. *et al.* Clustering algorithms: A comparative approach. *PloS one* **14**, e0210236 (2016).

49. Abanda, A., Mori, U. & Lozano, J. A. A review on distance based time series classification. *Data Min. Knowl. Discov.* **33**, 378–412 (2019).

50. Ramoni, M. F., Sebastiani, P. & Kohane, I. S. Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci.* **99**, 9121–9126 (2002).

51. McDowell, I. C. *et al.* Clustering gene expression time series data using an infinite Gaussian process mixture model. *PLoS Comput. Biol.* **14**, e1005896 (2018).

52. Everitt, B., Landau, S., Leese, M. & Stahl, D. *Cluster analysis*. (Wiley, 2011).

53. Milligan, G. W. & Cooper, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179 (1985).

54. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **63**, 411–423 (2001).

55. Kong, W., Hui, H. W. H., Peng, H. & Goh, W. W. B. Dealing with missing values in proteomics data. *PROTEOMICS* **22**, e2200092 (2022).

56. Rubin, D. B. Inference and Missing Data. *Biometrika* **63**, 581 (1976).
57. Pimentel, E. C. G., Edel, C., Emmerling, R. & Götz, K.-U. How imputation errors bias genomic predictions. *J. Dairy Sci.* **98**, 4131–4138 (2015).
58. Huisman, M. Imputation of Missing Item Responses: Some Simple Techniques. *Qual. Quant.* **34**, 331–351 (2000).
59. Titterington, D. M., Madow, W. G., Nisselson, H., Olkin, I. & Rubin, D. Incomplete Data in Sample Surveys. *J. Am. Stat. Assoc.* **80**, 768 (1985).
60. Dempster, A. & Rubin, D. Incomplete data in sample surveys. *Sample surveys* **2**, 3–10 (1983).
61. Daly, D. S. *et al.* Mixed-Effects Statistical Model for Comparative LC−MS Proteomics Studies. *J. Proteome Res.* **7**, 1209–1217 (2008).
62. Wang, J., Wang, P., Hedeker, D. & Chen, L. S. Using multivariate mixed-effects selection models for analyzing batch-processed proteomics data with non-ignorable missingness. *Biostatistics* **20**, 648–665 (2018).
63. Leitch, M. C., Mitra, I. & Sadygov, R. G. Generalized linear and mixed models for label-free shotgun proteomics. *Stat. Interface* **5**, 89–98 (2012).
64. Ji, H. & Liu, X. S. Analyzing 'omics data using hierarchical models. *Nat Biotechnol* **28**, 337–340 (2010).
65. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).
66. Klann, K. & Münch, C. PBLMM: Peptide-based linear mixed models for differential expression analysis of shotgun proteomics data. *J. Cell. Biochem.* **123**, 691–696 (2022).
67. Zhu, Y. *et al.* DEqMS: A Method for Accurate Variance Estimation in Differential Protein Expression Analysis*. *Mol. Cell. Proteom.* **19**, 1047–1057 (2020).
68. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
69. Yao, Q. *et al.* Osteoarthritis: pathogenic signaling pathways and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 56 (2023).
70. Zhai, G. Alteration of Metabolic Pathways in Osteoarthritis. *Metabolites* **9**, 11 (2019).
71. Saito, T. & Tanaka, S. Molecular mechanisms underlying osteoarthritis development: Notch and NF-κB. *Arthritis Res. Ther.* **19**, 94 (2017).
72. Haw, R., Hermjakob, H., D'Eustachio, P. & Stein, L. Reactome pathway analysis to enrich biological discovery in proteomics data sets. *PROTEOMICS* **11**, 3598–3613 (2011).
73. Wu, X., Hasan, M. A. & Chen, J. Y. Pathway and network analysis in proteomics. *J. Theor. Biol.* **362**, 44–52 (2014).
74. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **14**, 482–517 (2019).
75. Ma, J., Shojaie, A. & Michailidis, G. A comparative study of topology-based pathway enrichment analysis methods. *BMC Bioinform.* **20**, 546 (2019).

76. Nguyen, T.-M., Shafi, A., Nguyen, T. & Draghici, S. Identifying significantly impacted pathways: a comprehensive review and assessment. *Genome Biol.* **20**, 203 (2019).

77. Chowbina, S. R. *et al.* HPD: an online integrated human pathway database enabling systems biology studies. *BMC Bioinform.* **10**, S5 (2009).

78. Huang, H. *et al.* Identification of pathways and genes associated with synovitis in osteoarthritis using bioinformatics analyses. *Sci. Rep.* **8**, 10050 (2018).

79. Zhao, L. *et al.* Data-independent acquisition-based proteomics analysis correlating type 2 diabetes mellitus with osteoarthritis in total knee arthroplasty patients. *Medicine* **101**, e28738 (2022).

80. Ding, Y., Wang, L., Zhao, Q., Wu, Z. & Kong, L. MicroRNA-93 inhibits chondrocyte apoptosis and inflammation in osteoarthritis by targeting the TLR4/NF-κB signaling pathway. *Int. J. Mol. Med.* **43**, 779–790 (2019).

81. Lei, J. *et al.* Proteomic analysis of knee cartilage reveals potential signaling pathways in pathological mechanism of Kashin-Beck disease compared with osteoarthritis. *Sci. Rep.* **10**, 6824 (2020).

82. Su, S.-L. *et al.* Gene–gene interactions between TGF-β/Smad3 signalling pathway polymorphisms affect susceptibility to knee osteoarthritis. *BMJ Open* **5**, e007931 (2015).

83. Cherifi, C., Monteagudo, S. & Lories, R. J. Promising targets for therapy of osteoarthritis: a review on the Wnt and TGF-β signalling pathways. *Ther. Adv. Musculoskelet. Dis.* **13**, 1759720X211006959 (2021).

84. Loughlin, J. *et al.* Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. *Proc. Natl. Acad. Sci.* **101**, 9757–9762 (2004).

85. Li, Z. *et al.* p38MAPK Signaling Pathway in Osteoarthritis: Pathological and Therapeutic Aspects. *J. Inflamm. Res.* **15**, 723–734 (2022).

86. Ma, C.-Y., Chen, Y.-P. P., Berger, B. & Liao, C.-S. Identification of protein complexes by integrating multiple alignment of protein interaction networks. *Bioinformatics* **33**, 1681–1688 (2017).

87. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84**, 1–25 (2015).

88. Wu, Z., Liao, Q. & Liu, B. A comprehensive review and evaluation of computational methods for identifying protein complexes from protein–protein interaction networks. *Brief. Bioinform.* **21**, 1531–1548 (2019).

89. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein-Protein Interaction Detection: Methods and Analysis. *Int. J. Proteom.* **2014**, 147648 (2014).

90. Kann, M. G. Protein interactions and disease: computational approaches to uncover the etiology of diseases. *Brief. Bioinform.* **8**, 333–346 (2007).

91. Rivas, J. D. L. & Fontanillo, C. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *Plos Comput Biol* **6**, e1000807 (2010).

92. Raman, K. Construction and analysis of protein–protein interaction networks. *Autom. Exp.* **2**, 2 (2010).

93. Pedamallu, C. S. & Ozdamar, L. Modeling, Dynamics, Optimization and Bioeconomics I. *Springer Proc. Math. Stat.* 511–519 (2014) doi:10.1007/978-3-319-04849-9_30.

94. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2018).

95. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic acids Res.* **47**, D529–D541 (2018).

96. Toro, N. del *et al.* The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **50**, D648–D653 (2021).

97. Licata, L. *et al.* MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* **40**, D857–D861 (2012).

98. Alanis-Lobato, G. & Schaefer, M. H. Protein-Protein Interaction Networks. *Methods Mol. Biol.* 135–144 (2019) doi:10.1007/978-1-4939-9873-9_11.

99. Ren, J., Wang, J., Li, M. & Wu, F. Discovering essential proteins based on PPI network and protein complex. *Int. J. data Min. Bioinform.* **12**, 24–43 (2015).

100. Jalili, M. *et al.* Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks. *Front. Physiol.* **7**, 375 (2016).

101. He, X. & Zhang, J. Why Do Hubs Tend to Be Essential in Protein Networks? *PLoS Genet.* **2**, e88 (2006).

102. Zhong, J., Xiang, D. & Ma, X. Prediction and analysis of osteoarthritis hub genes with bioinformatics. *Ann Transl Medicine* **0**, 0–0 (2023).

103. Han, J.-D. J. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**, 88–93 (2004).

104. Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. & Gerstein, M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Comput. Biol.* **3**, e59 (2007).

105. Vallabhajosyula, R. R., Chakravarti, D., Lutfeali, S., Ray, A. & Raval, A. Identifying Hubs in Protein Interaction Networks. *PLoS ONE* **4**, e5344 (2009).

106. Schober, P., Boer, C. & Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia Analg.* **126**, 1763–1768 (2018).

107. Buzzatto-Leite, I., Afonso, J., Silva-Vignato, B., Coutinho, L. L. & Alvares, L. E. Differential gene co-expression network analyses reveal novel molecules associated with transcriptional dysregulation of key biological processes in osteoarthritis knee cartilage. *Osteoarthr. Cartil. Open* **4**, 100316 (2022).

108. Cao, J. *et al.* Weighted gene co-expression network analysis reveals specific modules and hub genes related to immune infiltration of osteoarthritis. *Ann. Transl. Med.* **0**, 0–0 (2021).

109. Gao, X., Sun, Y. & Li, X. Identification of key gene modules and transcription factors for human osteoarthritis by weighted gene co-expression network analysis. *Exp. Ther. Med.* **18**, 2479–2490 (2019).

110. Altenbuchinger, M., Weihs, A., Quackenbush, J., Grabe, H. J. & Zacharias, H. U. Gaussian and Mixed Graphical Models as (multi-)omics data analysis tools. *Biochim. Biophys. Acta (BBA) - Gene Regul. Mech.* **1863**, 194418 (2020).

111. Braeuning, A., Frenzel, F. & Lampen, A. On the necessity of careful interpretation of omics data. *Arch. Toxicol.* **92**, 2701–2702 (2018).

112. Yamada, R., Okada, D., Wang, J., Basak, T. & Koyama, S. Interpretation of omics data analyses. *J. Hum. Genet.* **66**, 93–102 (2021).

113. Midway, S. R. Principles of Effective Data Visualization. *Patterns* **1**, 100141 (2020).

114. Loeser, R. F., Collins, J. A. & Diekman, B. O. Ageing and the pathogenesis of osteoarthritis. *Nat. Rev. Rheumatol.* **12**, 412–420 (2016).

115. Zhang, Y. & Jordan, J. M. Epidemiology of Osteoarthritis. *Clin. Geriatr. Med.* **26**, 355–369 (2010).

116. Guermazi, A., Roemer, F. W. & Hayashi, D. Imaging of osteoarthritis: update from a radiological perspective. *Curr. Opin. Rheumatol.* **23**, 484–491 (2011).

117. Bedson, J. & Croft, P. R. The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. *BMC Musculoskelet. Disord.* **9**, 116–116 (2008).

118. Cross, M. *et al.* The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Ann. Rheum. Dis.* **73**, 1323 (2014).

119. Hunter, D. J. & Bierma-Zeinstra, S. Osteoarthritis. *Lancet* **393**, 1745–1759 (2019).

120. Collaborators, G. 2019 D. and I. *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).

121. Long, H. *et al.* Prevalence Trends of Site-Specific Osteoarthritis From 1990 to 2019: Findings From the Global Burden of Disease Study 2019. *Arthritis Rheumatol.* **74**, 1172–1183 (2022).

122. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (Lond., Engl.)* **392**, 1789–1858 (2018).

123. Bijlsma, J. W., Berenbaum, F. & Lafeber, F. P. Osteoarthritis: an update with relevance for clinical practice. *Lancet* **377**, 2115–2126 (2011).

124. Prieto-Alhambra, D. *et al.* Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. *Ann. Rheum. Dis.* **73**, 1659 (2014).

125. Turkiewicz, A. *et al.* Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthr. Cartil.* **22**, 1826–1832 (2014).

126. Bitton, R. The economic burden of osteoarthritis. *Am. J. Manag. care* **15**, S230-5 (2009).

127. Goldring, M. B. & Goldring, S. R. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis: Articular cartilage and subchondral bone. *Ann. N. York Acad. Sci.* **1192**, 230–237 (2010).

128. Felson, D. T. & Zhang, Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum.* **41**, 1343–1355 (1998).

129. Blagojevic, M., Jinks, C., Jeffery, A. & Jordan, K. P. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthr. Cartil.* **18**, 24–33 (2010).

130. Greene, M. A. & Loeser, R. F. Aging-related inflammation in osteoarthritis. *Osteoarthr. Cartil.* **23**, 1966–1971 (2015).

131. Reyes, C. *et al.* Association Between Overweight and Obesity and Risk of Clinically Diagnosed Knee, Hip, and Hand Osteoarthritis: A Population-Based Cohort Study. *Arthritis Rheumatol.* **68**, 1869–1875 (2016).

132. Dahaghin, S., Bierma-Zeinstra, S. M. A., Koes, B. W., Hazes, J. M. W. & Pols, H. A. P. Do metabolic factors add to the effect of overweight on hand osteoarthritis? The Rotterdam Study. *Ann. Rheum. Dis.* **66**, 916 (2007).

133. Raud, B. *et al.* Level of obesity is directly associated with the clinical and functional consequences of knee osteoarthritis. *Sci. Rep.* **10**, 3601 (2020).

134. Valdes, A. M. & Spector, T. D. Genetic epidemiology of hip and knee osteoarthritis. *Nat. Rev. Rheumatol.* **7**, 23–32 (2011).

135. Magnusson, K. *et al.* Genetic factors contribute more to hip than knee surgery due to osteoarthritis – a population-based twin registry study of joint arthroplasty. *Osteoarthr. Cartil.* **25**, 878–884 (2017).

136. Lieberthal, J., Sambamurthy, N. & Scanzello, C. R. Inflammation in joint injury and post-traumatic osteoarthritis. *Osteoarthr. Cartil.* **23**, 1825–1834 (2015).

137. FAIRBANK, T. J. Knee joint changes after meniscectomy. *J. bone Jt. Surg. Br. Vol.* **30B**, 664–70 (1948).

138. Buckwalter, J. A. & Mankin, H. J. Articular cartilage: degeneration and osteoarthritis, repair, regeneration, and transplantation. *Instr. course Lect.* **47**, 487–504 (1998).

139. Fox, A. J. S., Bedi, A. & Rodeo, S. A. The Basic Science of Human Knee Menisci: Structure, Composition, and Function. *Sports Heal.: A Multidiscip. Approach* **4**, 340–351 (2011).

140. Man, G. S. & Mologhianu, G. Osteoarthritis pathogenesis - a complex process that involves the entire joint. *J. Med. life* **7**, 37–41 (2013).

141. Tong, L. *et al.* Current understanding of osteoarthritis pathogenesis and relevant new approaches. *Bone Res.* **10**, 60 (2022).

142. Loeser, R. F., Goldring, S. R., Scanzello, C. R. & Goldring, M. B. Osteoarthritis: A disease of the joint as an organ. *Arthritis Rheum.* **64**, 1697–1707 (2012).

143. Wei, Y. & Bai, L. Recent advances in the understanding of molecular mechanisms of cartilage degeneration, synovitis and subchondral bone changes in osteoarthritis. *Connect. Tissue Res.* **57**, 245–261 (2016).

144. Robinson, W. H. *et al.* Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. *Nat. Rev. Rheumatol.* **12**, 580–592 (2016).

145. Barreto, G., Manninen, M. & Eklund, K. K. Osteoarthritis and Toll-Like Receptors: When Innate Immunity Meets Chondrocyte Apoptosis. *Biology* **9**, 65 (2020).

146. Sokolove, J. & Lepus, C. M. Role of inflammation in the pathogenesis of osteoarthritis: latest findings and interpretations. *Ther. Adv. Musculoskelet. Dis.* **5**, 77–94 (2013).

147. Goldring, M. B. Articular Cartilage Degradation in Osteoarthritis. *HSS J.* **8**, 7–9 (2012).

148. Cooke, M. E., Lawless, B. M., Jones, S. W. & Grover, L. M. Matrix degradation in osteoarthritis primes the superficial region of cartilage for mechanical damage. *Acta Biomater.* **78**, 320–328 (2018).

149. Li, G. *et al.* Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. *Arthritis Res. Ther.* **15**, 223 (2013).

150. Goldring, S. R. & Goldring, M. B. Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage–bone crosstalk. *Nat. Rev. Rheumatol.* **12**, 632–644 (2016).

151. Epsley, S. *et al.* The Effect of Inflammation on Bone. *Front. Physiol.* **11**, 511799 (2021).

152. Martin, E. *et al.* Incidental Meniscal Findings on Knee MRI in Middle-Aged and Elderly Persons. *N. Engl. J. Med.* **359**, 1108–1115 (2008).

153. Englund, M. *et al.* Meniscal tear in knees without surgery and the development of radiographic osteoarthritis among middle-aged and elderly persons: The multicenter osteoarthritis study. *Arthritis Rheum.* **60**, 831–839 (2009).

154. Roemer, F. W. *et al.* Semiquantitative assessment of focal cartilage damage at 3T MRI: a comparative study of dual echo at steady state (DESS) and intermediate-weighted (IW) fat suppressed fast spin echo sequences. *Eur. J. Radiol.* **80**, e126-31 (2010).

155. Roemer, F. W. *et al.* MRI-detected subchondral bone marrow signal alterations of the knee joint: terminology, imaging appearance, relevance and radiological differential diagnosis. *Osteoarthr. Cartil.* **17**, 1115–1131 (2009).

156. Englund, M. & Lohmander, L. S. Risk factors for symptomatic knee osteoarthritis fifteen to twenty-two years after meniscectomy. *Arthritis Rheum.* **50**, 2811–2819 (2004).

157. Englund, M., Roemer, F. W., Hayashi, D., Crema, M. D. & Guermazi, A. Meniscus pathology, osteoarthritis and the treatment controversy. *Nat Rev Rheumatol* **8**, 412–419 (2012).

158. Sinitcyn, P., Rudolph, J. D. & Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annu. Rev. Biomed. Data Sci.* **1**, 1–28 (2018).

159. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).

160. Wilson, T. H. B. A. . . . [and others] ; with problems by J. *Molecular biology of the cell*. (Fifth edition. New York : Garland Science, [2008] ©2008, 2008).

161. CRICK, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).

162. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17–R29 (2006).

163. Hartl, F. U. & Hayer-Hartl, M. Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science* **295**, 1852–1858 (2002).

164. Heijne, G. von. Signal sequences The limits of variation. *J. Mol. Biol.* **184**, 99–105 (1985).

165. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).

166. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261 (2003).

167. Robinson, C. V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450**, 973–982 (2007).

168. Olsen, C. A. Methods in Molecular Biology, Volume 615: Peptidomics, Methods and Protocols. Edited by Mikhail Soloviev. *Chembiochem* **11**, 1782–1782 (2010).

169. Bhutada, S. *et al.* Forward and reverse degradomics defines the proteolytic landscape of human knee osteoarthritic cartilage and the role of the serine protease HtrA1. *Osteoarthr Cartilage* (2022) doi:10.1016/j.joca.2022.02.622.

170. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

171. Bielajew, B. J. *et al.* Proteomic, mechanical, and biochemical characterization of cartilage development. *Acta Biomater.* **143**, 52–62 (2022).

172. Eyre, D. R. & Wu, J. J. Collagen structure and cartilage matrix integrity. *J. Rheumatol. Suppl.* **43**, 82–5 (1995).

173. Domon, B. & Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **312**, 212–217 (2006).

174. Cox, J. & Mann, M. Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. *Annu Rev Biochem* **80**, 273–299 (2011).

175. Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125 (2014).

176. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2021).

177. Deutsch, E. W. *et al.* The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **51**, D1539–D1548 (2022).

178. Miller, R. M. & Smith, L. M. Overview and considerations in bottom-up proteomics. *Analyst* **148**, 475–486 (2022).

179. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).

180. Olsen, J. V. & Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci.* **101**, 13417–13422 (2004).

181. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *PROTEOMICS* **11**, 2019–2026 (2011).

182. Zubarev, R. A., Zubarev, A. R. & Savitski, M. M. Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet? *J. Am. Soc. Mass Spectrom.* **19**, 753–761 (2008).

183. Gillet, L. C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **9**, 1–24 (2015).

184. Rozanova, S. *et al.* Quantitative Mass Spectrometry-Based Proteomics: An Overview. *Methods Mol. Biol. (Clifton, NJ)* **2228**, 85–116 (2021).

185. Doerr, A. DIA mass spectrometry. *Nat. Methods* **12**, 35–35 (2015).

186. Krasny, L. & Huang, P. H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omics* **17**, 29–42 (2020).

187. Strimbu, K. & Tavel, J. A. What are biomarkers? *Curr. Opin. HIV AIDS* **5**, 463–466 (2010).

188. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24**, 971–983 (2006).

189. Murphy, C. A. *et al.* The Meniscus in Normal and Osteoarthritic Tissues: Facing the Structure Property Challenges and Current Treatment Trends. *Annu. Rev. Biomed. Eng.* **21**, 1–27 (2019).

190. Roughley, P. J. & Mort, J. S. The role of aggrecan in normal and osteoarthritic cartilage. *J. Exp. Orthop.* **1**, 8 (2014).

191. Goldring, M. B. & Otero, M. Inflammation in osteoarthritis. *Curr. Opin. Rheumatol.* **23**, 471–478 (2011).

192. Cui, J. & Zhang, J. Cartilage Oligomeric Matrix Protein, Diseases, and Therapeutic Opportunities. *Int. J. Mol. Sci.* **23**, 9253 (2022).

193. Lohmander, L. S., Saxne, T. & Heinegård, D. K. Release of cartilage oligomeric matrix protein (COMP) into joint fluid after knee injury and in osteoarthritis. *Ann. Rheum. Dis.* **53**, 8 (1994).

194. Tseng, S., Reddi, A. H. & Cesare, P. E. D. Cartilage Oligomeric Matrix Protein (COMP): A Biomarker of Arthritis. *Biomark. Insights* **4**, BMI.S645 (2009).

195. Suri, S. & Walsh, D. A. Osteochondral alterations in osteoarthritis. *Bone* **51**, 204–211 (2012).

196. Stannus, O. *et al.* Circulating levels of IL-6 and TNF-α are associated with knee radiographic osteoarthritis and knee cartilage loss in older adults. *Osteoarthr. Cartil.* **18**, 1441–1447 (2010).

197. Spector, T. D. *et al.* Low-level increases in serum C-reactive protein are present in early osteoarthritis of the knee and predict progressive disease. *Arthritis Rheum.* **40**, 723–727 (1997).

198. Kaneko, S. *et al.* Interleukin–6 and interleukin–8 levels in serum and synovial fluid of patients with osteoarthritis. *Cytokines, Cell. Mol. Ther.* **6**, 71–79 (2009).

199. Fernández-Puente, P. *et al.* Identification of a Panel of Novel Serum Osteoarthritis Biomarkers. *J Proteome Res* **10**, 5095–5101 (2011).

200. Chisari, E., Yaghmour, K. M. & Khan, W. S. The effects of TNF-alpha inhibition on cartilage: a systematic review of preclinical studies. *Osteoarthr. Cartil.* **28**, 708–718 (2020).

201. Wang, X., Hunter, D., Xu, J. & Ding, C. Metabolic triggered inflammation in osteoarthritis. *Osteoarthr. Cartil.* **23**, 22–30 (2015).

202. Osta, B., Benedetti, G. & Miossec, P. Classical and Paradoxical Effects of TNF-α on Bone Homeostasis. *Front. Immunol.* **5**, 48 (2014).

203. Maeda, K. *et al.* Inflammation and Bone Metabolism in Rheumatoid Arthritis: Molecular Mechanisms of Joint Destruction and Pharmacological Treatments. *Int. J. Mol. Sci.* **23**, 2871 (2022).

204. Styrkarsdottir, U. *et al.* The CRTAC1 Protein in Plasma Is Associated With Osteoarthritis and Predicts Progression to Joint Replacement: A Large-Scale Proteomics Scan in Iceland. *Arthritis Rheumatol* **73**, 2025–2034 (2021).

205. Szilagyi, I. A. *et al.* Plasma proteomics identifies CRTAC1 as a biomarker for osteoarthritis severity and progression. *Rheumatol. (Oxf., Engl.)* **62**, 1286–1295 (2022).

206. Steck, E. *et al.* Chondrocyte secreted CRTAC1: A glycosylated extracellular matrix molecule of human articular cartilage. *Matrix Biol.* **26**, 30–41 (2007).

207. Ruiz-Romero, C., Fernández-Puente, P., Calamia, V. & Blanco, F. J. Lessons from the proteomic study of osteoarthritis. *Expert Rev Proteomic* **12**, 433–443 (2015).

208. Baker-LePain, J. C. & Lane, N. E. Role of bone architecture and anatomy in osteoarthritis. *Bone* **51**, 197–203 (2012).

209. Sellam, J. & Berenbaum, F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nat. Rev. Rheumatol.* **6**, 625–635 (2010).

210. Park, J. *et al.* Proteomic Analysis of the Meniscus Cartilage in Osteoarthritis. *Int J Mol Sci* **22**, 8181 (2021).

211. Englund, M., Guermazi, A. & Lohmander, L. S. The Meniscus in Knee Osteoarthritis. *Rheum Dis Clin N Am* **35**, 579–590 (2009).

212. Wang, L.-J., Zeng, N., Yan, Z.-P., Li, J.-T. & Ni, G.-X. Post-traumatic osteoarthritis following ACL injury. *Arthritis Res. Ther.* **22**, 57 (2020).

213. Ruiz-Romero, C. & Blanco, F. J. Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis. *Osteoarthr. Cartil.* **18**, 500–509 (2010).

214. Vangsness, C. T., Burke, W. S., Narvy, S. J., MacPhee, R. D. & Fedenko, A. N. Human knee synovial fluid cytokines correlated with grade of knee osteoarthritis--a pilot study. *Bulletin Nyu Hosp Jt Dis* **69**, 122–7 (2011).

215. Anderson, J. R., Phelan, M. M., Clegg, P. D., Peffers, M. J. & Rubio-Martinez, L. M. Synovial Fluid Metabolites Differentiate between Septic and Nonseptic Joint Pathologies. *J. Proteome Res.* **17**, 2735–2743 (2018).

216. Mateos, J. *et al.* Differential protein profiling of synovial fluid from rheumatoid arthritis and osteoarthritis patients using LC–MALDI TOF/TOF. *J. Proteom.* **75**, 2869–2878 (2012).

217. Anderson, J. R. *et al.* Optimization of Synovial Fluid Collection and Processing for NMR Metabolomics and LC-MS/MS Proteomics. *J. Proteome Res.* **19**, 2585–2597 (2020).

218. Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, (2019).

219. Anderson, J. R., Phelan, M. M., Foddy, L., Clegg, P. D. & Peffers, M. J. Ex Vivo Equine Cartilage Explant Osteoarthritis Model: A Metabolomics and Proteomics Study. *J Proteome Res* **19**, 3652–3667 (2020).

220. McNulty, A. L., Rothfusz, N. E., Leddy, H. A. & Guilak, F. Synovial fluid concentrations and relative potency of interleukin-1 alpha and beta in cartilage and meniscus degradation. *J Orthopaed Res* **31**, 1039–1045 (2013).

221. Geurts, J., Jurić, D., Müller, M., Schären, S. & Netzer, C. Novel Ex Vivo Human Osteochondral Explant Model of Knee and Spine Osteoarthritis Enables Assessment of Inflammatory and Drug Treatment Responses. *Int J Mol Sci* **19**, 1314 (2018).

222. Chang, W. *et al. shiny: Web Application Framework for R*. (2021).

223. Wadi, L., Meyer, M., Weiser, J., Stein, L. D. & Reimand, J. Impact of outdated gene annotations on pathway enrichment analysis. *Nat Methods* **13**, 705–6 (2016).

224. Folkesson, E. *et al.* Proteomic characterization of the normal human medial meniscus body using data-independent acquisition mass spectrometry. *J Orthop Res* (2020) doi:10.1002/jor.24602.

225. Mueller, M., Martens, L. & Apweiler, R. Annotating the human proteome: Beyond establishing a parts list. *Biochim. Biophys. Acta (BBA) - Proteins Proteom.* **1774**, 175–191 (2007).

226. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**, D503–D509 (2014).

227. Rydén, M. proteasy: Protease Mapping. (2022).

228. Lee, C., Bongcam-Rudloff, E., Sollner, C., Jahnen-Dechent, W. & Claesson-Welsh, L. Type 3 cystatins; fetuins, kininogen and histidine-rich glycoprotein. *Front Biosci* **Volume**, 2911 (2009).

229. Manderson, G. A. *et al.* Interactions of histidine-rich glycoprotein with immunoglobulins and proteins of the complement system. *Mol. Immunol.* **46**, 3388–3398 (2009).

230. Dabrowska, A. M., Tarach, J. S., Wojtysiak-Duma, B. & Duma, D. Fetuin-A (AHSG) and its usefulness in clinical practice. Review of the literature. *Biomed. Pap.* **159**, 352–359 (2015).

231. Cochran, B. J., Ong, K.-L., Manandhar, B. & Rye, K.-A. APOA1: a Protein with Multiple Therapeutic Functions. *Curr. Atheroscler. Rep.* **23**, 11 (2021).

232. Gao, S. *et al.* Histidine-rich glycoprotein ameliorates endothelial barrier dysfunction through regulation of NF-κB and MAPK signal pathway. *Br. J. Pharmacol.* **176**, 2808–2824 (2019).

233. Seto, J. *et al.* Accelerated Growth Plate Mineralization and Foreshortened Proximal Limb Bones in Fetuin-A Knockout Mice. *PLoS ONE* **7**, e47338 (2012).

234. Ouimet, M., Barrett, T. J. & Fisher, E. A. HDL and Reverse Cholesterol Transport. *Circ. Res.* **124**, 1505–1518 (2019).

235. Wu, M., Jia, B. & Li, M. Complement C3 and Activated Fragment C3a Are Involved in Complement Activation and Anti-Bacterial Immunity. *Front. Immunol.* **13**, 813173 (2022).

236. Zollinger, A. J. & Smith, M. L. Fibronectin, the extracellular glue. *Matrix Biol.* **60**, 27–37 (2017).

237. Proud, D. & Kaplan, A. P. Kinin Formation: Mechanisms and Role in Inflammatory Disorders. *Annu. Rev. Immunol.* **6**, 49–83 (1988).

238. Vries, P. S. de *et al.* Whole-genome sequencing study of serum peptide levels: the Atherosclerosis Risk in Communities study. *Hum. Mol. Genet.* **26**, 3442–3450 (2017).

239. Assirelli, E. *et al.* Complement Expression and Activation in Osteoarthritis Joint Compartments. *Front. Immunol.* **11**, 535010 (2020).

240. Struglics, A. *et al.* The complement system is activated in synovial fluid from subjects with knee injury and from patients with osteoarthritis. *Arthritis Res. Ther.* **18**, 223 (2016).

241. To, W. S. & Midwood, K. S. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. *Fibrogenesis Tissue Repair* **4**, 21 (2011).

242. Chevalier, X. Fibronectin, cartilage, and osteoarthritis. *Semin Arthritis Rheu* **22**, 307–318 (1993).

243. Rex, D. A. B. *et al.* A modular map of Bradykinin-mediated inflammatory signaling network. *J. Cell Commun. Signal.* **16**, 301–310 (2022).

244. Galvão, I., Sugimoto, M. A., Vago, J. P., Machado, M. G. & Sousa, L. P. Immunopharmacology and Inflammation. 3–32 (2018) doi:10.1007/978-3-319-77658-3_1.

245. Arooj, M. *et al.* Finding off-targets, biological pathways, and target diseases for chymase inhibitors via structure-based systems biology approach: Finding Off-Targets for Chymase Inhibitors. *Proteins: Struct., Funct., Bioinform.* **83**, 1209–1224 (2015).

246. Dvorak, H. F., Brown, L. F., Detmar, M. & Dvorak, A. M. Vascular permeability factor/vascular endothelial growth factor, microvascular hyperpermeability, and angiogenesis. *Am. J. Pathol.* **146**, 1029–39 (1995).

247. Vincenti, M. P. & Brinckerhoff, C. E. Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors. *Arthritis Res. Ther.* **4**, 157 (2002).

248. Li, H., Wang, D., Yuan, Y. & Min, J. New insights on the MMP-13 regulatory network in the pathogenesis of early osteoarthritis. *Arthritis Res. Ther.* **19**, 248 (2017).

249. Pap, T. & Korb-Pap, A. Cartilage damage in osteoarthritis and rheumatoid arthritis—two unequal siblings. *Nat. Rev. Rheumatol.* **11**, 606–615 (2015).

250. Ross, D. & Siegel, D. Functions of NQO1 in Cellular Protection and CoQ10 Metabolism and its Potential Role as a Redox Sensitive Molecular Switch. *Front. Physiol.* **8**, 595 (2017).

251. Jin, J. *et al.* Aldehyde Dehydrogenase 2 Ameliorates LPS-Induced Acute Kidney Injury through Detoxification of 4-HNE and Suppression of the MAPK Pathway. *J. Immunol. Res.* **2023**, 5513507 (2023).

252. Gong, D., Zhang, H. & Hu, S. Mitochondrial aldehyde dehydrogenase 2 activation and cardioprotection. *J. Mol. Cell. Cardiol.* **55**, 58–63 (2013).

253. Liu, L. *et al.* The role of oxidative stress in the development of knee osteoarthritis: A comprehensive research review. *Front. Mol. Biosci.* **9**, 1001212 (2022).

254. Gu, J.-Y., Han, F., Chen, S.-Y. & Zhang, Q. Research progress and hot spot analysis related to oxidative stress and osteoarthritis: a bibliometric analysis. *BMC Musculoskelet. Disord.* **24**, 411 (2023).

255. Courties, A., Sellam, J. & Berenbaum, F. Metabolic syndrome-associated osteoarthritis. *Curr. Opin. Rheumatol.* **29**, 214–222 (2017).

256. Duell, E. J. *et al.* Genetic variation in alcohol dehydrogenase (ADH1A , ADH1B , ADH1C , ADH7) and aldehyde dehydrogenase (ALDH2), alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Carcinogenesis* **33**, 361–367 (2012).

257. Ansell, R., Granath, K., Hohmann, S., Thevelein, J. M. & Adler, L. The two isoenzymes for yeast NAD+-dependent glycerol 3-phosphate dehydrogenase encoded by GPD1 and GPD2 have distinct roles in osmoadaptation and redox regulation. *EMBO J.* **16**, 2179–2187 (1997).

258. Sebastian, K. *et al.* Characterization of SLCO5A1/OATP5A1, a Solute Carrier Transport Protein with Non-Classical Function. *PLoS ONE* **8**, e83257 (2013).

259. Shan, Y. *et al.* Targeting HIBCH to reprogram valine metabolism for the treatment of colorectal cancer. *Cell Death Dis.* **10**, 618 (2019).

260. Xu, M. *et al.* Molecular mechanism of valine and its metabolite in improving triglyceride synthesis of porcine intestinal epithelial cells. *Sci. Rep.* **13**, 2933 (2023).

261. Zanini, I. M. Y., Soneson, C., Lorenzi, L. E. & Azzalin, C. M. Human cactin interacts with DHX8 and SRRM2 to assure efficient pre-mRNA splicing and sister chromatid cohesion. *J. Cell Sci.* **130**, 767–778 (2017).

262. Felisberto-Rodrigues, C. *et al.* Structural and functional characterisation of human RNA helicase DHX8 provides insights into the mechanism of RNA-stimulated ADP release. *Biochem. J.* **476**, 2521–2543 (2019).

263. Snyder, E. M. *et al.* APOBEC1 complementation factor (A1CF) is dispensable for C-to-U RNA editing in vivo. *RNA* **23**, 457–465 (2017).

264. Nikolaou, K. C. *et al.* The RNA-Binding Protein A1CF Regulates Hepatic Fructose and Glycerol Metabolism via Alternative RNA Splicing. *Cell Rep.* **29**, 283-300.e8 (2019).

265. Appel, L.-M. *et al.* PHF3 regulates neuronal gene expression through the Pol II CTD reader domain SPOC. *Nat. Commun.* **12**, 6078 (2021).

266. Feng, H., Danfelter, M., Strömqvist, B. & Heinegård, D. Extracellular Matrix in Disc Degeneration. *J. Bone Jt. Surg.* **88**, 25–29 (2006).

267. Hardingham, T. E., Fosang, A. J. & Dudhia, J. The structure, function and turnover of aggrecan, the large aggregating proteoglycan from cartilage. *Eur. J. Clin. Chem. Clin. Biochem. : J. Forum Eur. Clin. Chem. Soc.* **32**, 249–57 (1994).

268. Pérez-García, S. *et al.* Profile of Matrix-Remodeling Proteinases in Osteoarthritis: Impact of Fibronectin. *Cells* **9**, 40 (2019).

269. Chen, Y. *et al.* HAPLN1 Affects Cell Viability and Promotes the Pro-Inflammatory Phenotype of Fibroblast-Like Synoviocytes. *Front. Immunol.* **13**, 888612 (2022).

270. Wilson, R. *et al.* Comprehensive Profiling of Cartilage Extracellular Matrix Formation and Maturation Using Sequential Extraction and Label-free Quantitative Proteomics*. *Mol. Cell. Proteom.* **9**, 1296–1313 (2010).

271. Chen, Q., Zhang, Y., Johnson, D. M. & Goetinck, P. F. Assembly of a Novel Cartilage Matrix Protein Filamentous Network: Molecular Basis of Differential Requirement of von Willebrand Factor A Domains. *Mol. Biol. Cell* **10**, 2149–2162 (1999).

272. Gao, J.-B. *et al.* Fibulin-5 protects the extracellular matrix of chondrocytes by inhibiting the Wnt/β-catenin signaling pathway and relieves osteoarthritis. *Eur. Rev. Méd. Pharmacol. Sci.* **24**, 5249–5258 (2020).

273. Karlsson, C. *et al.* Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthr. Cartil.* **18**, 581–592 (2010).

274. Papke, C. L. & Yanagisawa, H. Fibulin-4 and fibulin-5 in elastogenesis and beyond: Insights from mouse and human studies. *Matrix Biol.* **37**, 142–149 (2014).

275. McShane, B. B., Gal, D., Gelman, A., Robert, C. & Tackett, J. L. Abandon Statistical Significance. *Am Statistician* **73**, 235–245 (2019).

276. Staggs, V. S. Why statisticians are abandoning statistical significance. *Res. Nurs. Heal.* **42**, 159–160 (2019).

277. Kerin, A., Patwari, P., Kuettner, K., Cole, A. & Grodzinsky, A. Molecular basis of osteoarthritis: biomechanical aspects. *Cell. Mol. Life Sci. CMLS* **59**, 27–35 (2002).

278. Pavelka, N. *et al.* Statistical Similarities between Transcriptomics and Quantitative Shotgun Proteomics Data *. *Mol. Cell. Proteom.* **7**, 631–644 (2008).

279. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).

280. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220–220 (2010).

Paper I

Bioinformatics, 37(20), 2021, 3491–3493
doi: 10.1093/bioinformatics/btab373
Advance Access Publication Date: 12 May 2021
Original Paper

OXFORD

Structural bioinformatics

ProteoMill: efficient network-based functional analysis portal for proteomics data

Martin Rydén [1,*], **Martin Englund**[1,†] **and Neserin Ali**[2,†]

[1]Department of Clinical Sciences Lund, Orthopedics, Clinical Epidemiology Unit, Lund University, SE-22185 Lund, Sweden and [2]Department of Clinical Sciences Lund, Rheumatology and Molecular Skeletal Biology, Lund University, SE-22184 Lund, Sweden

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the last two authors should be regarded as Joint Last Authors.

Associate Editor: Pier Luigi Martelli

Received on February 7, 2021; revised on April 25, 2021; editorial decision on May 6, 2021; accepted on May 11, 2021

Abstract

Summary: Functional analysis has become a common approach to incorporate biological knowledge into the analysis of omics data, and to explore molecular events that govern a disease state. It is though only one step in a wider analytical pipeline that typically requires use of multiple individual analysis software. There is currently a need for a well-integrated omics analysis tool that performs all the steps. The ProteoMill portal is developed as an R Shiny application and integrates all necessary steps from data-upload, converting identifiers, to quality control, differential expression and network-based functional analysis into a single fast, interactive easy to use workflow. Further, it maintains annotation data sources up to date, overcoming a common problem with use of outdated information and seamlessly integrates multiple R-packages for an improved user-experience. The functionality provided in this software can benefit researchers by facilitating the exploratory analysis of proteomics data.

Availability and implementation: ProteoMill is available at https://proteomill.com.

Contact: martin.ryden@med.lu.se

1 Introduction

The large amounts of data generated from omics experiments have stressed the need for methods to reveal and extract critical components of dynamic biological systems in a readable manner, which connects to the specific study question. Expression data that are derived from high throughput analysis have multiple levels of biological features connected to it. In a real biological environment, the physical, genetic, regulatory and functional properties of a molecular set work together in a response to environmental stimuli. Holistically evaluating these attributes is a way to reveal the intercommunication between these properties and to provide a biological context. However, this task encompasses some impending challenges, including differences in biomolecule identification, data dimensionality reduction, biological contextualization, statistical analysis and data visualization and this differs among the various types of individual datasets.

Existing omics analysis tools are typically specialized for individual parts of the analysis workflow and differences in data format standards means the tools do not integrate well when used as part of an analysis workflow. This requires the researcher not only to have knowledge of the different individual software, but also knowing how to format the generated output from one software for use in the next software. This often poses a time-consuming task, particularly for researchers with little computational experience or little experience with the software(s) in question and is prone to errors.

Omics analysis platforms such as Perseus (Tyanova and Cox, 2018) and Qlucore (Qlucore, 2021) offer thorough analytical and explorative features, but require users to download and install their software and is not open source. While there are many existing web-based omics tools which are able to perform individual parts of an analysis workflow (Efstathiou *et al.*, 2017; Kuleshov *et al.*, 2016; Luo *et al.*, 2017; Merico *et al.*, 2010; Perlasca *et al.*, 2019; Schweppe *et al.*, 2017; Zheng and Wang, 2008), many lack the ability to perform complete pipelines in fast, interactive web-environments. Reimand *et al.* lists the protocols and time consumption for popular enrichment software, with the time expense ranging from minutes to several hours (Reimand *et al.*, 2019). In contrast, the run time for ProteoMill functions are a few seconds at the most, as described in Table 1.

Another important but often overlooked aspect for generating reliable and biologically relevant results is the quality of annotation data, and, by extension, a tool's ability to maintain annotation data sources up to date. Lina Wadi *et al.* reported that 67% of publications in their survey referenced software using outdated annotation data (Wadi *et al.*, 2016). Web-based tools have an inherit advantage in that back-end data sources can be dynamically updated without requiring manual action by the user (such as downloading and installing software).

Analysis of proteomic data faces additional challenges (Kirik *et al.*, 2012). Different gene- and protein level identifier types are utilized in the various omics tools, which often require the researcher to convert between identifier types before proceeding to the next step of the analysis. This can result in loss of data since there can exist one-to-many mappings between two identifier types or that an identifier cannot be mapped between two identifier types (Reimand *et al.*, 2019). Furthermore, a frequent concern in mass spectrometry-derived data is the abundance of missing values (Lazar *et al.*, 2016; Wang *et al.*, 2017).

Thus, a tool that could help to transform the biological research into integrated framework is preferred. The aim of this study is to describe a newly developed software that addresses many of the existing shortcomings. The fundamental concepts of this software are to provide sets of well-integrated, easy-to-use and to a large extent automated functions for exploratory analysis of proteomic data.

2 Materials and methods

2.1 Architecture

ProteoMill runs as a web application using Shiny Server and is hosted on Amazon Web Services. The software is developed in R (version 3.6.1) and the interface was created using the R-package Shiny (Chang *et al.*, 2021) and shinydashboard (Chang and Ribeiro, 2018) (version 0.7.1) with a customized CSS theme. Animations were created using jQuery and the library animejs. Plotly (Plotly Technologies Inc., 2015), ggplot2 (Wickham, 2009), heatmaply (Galili *et al.*, 2018), networkD3 (Allaire *et al.*, 2017) and visNetwork (Almende et al., 2019) were used for plotting.

2.2 Identifier conversion

The Bioconductor packages AnnotationDbi (Pagès *et al.*, 2020) and ensembldb (Rainer *et al.*, 2019) was used for converting between identifiers. The identifier type of the user's uploaded data is automatically recognized and converted to four different identifier types (where applicable). This way, the user can choose to display protein labels as any of the five identifier types, but do not need to worry about manually converting between identifiers.

2.3 Data quality control

Principal component analysis (PCA) was implemented using the R-package stats. Another package, mixOmics, was used for multilevel PCA.

2.4 Differential expression analysis

Two R-packages, limma (Ritchie *et al.*, 2015) and DESeq2 (Love *et al.*, 2014) were implemented for differential expression analysis. Each package is commonly used for fitting gene-wise linear models to expression data. limma was originally developed with a primary focus on the analysis of microarray data, while DESeq2 for the analysis of RNA-seq data and is based on the negative binomial distribution.

Differential expression analysis is conducted by specifying two contrasts and choosing a paired or non-paired design. The results are evaluated by inspecting the table in the 'Differential expression' tab.

The results are displayed as estimated by the specific software, using the software's default settings for shrinkage parameters, correction for multiple testing, significance level and etc. For example, the correction for multiple testing is done using the Benjamini–Hochberg method and is applied to the tests performed within one run of the analysis and not with respect to all tests performed within one family of hypotheses in a study, which sometimes may be misleading (Ranstam, 2016). The user needs to verify if these setting are appropriate for the specific analysis done.

2.5 Functional enrichment and network analysis

The hypergeometric distribution was used to calculate the probability of protein list overlap.

$$P = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad (1)$$

In this formula, N is the total number of proteins in the background distribution, M is the number of proteins in the background distribution annotated to a pathway, n is the total number of selected proteins of interest and x are the proteins of interest annotated to a pathway.

Pathway data and interaction data are dynamically collected from Reactome (Fabregat *et al.*, 2018) and STRING (Szklarczyk *et al.*, 2015) (https://reactome.org/download-data). MD5sum hashes are used to ensure that the local database is up to date.

For each entry in the main pathway data file, the top-level parent pathway was annotated. This was done by creating a directed acyclic graph object using the R-package igraph (Csárdi and Nepusz, 2006).

2.6 Data sources

An important aspect of this software is to maintain data sources up to date. This is done by using an automated workflow at a bi-monthly interval. Data are collected from the two primary data sources, Reactome (Fabregat *et al.*, 2018) for pathway data and STRING (Szklarczyk *et al.*, 2015) for protein interaction data. These data are then structured to a predefined format, making it possible to integrate them in the analysis.

3 Results

The presented software, ProteoMill, proposes a unique approach to conducting explorative analysis of proteomics data. The data visualization capabilities present in this software are designed to make it possible even for researchers without any particular computational training to gain insights about the biological meaning of their data. Many of the graphical components are interactive, which is a useful feature for analysing protein interactions and selecting subnetworks of interest.

A common goal in many of ProteoMill's functionalities is to reduce data complexity, and to provide a framework for extracting elements of biological relevance. PCA reduces a dataset of hundreds or thousands of expression datapoints into a single datapoint for each condition, plotted in 2–3 principal components, which in turn describes the dimensions with largest variability. The datapoints cluster together based on the similarity of their expression profiles.

Categorizing proteins into biological entities, described as pathways, is another way to reduce complexity and make sense of one's data. Network graphs produced from interaction data can be difficult to interpret. In ProteoMill, pathways are used to categorize and label groups of interacting proteins, and as a way to inspect subnetworks based on these common biological themes.

The integrated enrichment- and network analysis provides a way for users to simultaneously explore functional analysis output and interaction data, and this feature has been specifically designed to easily identify and select subnetworks of interest for further analysis.

3.1 Reproducibility

ProteoMill supports the use of reproducibility tokens as a simple way to load settings and database versions from a prior session. The token contains information about all user defined settings that affect the outcome of the analysis—every statistical result and its graphical representations. The token also contains an MD5sum hash for the uploaded dataset and warns the user if the uploaded file is not identical to the file used in the previous session.

3.2 Performance

To assess the performance of ProteoMill, we measured the execution speed of its most prominent functions directly on the server (Table 1), using a publicly available dataset consisting of 12 samples and 12 320 proteins (Wertheim *et al.*, 2009). The time elapsed for rendering plots depends on the client-side machine and browser. The column labeled

Table 1. Benchmark results of ProteoMill functions

Function	Exec. time (ms)	Total time (ms)
Upload data	2560	—
Set missing value cutoff	90	—
Differential expression (limma)	215	—
Differential expression (DESeq2)	5042	—
Pathway over-representation	367	—
Interaction network[a]	53	1041
Interaction network[b]	58	5580

[a]Using 162 nodes and 582 edges.

[b]Using 1356 nodes and 17 718 edges.

'Exec. Time' describe the elapsed time of server-side calculations/data sub-setting operations and the column 'Total time' also describes the rendering time as measured on a 2018 MacBook Pro (2.2 GHz 6-Core Intel Core i7).

4 Discussion

The integrated features in this software provide powerful visualization strategies for the exploration of omics data, with a particular focus on the management and manipulation of proteomics data. By using this platform, researchers can expect to discover biologically relevant rendering of their data through results aggregated from reliable and up-to-date data sources.

The software offers innovative strategies to interactively explore quantitative proteomics data in a comprehensive workflow from data-upload to network analysis. It has a strong focus on well-maintained data sources, computational efficiency and user-friendliness.

Importantly, ProteoMill utilizes many existing R packages for statistical analysis and pathway annotation that are standard in the field. However, these methods are strongly focused on estimation of P-values and classifications of results based on P-value thresholds. This is an unfavorable approach to use of statistical methods and there is a need to move further in better estimation methods and expressing uncertainty (Benjamin *et al.*, 2018).

Acknowledgements

The authors would like to thank Aleksandra Turkiewicz, PhD for helpful discussions and critical reading of this paper.

Funding

This work was supported by 'The Swedish Research Council (Vetenskapsrådet), grant number 542-2014-2348' and 'Kockska stiftelsen för medicinsk forskning (Fromma)'.

Conflict of Interest: none declared.

Data availability

The tool itself is available on https://proteomill.com/. All demonstrational data can be downloaded directly from the website by clicking Demo data, selecting a dataset and clicking 'Download'.

Code availability

The source code is available at https://github.com/martinry/ProteoMill/.

References

Allaire,J.J. *et al.* (2017) NetworkD3: D3 JavaScript network graphs from R. Last accessed 23 April, 2021.

Almende, B.V. *et al.* (2019) visNetwork: network visualization using 'vis.js' Library. Last accessed 23 April, 2021.

Benjamin,D.J. *et al.* (2018) Redefine statistical significance. *Nat. Hum. Behav.*, **2**, 6–10.

Chang,W. *et al.* (2021) Shiny: web application framework for R. Last accessed 23 April, 2021.

Chang,W. and Ribeiro,B.B. (2018) shinydashboard: create dashboards with 'Shiny'. Last accessed 23 April, 2021.

Csárdi,G. and Nepusz,T. (2006) The igraph software package for complex network research.

Efstathiou,G. *et al.* (2017) ProteoSign: an end-user online differential proteomics statistical analysis platform. *Nucleic Acids Res.*, **45**, gkx444.

Fabregat,A. *et al.* (2018) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.

Galili,T. *et al.* (2018) Heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, **34**, 1600–1602.

Kirik,U. *et al.* (2012) Multimodel pathway enrichment methods for functional evaluation of expression regulation. *J. Proteome Res.*, **11**, 2955–2967.

Kuleshov,M.V. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.

Lazar,C. *et al.* (2016) Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J. Proteome Res.*, **15**, 1116–1125.

Love,M.I. *et al.* (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

Luo,W. *et al.* (2017) Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res.*, **45**, W501–W508.

Merico,D. *et al.* (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, **5**, e13984.

Pagès,H. *et al.* (2020) AnnotationDbi: manipulation of SQLite-based annotations in Bioconductor. Last accessed 23 April, 2021.

Perlasca,P. *et al.* (2019) UNIPred-Web: a web tool for the integration and visualization of biomolecular networks for protein function prediction. *BMC Bioinformatics*, **20**, 422.

Plotly Technologies Inc. (2015) *Collaborative Data Science*. Montréal, QC: Plotly Technologies Inc.

Qlucore AB. (2021) Qlucore omics explorer.

Rainer,J. *et al.* (2019) ensembldb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics*, **35**, 3151–3153.

Ranstam,J. (2016) Multiple P-values and Bonferroni correction. *Osteoarthr. Cartilage*, **24**, 763–764.

Reimand,J. *et al.* (2019) Pathway enrichment analysis and visualization of omics data using g: profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.*, **14**, 482–517.

Ritchie,M.E. *et al.* (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.

Schweppe,D.K. *et al.* (2017) BioPlex display: an interactive suite for large--scale, AP-MS protein-protein interaction data. *J Proteome Res.* **17**, 722–726.

Szklarczyk,D. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.

Tyanova,S. and Cox,J. (2018) Cancer systems biology: methods and protocols. *Methods Mol. Biol.*, **1711**, 133–148.

Wadi,L. *et al.* (2016) Impact of outdated gene annotations on pathway enrichment analysis. *Nat. Methods*, **13**, 705–706.

Wang,J. *et al.* (2017) In-depth method assessments of differentially expressed protein detection for shotgun proteomics data with missing values. *Sci. Rep.-uk*, **7**, 3367.

Wertheim,G.B.W. *et al.* (2009) The Snf1-related kinase, Hunk, is essential for mammary tumor metastasis. *Proc. Natl. Acad. Sci.*, **106**, 15855–15860.

Wickham,H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Zheng,Q. and Wang,X.-J. (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res.*, **36**, W358–W363.

Paper II

Received: 29 January 2023 | Revised: 11 April 2023 | Accepted: 4 May 2023

DOI: 10.1002/pmic.202300040

Proteomics
Proteomics and Systems Biology

RESEARCH ARTICLE

Identification and quantification of degradome components in human synovial fluid reveals an increased proteolytic activity in knee osteoarthritis patients vs controls

Martin Rydén[1] | Aleksandra Turkiewicz[1] | Patrik Önnerfjord[2] | Jon Tjörnstrand[1,3] | Martin Englund[1] | Neserin Ali[1]

[1]Faculty of Medicine, Department of Clinical Sciences, Lund, Orthopedics, Clinical Epidemiology Unit, Lund University, Lund, Sweden

[2]Faculty of Medicine, Department of Clinical Sciences, Lund, Rheumatology and Molecular Skeletal Biology, Lund University, Lund, Sweden

[3]Department of Orthopedics, Skåne University Hospital, Lund, Sweden

Correspondence
Martin Rydén, Faculty of Medicine, Department of Clinical Sciences, Lund University, Lund, Sweden.
Email: martin.ryden@med.lu.se

Funding information
H2020 European Research Council, Grant/Award Number: 771121; Reumatikerförbundet; Österlund Foundation; Foundation for Research in Rheumatology; Vetenskapsrådet; Greta och Johan Kocks stiftelser; Faculty of Medicine, Lund University, Sweden; Governmental Funding of Clinical Research program within the National Health Service (ALF); IAB Lundberg Foundation

Abstract

Synovial fluid (SF) may contain cleavage products of proteolytic activities. Our aim was to characterize the degradome through analysis of proteolytic activity and differential abundance of these components in a peptidomic analysis of SF in knee osteoarthritis (OA) patients versus controls ($n = 23$). SF samples from end-stage knee osteoarthritis patients undergoing total knee replacement surgery and controls, that is, deceased donors without known knee disease were previously run using liquid chromatography mass spectrometry (LC-MS). This data was used to perform new database searches generating results for non-tryptic and semi-tryptic peptides for studies of degradomics in OA. We used linear mixed models to estimate differences in peptide-level expression between the two groups. Known proteolytic events (from the MEROPS peptidase database) were mapped to the dataset, allowing the identification of potential proteases and which substrates they cleave. We also developed a peptide-centric R tool, *proteasy*, which facilitates analyses that involve retrieval and mapping of proteolytic events. We identified 429 differentially abundant peptides. We found that the increased abundance of cleaved APOA1 peptides is likely a consequence of enzymatic degradation by metalloproteinases and chymase. We identified metalloproteinase, chymase, and cathepsins as the main proteolytic actors. The analysis indicated increased activity of these proteases irrespective of their abundance.

KEYWORDS
degradomics, osteoarthritis, peptidomics, proteomics, synovial fluid

Abbreviations: A2M, alpha-2-macroglobulin; ACAN, aggrecan core protein; ACTB, actin; ADAMTS, a disintegrin and metalloproteinase with thrombospondin motifs; AHSG, alpha-2-HS-glycoprotein; ALB, albumin; APOA1, apolipoprotein A-I; APOB, apolipoprotein B; C3, complement C3; CFB, complement factor B; CLUAP1, clusterin-associated protein 1; CRTAC1, cartilage acidic protein 1; DCN, decorin; DIA, data-independent acquisition; ECM, extracellular matrix; FGA, fibrinogen alpha chain; FN1, fibronectin; HRG, histidine-rich glycoprotein; IL, interleukin; ITIH1, inter-alpha-trypsin inhibitor; ITIH2, inter-alpha-trypsin inhibitor; KNG1, kininogen-1; MMP, matrix metalloproteinase; OA, osteoarthritis; PEBP1, phosphatidylethanolamine-binding protein 1; PLG, plasminogen; SERPINA3, alpha-1-antichymotrypsin; SF, synovial fluid; SOD3, extracellular superoxide dismutase; TNF, tissue necrosis factor.

1 | INTRODUCTION

Osteoarthritis (OA) is a degenerative joint disease and a common cause of joint damage. It is especially common in middle-aged and elderly people and is estimated to affect more than 25% of adults [1]. Currently there are no therapies which can cure or prevent the disease. The pathological changes that characterize knee osteoarthritis (OA) are typically the gradual breakdown of articular cartilage, low-grade inflammation of the synovium, bone alterations, and meniscus degeneration [2]. Although many factors contribute to the disease progression, proteases, and inflammatory cytokines are considered a main contributor to the catabolic processes of OA, including interleukin 1β (IL-1β), tissue necrosis factor-α (TNF-α), IL-6, IL-12, IL-15 contributing to the increased expression of matrix metalloproteinases (MMPs), which in turn drive extracellular matrix (ECM) degradation [3–6].

Using targeted approaches limited to specific proteases have previously been reported to be involved in the progression of OA [4, 7]. However, a screening approach that analyses all the active proteases has to the best of our knowledge not been reported. Thus, the relative abundance of peptides in OA and its associated protease activity remains understudied. Further research in this area is needed to better understand the role of proteases in OA disease progression.

Synovial fluid (SF) is in direct contact with cartilage, synovium, and meniscus in the joint cavities where it acts like a lubricant by reducing friction and has additional metabolic and regulatory functions [8].

Changes in SF composition has previously been described to be associated with joint disease [9–11] but only few proteomics studies comparing SF composition in knee OA patients versus healthy controls exist [12–17].

Peptidomics can provide information regarding the proteolytic activity that generated the observed endogenously cleaved protein fragments [18]. The use of quantitative peptidomics allows studies on relative peptide levels as a result of an increased or decreased proteolytic activity. The standard peptidomics approach is to isolate peptides from the rest of the protein bulk, for example, by ethanol precipitation, where smaller peptides stay in the supernatant while larger peptides and proteins end up in the protein pellet, or by ultrafiltration where the filtrate contains the peptidome. This usually results in very few stable peptides that could be further used in a quantitative manner [19, 20]. Studying the cleaving sites of relatively intact proteins may be seen as a lost treasure that is usually overlooked in proteomics due to the need for artificial cleavages, usually by trypsin, to be able to detect peptides with a bottom-up mass spectrometry (MS) method.

We hypothesized that endogenously cleaved protein fragments in SF could reflect the changed protease activity and provide important new insights about the OA disease process. Thus, our aim was to characterize the degradome of human knee SF in knee OA versus controls. Specifically, we aimed to (1) identify and quantify endogenously cleaved peptides potentially involved in the OA process, (2) identify proteases whose proteolytic activity may impact the host proteins of these peptides.

Significance of the Study

This study provides a broad array of differentially abundant endogenously cleaved peptides and their potential cleaving actor in human SF. This study demonstrates that the proteolytic activity of the predicted proteases extends beyond the extracellular matrix of the surrounding tissues and can also affect factors such as chylomicron assembly potentially leading to a hampered homeostasis.

2 | MATERIALS AND METHODS

2.1 | Human synovial fluid samples

In this study we reused the raw MS data files from a previous study [14] to extract semi- and non-tryptic peptide data. Briefly, the raw files were generated from SF obtained from end-stage medial compartment knee OA patients undergoing knee arthroplasty in the year 2017 in the Skåne region, Sweden, (n = 11, age range 55–80 years, eight women and three men) and deceased (between year 2017 and 2018, from the same geographical area as the patients) human donors without known chronic knee disease (n = 12, age range 19–79 years, five women and seven men) [14]. The latter group will hereafter be referred to as controls. Informed consent has been obtained for all samples included in this analysis. The sample collection and analysis have been approved by the ethical review committee of Lund and have been carried out in accordance with relevant guidelines and regulations by the Declaration of Helsinki principles. Only SF samples in which we did not detect visible blood contamination were included. One of the original 13 control samples was discarded due to a random error during the identification search in Peaks Studio X.

2.2 | LC-MS/MS proteomics

The samples were analyzed using a nanoLC-system (EASY-nLC 1000) coupled to a mass spectrometer (Thermo Scientific Q-Exactive HFX™) using data-independent acquisition (DIA) [14]. The raw MS files has been deposited to ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023708. In the referenced study we reported analysis of proteomics composition based on tryptic peptides. In the current study, we focus on a new aim, targeting endogenously cleaved peptides, that is, semi- and non-tryptic peptides only. For this purpose, we reused the raw MS data and searched it with no enzyme specification so that semi-tryptic and non-tryptic peptides could be extracted. The MS data was searched in the Peaks Studio X software with non-specific cleave sites specified, and post-translational modifications carbamidomethylation (C), oxidation (M), and deamidation (NQ) were selected. Abundances were extracted

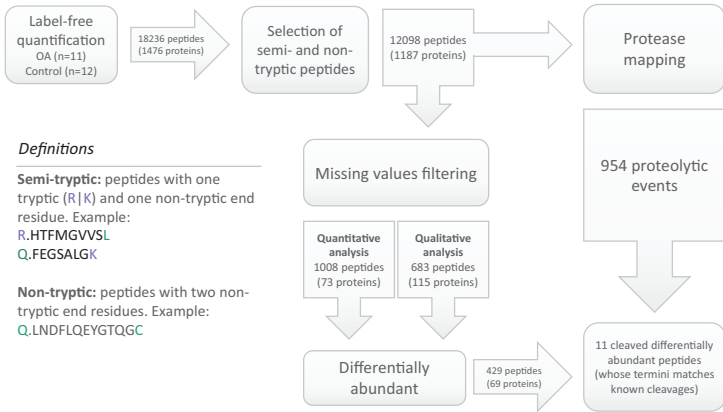

FIGURE 1 Schematic overview of the study.

using FDR settings peptide −10lgP > = 15, protein −10lgP > = 20, PTM Ascore > = 0. Only peptide sequences with one or more non-tryptic terminus were used (Figure 1). The data were normalized to reduce technical biases using quantile normalization in the NormalyzerDE R-package [21].

2.3 | Statistical (quantitative) analysis

The inclusion criterion for the quantitative analysis was for a peptide to be quantified in at least seven samples in each group. After filtering for missing values, 1008 peptide sequences remained and were included in the quantitative analysis (Figure 1). We conducted the statistical analysis using mixed linear regression models in R using the lme4 package on base-2 log-transformed intensity as the response. A separate model was fitted to each protein, including all peptides from this protein. Age, sex, disease status, and peptide were used as fixed effects terms with interactions between disease status and peptide. The subject was included as random effects term. Contrasts between groups (OA vs. controls) were specified using the emmeans package and are reported with 95% confidence intervals based on restricted maximum likelihood estimates using the Kenward-Rogers method for estimation of degrees of freedom. Peptides that had a 95% confidence interval of the base 2 log fold-change not spanning zero were considered differentially abundant. Although the comparison was made on the peptide level, much of the biological meaning is found on the protein level. Therefore, we mainly describe and emphasize the "host protein," that is, the annotated protein representing a peptide used in the analysis. Given the exploratory nature of the study and the use of mixed models that minimize the multiplicity problem [22], we did not apply any further corrections for multiplicity, but rather we report all derived estimates to inform future studies and meta-analyses.

2.4 | Qualitative analysis

In the set of peptide sequences with fewer than seven quantified samples in either group, we identified 683 peptides (115 proteins) of interest that were used in a qualitative analysis (Figure 1). The peptides were ranked based on the difference in the number of samples they had been quantified in, that is, having a non-missing value. For example, a sequence quantified in eleven OA samples and only in one control sample, or vice versa, would have a high rank. Host proteins of peptides with an absolute difference in the number of quantified samples greater than or equal to seven were defined as qualitatively differentially abundant.

2.5 | Protease mapping

We used the protease database MEROPS [23] to identify proteolytic cleavage sites in the dataset. The entire set of SQL statements used to construct the MEROPS database was downloaded from https://ebi.ac.uk/pub/databases/merops/current_release/meropsweb121.tar.gz on April 10, 2022. We exported the SQL table Substrate search in text format. Only human entries were kept, totaling 5188 rows of proteolytic events, 1678 unique proteases, and 15168 unique substrates. We mapped MEROPS IDs to Uniprot accessions and Gene names using the human reference proteome downloaded from the Uniprot website (https://www.uniprot.org/proteomes/UP000005640) on April 13, 2022. In the SF dataset, for each peptide sequence, the residue numbers (start and end position) of the protein sequence where cleaving occurred was annotated. 12,098 peptide sequences were used as input. The start and end cleavage positions were further mapped to the residue number of known cleavage sites in the MEROPS database. That resulted in a dataset containing information on the substrates, proteases, cleaved residue and their position, and the

peptide sequences. The MEROPS substrate names were converted to UniProt IDs and annotated as reviewed (UniProtKB/Swiss-Prot) or unreviewed (UniProtKB/TrEMBL), referring to their level of curation and annotation in the UniProt knowledgebase. The proteases that were extracted from the MEROPS database were further searched against the current dataset to be able to identify which proteases that actually could be detected in this cohort. These proteases will be from here on referred to as "found in SF."

2.6 | R package "proteasy"

To facilitate the workflow and reproducibility of this and future analyses, the first author (MR) developed an R-package, which is available on GitHub (https://github.com/martinry/proteasy) and was accepted to Bioconductor on June 28th, 2022. The main function, findProteases, takes a set of UniProt accessions, peptide sequences, and name of studied organism as input, and outputs the equivalent of Tables S1–S3 of this manuscript, detailing the proteases, substrates, and cleavages of potential relevance for the input. Full documentation of the package is available in the package manual, with examples provided in the accompanied vignette.

2.7 | Pathway and network analysis

Functional analysis of pathways was conducted using a local database containing pathway data from REACTOME [24] and using a custom script for pathway analysis from a previously published study [25]. Inclusion criteria for pathway analysis was host proteins whose peptides were differentially expressed, as previously defined. The "background proteome" was set to host proteins of all identified peptides in the dataset. Only sets of differentially abundant proteins larger than three were included.

Proteases and substrates were also visualized as an interaction network using the igraph [26] library in R. Here, we included all proteases found in SF in the current dataset, the substrates they cleave, and annotated (where applicable) whether the protein's abundance increased or decreased, according to the definitions stated above.

3 | RESULTS

3.1 | Quantitative analysis

Using our definitions for determining differential abundance, where we aggregated the results of quantitative and qualitative analysis, we identified 429 peptides (from 69 host proteins) as differentially abundant (Figure 2, Table S4 and S5). In the quantitative analysis 271 peptides (59 proteins) were differentially abundant between OA and normal group (Table S4). In this set, 162 peptides (37 proteins) were increased, and 109 peptides (30 proteins) were decreased (Figure 2). Host proteins of the most increased peptides (log 2 FC > = 2.5) were histidine-rich glycoprotein (HRG), alpha-2-HS-glycoprotein (AHSG), and apolipoprotein A1 (APOA1). Another set of host proteins of highly increased peptides ($2 < \log 2\ FC < 2.5$) were fibronectin (FN1), kininogen-1 (KNG1), Inter-alpha-trypsin inhibitor (ITIH1), Apolipoprotein B (APOB), and complement C3 (C3).

3.2 | Qualitative analysis

In the qualitative data analysis, the largest difference was observed for HRG, plasminogen (PLG), FN1, AHSG and inter-alpha-trypsin inhibitor (ITIH2) that were found in OA but not control samples, and decorin (DCN) and extracellular superoxide dismutase (SOD3) found in control samples but not in OA samples (Table S5).

3.3 | Protease mapping

We identified entries corresponding to 954 proteolytic events (Table S3) based on our mapping of peptide ends to the MEROPS database. A majority, 736 cleaved at the N-terminal position. One hundred ninety-two proteases (57 reviewed) were identified as potentially cleaving actors (Table S2). Out of those, 11 proteases were also found in SF (Figure 2), one of which complement factor B (CFB) contains peptides increased in both the quantitative and qualitative analysis. Fifty-seven proteases (corresponding to 243 proteolytic events) had status "reviewed." The cleaving activity of these proteases were annotated for 61 substrates (Table S1). The protease families that accounted for most of the 243 proteolytic events were metalloproteinases (56 events—about 23%), cathepsins (15 events—about 6%), and caspases (10 events—about 4%).

We found 33 reviewed proteases potentially acting on substrates of differentially abundant peptides (Table S7). Among these were metalloproteinases—MMP1, MMP2, MMP3, MMP7, MMP9, MMP8, MMP11, MMP12, MMP14, MMP26, and ADAMTS5 (ADAMTS5); caspases—CASP3, CASP4, and CASP6; cathepsins—CTSB, CTSD, CTSE, and CTSG; calpains—CAPN1 and CAPN2; carboxypeptidases—CPN1, CPB2, and CPM. Additional proteases were CELA1, CFB, CMA1, ELANE, GZMB, KLK1, MEP1A, MEP1B, PLAU, and PREP. No cleaving proteases were found acting on the most highly increased host proteins HRG or AHSG. Cleaving proteases acting on APOA1 were chymase (CMA1), macrophage metalloelastase (MMP12) and matrilysin (MMP7), but neither of these proteases were found in SF. C3 was cleaved by nine proteases, four of which were detected in SF. FN1 was cleaved by 20 proteases, one of which, MMP2 was detected in SF. KNG1 was cleaved by 35 proteases, three of which; cathepsin D (CTSD), carboxypeptidase N catalytic chain (CPN1), and carboxypeptidase B2 (CPB2) were detected in SF (Figure 3).

The qualitatively increased proteins with annotated cleaving events were FN1 (75 events mainly involving ELANE, MMP2, MMP8, MMP9, MMP12, A disintegrin and metalloproteinase with thrombospondin motifs 5 but also multiple instances of unreviewed proteases highly

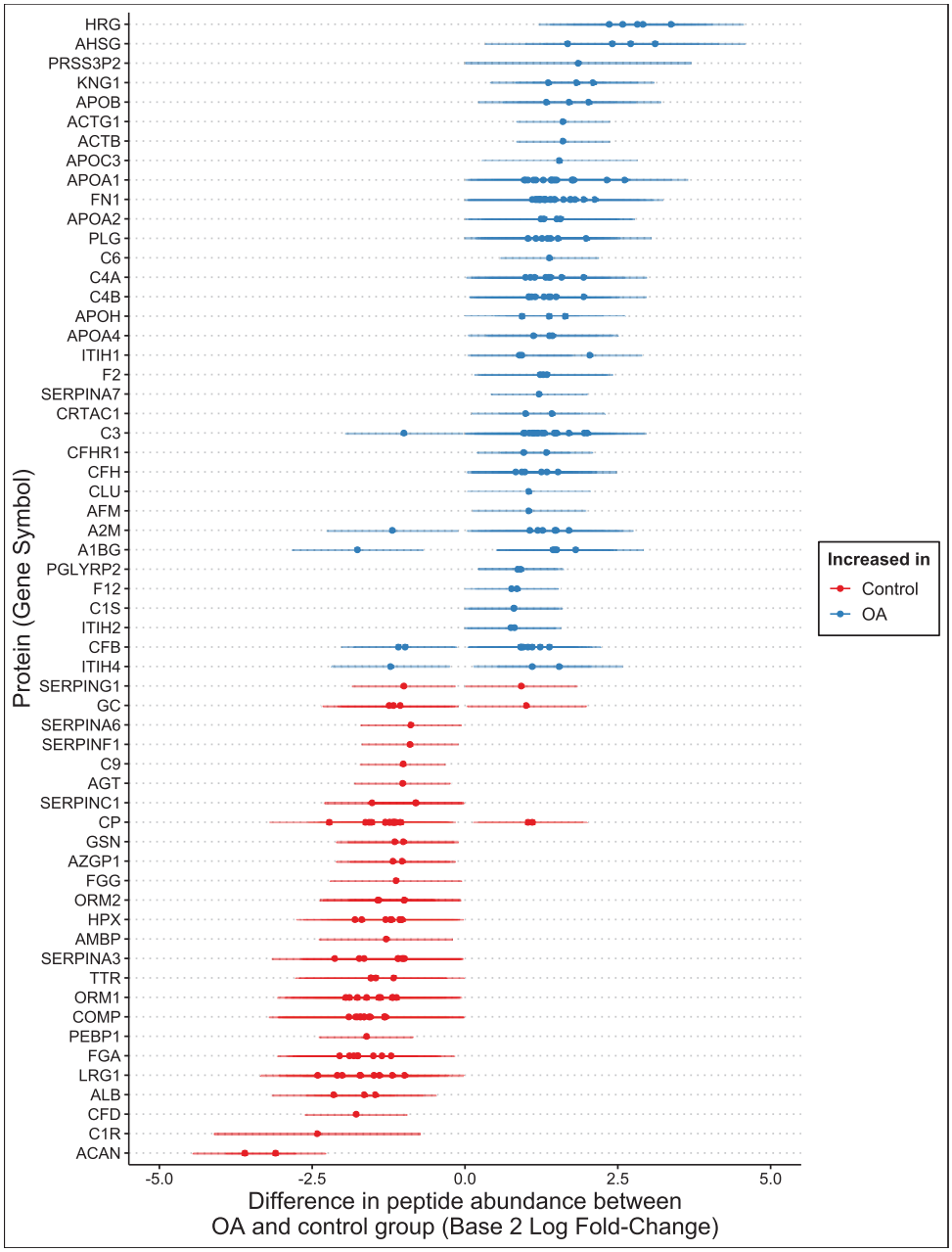

FIGURE 2 Differentially abundant peptides (based solely on the quantitative analysis), with 95% confidence intervals, aggregated on protein level (y-axis), and sorted by mean estimate of peptides within host protein. Only peptides with confidence interval not spanning zero included. Blue: differentially abundant peptides more abundant in OA group, Red: differentially abundant peptides more abundant in Control group.

FIGURE 3 Interaction network of proteases found in SF, and cleaved substrates, annotated by increased or decreased differential abundance in OA versus controls.

similar to neutrophil collagenase), CFB (a protease cleaving C3, see Figure 3), C3, KNG1, FN1, alpha-2-macroglobulin (A2M). The qualitatively decreased proteins with annotated cleaving events were aggrecan core protein (ACAN), alpha-1-antichymotrypsin (SERPINA3), fibrinogen alpha chain (FGA) and phosphatidylethanolamine-binding protein 1 (PEBP1) (Table S7).

3.4 | Pathway analysis

Pathway analysis of host proteins of peptides with increased levels in OA group resulted in 65 pathways with three or more differentially abundant proteins in each set (Table 1). These results suggested OA protein activity predominantly in pathways related to immune system (particularly complement activation), transport of small molecules, and hemostasis. The pathways with complete overlap between differentially abundant proteins and background sets were "Terminal pathway of complement" and "Alternative complement activation."

3.5 | Comparison with tryptic data

We assessed whether increased levels of a peptide was due to elevated enzymatic degradation of its host-protein in OA or greater abundance of the intact host-protein in OA, by contrasting our quantitative results with our previous study that compared the same late-stage OA and healthy controls [14], but in which protein abundances were calculated from intensities of tryptic peptides (Figures S1 and S2). We found that 49 host-proteins of differentially abundant semi- and non-tryptic peptides in the current study were not differentially abundant in the previous study (Tables S4 and S5). For example, proteins APOA1, ITIH1 and C3 were not differentially abundant in the previous study but are host-proteins of highly increased peptides in the current analysis.

4 | DISCUSSION

In this study of endogenously cleaved peptides in human end-stage knee OA versus knee-healthy controls, we identified 69 host

TABLE 1 Pathway results for host proteins of differentially abundant peptides.

Pathway	Category	Ratio
Terminal pathway of complement	Immune system	8/8 (100%)
Alternative complement activation	Immune system	4/4 (100%)
Activation of C3 and C5	Immune system	6/7 (86%)
Chylomicron remodeling	Transport of small molecules	5/7 (71%)
Dissolution of Fibrin Clot	Hemostasis	4/6 (67%)
Integrin signaling	Hemostasis	4/6 (67%)
GRB2:SOS provides linkage to MAPK signaling for Integrins	Hemostasis	4/6 (67%)
p130Cas linkage to MAPK signaling for integrins	Hemostasis	4/6 (67%)
MyD88 deficiency (TLR2/4)	Disease	4/6 (67%)
IRAK4 deficiency (TLR2/4)	Disease	4/6 (67%)
MyD88:MAL(TIRAP) cascade initiated on plasma membrane	Immune system	4/6 (67%)
MAP2K and MAPK activation	Signal transduction	7/11 (64%)
Signaling by high-kinase activity BRAF mutants	Disease	7/11 (64%)
Regulation of TLR by endogenous ligand	Immune system	5/8 (63%)
Chylomicron assembly	Transport of small molecules	5/8 (63%)
Amyloid fiber formation	Metabolism of proteins	13/22 (59%)
Signaling by moderate kinase activity BRAF mutants	Disease	7/12 (58%)
Paradoxical activation of RAF signaling by kinase inactive BRAF	Disease	7/12 (58%)
Signaling downstream of RAS mutants	Disease	7/12 (58%)
Retinoid metabolism and transport	Metabolism	10/18 (56%)
Peptide ligand-binding receptors	Signal transduction	5/9 (56%)
Signaling by RAF1 mutants	Disease	6/11 (55%)
Regulation of complement cascade	Immune system	22/43 (51%)
Clathrin-mediated endocytosis	Vesicle-mediated transport	8/16 (50%)
Signaling by BRAF and RAF1 fusions	Disease	7/14 (50%)
Antimicrobial peptides	Immune system	6/12 (50%)
Cargo recognition for clathrin-mediated endocytosis	Vesicle-mediated transport	5/10 (50%)
Platelet degranulation	Hemostasis	26/58 (45%)
Intrinsic pathway of fibrin clot formation	Hemostasis	8/18 (44%)
Common pathway of fibrin clot formation	Hemostasis	7/16 (44%)
ER-phagosome pathway	Immune system	5/12 (42%)
Post-translational protein phosphorylation	Metabolism of proteins	19/54 (35%)
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	Metabolism of proteins	21/63 (33%)
G alpha (i) signaling events	Signal transduction	4/10 (40%)
Iron uptake and transport	Transport of small molecules	4/11 (36%)
Scavenging of heme from plasma	Vesicle-mediated transport	6/20 (30%)
Initial triggering of complement	Immune system	5/20 (25%)
Degradation of the extracellular matrix	Extracellular matrix organization	5/23 (22%)
Integrin cell surface interactions	Extracellular matrix organization	8/38 (21%)
ECM proteoglycans	Extracellular matrix organization	5/29 (17%)
Neutrophil degranulation	Immune system	22/129 (17%)

Note: Ratio is the ratio of proteins in the set of host proteins of differentially abundant peptides overlapping with the background proteome set. All host proteins of peptides (non-tryptic, semi-tryptic, and tryptic) identified in SF were used as background proteome, and only pathways with more than three proteins in a set were included.

proteins of 429 differentially abundant semi- or non-tryptic peptides, a majority of which were increased in OA, suggesting an increased proteolytic activity of the proteases that potentially cleave these peptides. Strongest support for such cleaving event to have happened exists for those proteases that we also could identify in SF (Figure 3). The majority of these proteases were metalloproteinases, cathepsins, and carboxypeptidases.

To facilitate retrieval of cleavage sites, we developed an R-package named proteasy. Similar to existing tools such as TopFind and Proteasix, proteasy exists for the purpose of retrieving data about proteases by mapping peptide termini positions to known sites where a protease cleaves. The main function of proteasy, *findProteasy*, works similarly to the "observed" finding mode of Proteasix, but with some important differences. Proteasix makes use of additional knowledgebases to MEROPS, but a large overlap between these is to be expected. Proteasy only returns proteases from the same organism as input whereas Proteasix also returns proteases for multiple species. Proteasy searches only the exact endopeptidase whereas Proteasix searches a window of the endopeptidase + 1, 2, and 3 AA. Differences between the methods may also be attributed to use of different versions of the MEROPS database. The proteasy R-package utilizes data derived from the MEROPS database, and is limited to the entries therein. The MEROPS database is a manually curated knowledgebase with information about proteolytic enzymes, their inhibitors and substrates. The main function in the package allows for batch identification of possible proteases for a set of substrates (protein IDs and peptide sequences), and may serve as a useful tool in peptide-centric analyses.

In both the qualitative and quantitative comparisons, we found peptide levels of HRG to differ the most between the OA and the control group (Table S5 and S6). HRG has been reported to be increased in OA in previous proteomics studies [27, 28]. HRG is an abundant protein in plasma and has been referred to as "the Swiss army knife" of plasma due to its involvement in many biological processes and ability to interact with multiple ligands simultaneously [29]. We also observed increased protease activity of AHSG, another glycoprotein, known to influence the mineral phase of bone [30]. The extensive proteolysis these proteins undergo may negatively impact the regulatory functions they are involved in. These two proteins, together with KNG1, form the type 3 subgroup within the human cystatin superfamily of cysteine protease inhibitors [31]. This subgroup has been implicated in angiogenesis [32], the formation of new blood vessels and may be useful in treatment of diseases with extensive angiogenesis. In OA progression, vascular growth is increased in the synovium, osteophytes and menisci which contributes to the development of synovitis, osteochondral damage, osteophyte formation and meniscal pathology [33].

We examined whether increased levels of a peptide was due to elevated enzymatic degradation of its host-protein in OA, or if the increase was a result of greater abundance of the intact protein in OA, by contrasting our results with a previous study conducted on the same samples, but where quantification was done solely on tryptic peptides. We hypothesized that a protein which is increased/upregulated in OA versus controls in both studies is likely due to greater quantities of the intact protein, while an increase of semi- and non-tryptic peptides alone is likely a consequence of degradation. We found that one such protein, APOA1, was not upregulated in the study based on tryptic peptides, but the semi-tryptic peptides for which it was annotated in the current study were highly increased. Further, we found CMA1 as a possible cleaving protease matching APOA1 peptides (Table S7).

The two largest subnetworks from the 32 differentially cleaved host proteins belongs to lipoprotein particles complexes, such as VLDL and LDL and the complement factor cascade. We found the proteases and substrates acting on the lipoprotein particle complexes and complement cascade to be important factors in OA development (Table 1). Two pathways with involvement in transport of small molecules; "Chylomicron remodeling," "Chylomicron assembly" ranked highly due to the high representation of apolipoproteins. The proteases that potentially cleaved these apolipoproteins were MMP7 and MMP12 (Table S6) that are commonly known to cleave ECM proteins. MMPs have previously been shown to be able to inhibit apolipoproteins functionality by cleaving them [34]. This suggests that the MMPs do not only act as degrading enzymes of the surrounding tissues but could also inhibit the lipid metabolism in that environment [34, 35]. Chymase has also been identified to be an important protease cleaving apolipoproteins [36]. The main producer of chymase is mast cells [37]. Mast cells have previously been detected to be in a higher range in synovium of OA in comparison to RA [38] and to be associated with radiographic damage in OA [38]. These findings suggest an increased proteolytic activity not just against the surrounding tissues but also plasma proteins that are essential to maintain the homeostasis of the joints.

Most of the differentially cleaved proteins (Table S6) were proteins involved in the extracellular cellular matrix assembly; fibronectin, cartilage acidic protein 1 (CRTAC1), clusterin (CLUAP1), aggrecan and actin (ACTB). Also plasma proteins; C3, albumin (ALB) were differentially cleaved. Proteases that were found to actively cleave some of these proteins were MMPs, chymase, caspase, elastase, granzyme, but no known cleaving site was detected for CRTAC1 [39]. CRTAC1 is known to be abundant in cartilage and CRTAC1 as a protein of an interest to be detected both in early and late-stage OA that have showed a trend of increase in late-stage OA [39]. The semi-tryptic peptides that were found to be differentially expressed in this study are fragments from the same part of CRTAC1, which means a potential to multiple cleaving activities in this specific region of the protein.

A study by Abji et al. examined the proteases present in SF using flow cytometry [40]. They identified 42 proteases identified in psoriatic arthritis, rheumatoid arthritis, and OA patients. Comparing these results to our findings, 11 of these proteases were also identified in the current study, a majority of which were MMPs.

Several semi-tryptic peptides that are derived from fibronectin could be found in the MEROPS database to be cleaved by several proteases, such neutrophile elastase. Neutrophile elastase induces chondrocytes apoptosis and facilitates the OA occurrence via caspase signaling pathway. Fibronectin, a glycoprotein abundant in normal cartilage [41], is one of many ECM proteins that may originate from cartilage degradation.

Peptides of fibrinogen proteins, fibronectin, actin, and phosphatidylethanolamine-binding protein contributed to the high ranking of multiple signaling pathways related to the RAF kinase family. RAF kinases have been of interest in the study of multiple diseases, particularly in cancer research [42], and in the case of OA the Ras/Raf/MEK/ERK pathway has been reported to induce expression of MMP13 and association with OA in mice [43].

This study encompasses a broad profiling of endogenous peptides in synovial fluid. The strengths this study offers are the use of human SF samples from both OA patients and knee-healthy controls, robust statistical methodology and advanced bioinformatics operations to highlight important insights about the actors involved in the endogenous cleaving of peptides, and how they may relate to OA disease development.

The field of peptidomics focuses on peptides which often display biological activity, such as hormones, cytokines, toxins, neuropeptides and alike, which are generated from larger precursors, as well as biomarker-type peptides that may not have any bioactivity but are indicative of a particular pathology. Commonly employed enrichment protocols for such studies include ultracentrifugation, filtration, precipitation, or other type of small peptide fragment enrichment [44]. A single universal system suitable for extraction and separation of all classes of peptidomes is yet to be reported. Enrichment techniques are typically designed to target peptides with specific physiochemical properties, and are not equally suitable for all peptide classes [45]. These protocols enhance the detection of smaller peptides, generally enabling detection of a few hundred up to a thousand peptides. To cover a higher cleaving action usually multiple enrichment protocols need to be addressed. Therefore, our strategy in this study was to use a protocol that eliminates enrichment selection of a specific subgroup of peptides. By studying the fragments/peptides of the of the proteins that remain in the pellet after ethanol precipitation, we were able to identify a larger number of peptides that was not cleaved by trypsin. In our analysis, we identified 12,098 peptides with at least one non-tryptic end site. Studying a larger set of peptides could then better reflect the proteolytic activity. This strategy comes at a cost in that small protein fragments are eliminated during the precipitation stage. Moreover, endogenous proteases that target tryptic cleavage sites cannot be distinguished from trypsin.

Missing data on peptide abundance are a known issue in MS analyses. Not to discard these data, we performed a quantitative analysis based on the patterns of missingness. We acknowledge that our definition of qualitatively differentially abundant peptides is arbitrary, but we believe it provides more information than discarding the missing data. For transparency, we include the information on missing data pattern for all identified peptides, including those not suitable for statistical analysis (Table S8). But even with such a drawback still the amount of quantified semi-tryptic peptides that were detected is to our knowledge the highest number of identified semi-tryptic peptides identified differing between OA and controls [20].

Some limitations for the pathway analysis are that ranking was done by presence of differential peptides in relation to annotated peptides in the background set, a metric which does not necessarily take into account the complex relationships present in biological pathways. Another limitation is the somewhat abstract concept of a pathway, which is not always representative of the underlying biology, and is subject to interpretation [46].

Finally, many cleaved peptides are not identifiable from existing databases. This may have several explanations. The quality of peptide identification relies on estimating a false discovery rate using a decoy-target approach, but may still include false positive identifications. Moreover, incomplete annotation is still a limitation for protease databases. For each proteolytic event, MEROPS provides a list of selected references, and currently contains over 10,000 such references [23]. Inevitably, the cost of avoiding annotation error by availability of peer-reviewed references limits the coverage of the knowledgebase.

In summary, we have performed a discovery-based profiling of the SF peptidome in OA and healthy subjects, and studied the role of proteases and inflammatory cytokines which are the potential cleaving actors. We developed an R-package to facilitate analyses of proteases and substrates. Our findings suggest the increased proteolytic activity associated with OA catabolism is not restricted to the ECM of the surrounding tissues, but may also implicate homeostasis by external factors through mechanisms such as chylomicron remodeling and assembly.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement #771121), the Swedish Research Council, the Foundation for Research in Rheumatology (FOREUM), the IAB Lundberg Foundation, the Greta and Johan Kock Foundation, the Swedish Rheumatism Association, the Österlund Foundation, the Governmental Funding of Clinical Research program within the National Health Service (ALF), and the Faculty of Medicine, Lund University, Sweden. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

ASSOCIATED DATA

The raw MS files has been deposited to ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023708.

DATA AVAILABILITY STATEMENT

The raw MS files has been deposited to ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD023708.

ORCID

Martin Rydén https://orcid.org/0000-0001-6968-4314

REFERENCES

1. Chen, D., Shen, J., Zhao, W., Wang, T., Han, L., Hamilton, J. L., & Im, H-J. (2017). Osteoarthritis: Toward a comprehensive understanding of pathological mechanism. *Bone Res.*, *5*, 16044.
2. Loeser, R. F., Goldring, S. R., Scanzello, C. R., & Goldring, M. B. (2012). Osteoarthritis: A disease of the joint as an organ. *Arthritis and Rheumatism*, *64*, 1697–1707.
3. Maldonado, M., & Nam, J. (2013). The role of changes in extracellular matrix of cartilage in the presence of inflammation on the pathology of osteoarthritis. *BioMed Research International*, *2013*, 1.
4. Struglics, A., Larsson, S., Pratta, M. A., Kumar, S., Lark, M. W., & Lohmander, L. S. (2006). Human osteoarthritis synovial fluid and joint cartilage contain both aggrecanase- and matrix metalloproteinase-generated aggrecan fragments. *Osteoarthritis and Cartilage*, *14*, 101–113.
5. El-Sayed, M. E., Asmaa, K. F., & Samar, E.-B. S. (2019). The role of matrix metalloproteinases in osteoarthritis pathogenesis: An updated review. *Life Sciences*, *234*, 116786.
6. Kapoor, M., Martel-Pelletier, J., Lajeunesse, D., Pelletier, J.-P., & Fahmi, H. (2011). Role of proinflammatory cytokines in the pathophysiology of osteoarthritis. *Nature Reviews Rheumatology*, *7*, 33–42.
7. Tío, L., Martel-Pelletier, J., Pelletier, J.-P., Bishop, P. N., Roughley, P., Farran, A., Benito, P., & Monfort, J. (2014). Characterization of opticin digestion by proteases involved in osteoarthritis development. *Joint, Bone, Spine*, *81*, 137–141.
8. Hui, A. Y., Mccarty, W. J., Masuda, K., Firestein, G. S., & Sah, R. L. (2012). A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *4*, 15–37.
9. Vincent, T. L. (2022). OA synovial fluid: Biological insights into a whole-joint disease. *Osteoarthr Cartilage*, *30*, 765–766.
10. Liao, W., Li, Z., Li, T., Zhang, Q., Zhang, H., & & Wang, X. (2018). Proteomic analysis of synovial fluid in osteoarthritis using SWATH-mass spectrometry. *Molecular Medicine Reports*, *17*, 2827–2836.
11. Gobezie, R., Kho, A., Krastins, B., Sarracino, D. A., Thornhill, T. S., Chase, M., Millett, P. J., & Lee, D. M. (2007). High abundance synovial fluid proteome: Distinct profiles in health and osteoarthritis. *Arthritis Research & Therapy*, *9*, R36–R36.
12. Barreto, G., Soliymani, R., Baumann, M., Waris, E., Eklund, K. K., Zenobi-Wong, M., & Lalowski, M. (2018). Functional analysis of synovial fluid from osteoarthritic knee and carpometacarpal joints unravels different molecular profiles. *Rheumatology*, *58*, 897–907.
13. Bennike, T., Ayturk, U., Haslauer, C. M., Froehlich, J. W., Proffen, B. L., Barnaby, O., Birkelund, S., Murray, M. M., Warman, M. L., Stensballe, A., & Steen, H. (2014). A normative study of the synovial fluid proteome from healthy porcine knee joints. *Journal of Proteome Research*, *13*, 4377–4387.
14. Ali, N., Turkiewicz, A., Hughes, V., Folkesson, E., Tjörnstand, J., Neuman, P., Önnerfjord, P., & Englund, M. (2022). Proteomics profiling of human synovial fluid suggests increased protein interplay in early-osteoarthritis (OA) that is lost in late-stage OA. *Molecular & Cellular Proteomics*, *21*, 100200.
15. Balakrishnan, L., Nirujogi, R. S., Ahmad, S., Bhattacharjee, M., Manda, S. S., Renuse, S., Kelkar, D. S., Subbannayya, Y., Raju, R., Goel, R., Thomas, J. K., Kaur, N., Dhillon, M., Tankala, S. G., Jois, R., Vasdev, V., Ramachandra, Y., Sahasrabuddhe, N. A., Prasad, T. K., … Pandey, A. (2014). Proteomic analysis of human osteoarthritis synovial fluid. *Clinical Proteomics*, *11*, 6–6.
16. Kamphorst, J. J., Van Der Heijden, R., Degroot, J., Lafeber, F. P. J. G., Reijmers, T. H., Van El, B., Tjaden, U. R., Van Der Greef, J., & Hankemeier, T. (2007). Profiling of endogenous peptides in human synovial fluid by NanoLC—MS: Method validation and peptide identification. *Journal of Proteome Research*, *6*, 4388–4396.
17. Ritter, S. Y., Subbaiah, R., Bebek, G., Crish, J., Scanzello, C. R., Krastins, B., Sarracino, D., Lopez, M. F., Crow, M. K., Aigner, T., Goldring, M. B., Goldring, S. R., Lee, D. M., Gobezie, R., & Aliprantis, A. O. (2013). Proteomic analysis of synovial fluid from the osteoarthritic knee: Comparison with transcriptome analyses of joint tissues. *Seminars in Arthritis and Rheumatism*, *65*, 981–992.
18. Foreman, R. E., George, A. L., Reimann, F., Gribble, F. M., & Kay, R. G. (2021). Peptidomics: A review of clinical applications and methodologies. *Journal of Proteome Research*, *20*, 3782–3797.
19. Mahendran, S. M., Keystone, E. C., Krawetz, R. J., Liang, K., Diamandis, E. P., & Chandran, V. (2019). Elucidating the endogenous synovial fluid proteome and peptidome of inflammatory arthritis using label-free mass spectrometry. *Clinical Proteomics*, *16*, 23.
20. Fernández-Puente, P., González-Rodríguez, L., Calamia, V., Picchi, F., Lourido, L., Camacho-Encina, M., Oreiro, N., Rocha, B., Paz-González, R., Marina, A., García, C., Blanco, F. J., & Ruiz-Romero, C. (2019). Analysis of endogenous peptides released from osteoarthritic cartilage unravels novel pathogenic markers*. *Molecular & Cellular Proteomics*, *18*, 2018–2028.
21. Willforss, J., Chawade, A., & Levander, F. (2019). NormalyzerDE: Online tool for improved normalization of omics expression data and high-sensitivity differential expression analysis. *Journal of Proteome Research*, *18*, 732–740.
22. Ji, H., & Liu, X. S. (2010). Analyzing 'omics data using hierarchical models. *Nature Biotechnology*, *28*, 337–340.
23. Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2014). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, *42*, D503–D509.
24. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., … D'eustachio, P. (2017). The reactome pathway knowledgebase. *Nucleic Acids Research*, *46*, D649–D655.
25. Rydén, M., Englund, M., & Ali, N. (2021). ProteoMill: Efficient network-based functional analysis portal for proteomics data. *Bioinformatics*, *37*, 3491–3493.
26. Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Interjournal*, Complex Systems, 1695.
27. Fernández-Puente, P., Mateos, J., Fernández-Costa, C., Oreiro, N., Fernández-López, C., Ruiz-Romero, C., & Blanco, F. J. (2011). Identification of a panel of novel serum osteoarthritis biomarkers. *Journal of Proteome Research*, *10*, 5095–5101.
28. Chae, D.-S., Kim, M. E., Kang, K.-Y., Lee, N. Y., Lee, W. S., & Lee, J. S. (2021). Quantitative proteomic analysis comparing grades ICRS1 and ICRS3 in patients with osteoarthritis. *Experimental and Therapeutic Medicine*, *22*, 1470.
29. Poon, I. K. H., Patel, K. K., Davis, D. S., Parish, C. R., & Hulett, M. D. (2011). Histidine-rich glycoprotein: The Swiss Army knife of mammalian plasma. *Blood*, *117*, 2093–2101.
30. Brylka, L., & Jahnen-Dechent, W. (2013). The role of fetuin-A in physiological and pathological mineralization the role of fetuin-A in physiological and pathological mineralization. *Calcified Tissue International*, *93*, 355–364.
31. Rizzu, P., & Baldini, A. (1995). Three members of the human cystatin gene superfamily, AHSG, HRG, and KNG, map within one megabase of genomic DNA at 3q27. *Cytogenetic and Genome Research*, *70*, 26–28.
32. Lee, C., Bongcam-Rudloff, E., Sollner, C., Jahnen-Dechent, W., & Claesson-Welsh, L. (2009). Type 3 cystatins; fetuins, kininogen and histidine-rich glycoprotein. *Frontiers in Bioscience*, *14*, 2911–2922.
33. Mapp, P. I., & Walsh, D. A. (2012). Mechanisms and targets of angiogenesis and nerve growth in osteoarthritis. *Nature Reviews Rheumatology*, *8*, 390–398.
34. Park, J. H., Park, S-M., Park, S-H., Cho, K-H., & Lee, S-T. (2008). Cleavage and functional loss of human apolipoprotein E by digestion of matrix metalloproteinase-14. *Proteomics*, *8*, 2926–2935.

35. Sanchez, C., Bay-Jensen, A.-C., Pap, T., Dvir-Ginzberg, M., Quasnichka, H., Barrett-Jolley, R., Mobasheri, A., & Henrotin, Y. (2017). Chondrocyte secretome: A source of novel insights and exploratory biomarkers of osteoarthritis. *Osteoarthr Cartilage*, *25*, 1199–1209.
36. He, A., & Shi, G-P. (2012). Mast cell chymase and tryptase as targets for cardiovascular and metabolic diseases. *Current Pharmaceutical Design*, *19*, 1114–1125.
37. Caughey, G. H. (2007). Mast cell tryptases and chymases in inflammation and host defense. *Immunological Reviews*, *217*, 141–154.
38. Lange-Brokaar, B. J. E., Kloppenburg, M., Andersen, S. N., Dorjée, A. L., Yusuf, E., Herb-van Toorn, L., Kroon, H. M., Zuurmond, A-M., Stojanovic-Susulic, V., Bloem, J. L., Nelissen, R. G. H. H., Toes, R. E. M., & Ioan-Facsinay, A. (2016). Characterization of synovial mast cells in knee osteoarthritis: Association with clinical parameters. *Osteoarthr Cartilage*, *24*, 664–671.
39. Styrkarsdottir, U., Lund, S. H., Saevarsdottir, S., Magnusson, M. I., Gunnarsdottir, K., Norddahl, G. L., Frigge, M. L., Ivarsdottir, E. V., Bjornsdottir, G., Holm, H., Thorgeirsson, G., Rafnar, T., Jonsdottir, I., Ingvarsson, T., Jonsson, H., Sulem, P., Thorsteinsdottir, U., Gudbjartsson, D., & Stefansson, K. (2021). The CRTAC1 protein in plasma is associated with osteoarthritis and predicts progression to joint replacement: A large-scale proteomics scan in Iceland. *Arthritis & Rheumatology (Hoboken, NJ)*, *73*, 2025–2034.
40. Abji, F., Rasti, M., Gómez-Aristizábal, A., Muytjens, C., Saifeddine, M., Mihara, K., Motahhari, M., Gandhi, R., Viswanathan, S., Hollenberg, M. D., Oikonomopoulou, K., & Chandran, V. (2021). Proteinase-mediated macrophage signaling in psoriatic arthritis. *Frontiers in Immunology*, *11*, 629726.
41. Chevalier, X. (1993). Fibronectin, cartilage, and osteoarthritis. *Seminars in Arthritis and Rheumatism*, *22*, 307–318.
42. Matallanas, D., Birtwistle, M., Romano, D., Zebisch, A., Rauch, J., Von Kriegsheim, A., & Kolch, W. (2011). Raf family kinases: Old dogs have learned new tricks. *Genes & Cancer*, *2*, 232–260.
43. Xu, L., Peng, H., Wu, D., Hu, K., Goldring, M. B., Olsen, B. R., & Li, Y. (2005). Activation of the discoidin domain receptor 2 induces expression of matrix metalloproteinase 13 associated with osteoarthritis in mice*◆. *Journal of Biological Chemistry*, *280*, 548–555.
44. Olsen, C. A. (2010). Methods in molecular biology, volume 615: Peptidomics, methods and protocols. Edited by Mikhail Soloviev. *ChemBioChem*, *11*, 1782–1782.
45. Bhutada, S., Li, L., Willard, B., Muschler, G., Piuzzi, N., & Apte, S. S. (2022). Forward and reverse degradomics defines the proteolytic landscape of human knee osteoarthritic cartilage and the role of the serine protease HtrA1. *Osteoarthr Cartilage*, *30*, 1091–1102.
46. Kirik, U., Cifani, P., Albrekt, A.-S., Lindstedt, M., Heyden, A., & Levander, F. (2012). Multimodel pathway enrichment methods for functional evaluation of expression regulation. *Journal of Proteome Research*, *11*, 2955–2967.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202300040 in the Supporting Information section at the end of the article.

How to cite this article: Rydén, M., Turkiewicz, A., Önnerfjord, P., Tjörnstrand, J., Englund, M., & Ali, N. (2023). Identification and quantification of degradome components in human synovial fluid reveals an increased proteolytic activity in knee osteoarthritis patients vs controls. *Proteomics*, *23*, e2300040. https://doi.org/10.1002/pmic.202300040

Paper III

Received: 24 January 2023 | Revised: 5 May 2023 | Accepted: 18 May 2023

DOI: 10.1002/jor.25633

RESEARCH ARTICLE

Journal of Orthopaedic Research®

A human meniscus explant model for studying early events in osteoarthritis development by proteomics

Martin Rydén[1] | Karin Lindblom[2] | Aida Yifter-Lindgren[2] | Aleksandra Turkiewicz[1] | Anders Aspberg[2] | Viveka Tillgren[2] | Martin Englund[1] | Patrik Önnerfjord[2]

[1]Department of Clinical Sciences Lund, Orthopaedics, Clinical Epidemiology Unit, Faculty of Medicine, Lund University, Lund, Sweden

[2]Department of Clinical Sciences Lund, Section for Rheumatology and Molecular Skeletal Biology, Faculty of Medicine, Lund University, Lund, Sweden

Correspondence
Martin Rydén, Department of Clinical Sciences Lund, Clinical Epidemiology Unit, Lund University, Remissgatan 4, 222 42 Lund, Sweden.
Email: martin.ryden@med.lu.se

Funding information
Foundation for Movement Disabilities in Skåne; Alfred Österlund Foundation; Stiftelsen Olle Engkvist Byggmästare; Reumatikerförbundet; IngaBritt och Arne Lundbergs Forskningsstiftelse; H2020 European Research Council, Grant/Award Number: 771121; Vetenskapsrådet; ALF agreement between the Swedish government and the county; Anna-Greta Crafoords Stiftelse för Reumatologisk Forskning; Greta och Johan Kocks stiftelser

Abstract

Degenerative meniscus lesions have been associated with both osteoarthritis etiology and its progression. We, therefore, sought to establish a human meniscus ex vivo model to study the meniscal response to cytokine treatment using a proteomics approach. Lateral menisci were obtained from five knee-healthy donors. The meniscal body was cut into vertical slices and further divided into an inner (avascular) and outer region. Explants were either left untreated (controls) or stimulated with cytokines. Medium changes were conducted every 3 days up to Day 21 and liquid chromatography–mass spectrometry was performed at all the time points for the identification and quantification of proteins. Mixed-effect linear regression models were used for statistical analysis to estimate the effect of treatments versus control on protein abundance. Treatment by IL1ß increased release of cytokines such as interleukins, chemokines, and matrix metalloproteinases but a limited catabolic effect in healthy human menisci explants. Further, we observed an increased release of matrix proteins (collagens, integrins, prolargin, tenascin) in response to oncostatin M (OSM) + tumor necrosis factor (TNF) and TNF+interleukin-6 (IL6) + sIL6R treatments, and analysis of semitryptic peptides provided additional evidence of increased catabolic effects in response to these treatments. The induced activation of catabolic processes may play a role in osteoarthritis development.

KEYWORDS
cytokines, explants, meniscus, osteoarthritis, proteomics

1 | INTRODUCTION

Osteoarthritis (OA) is the most common form of arthritis and a major source of pain, disability, and socioeconomic cost worldwide.[1] However, the detailed molecular mechanisms involved in OA initiation and progression are still poorly understood, there is no available cure and no biomarkers in use within clinical practice. Knee OA is often the pathological response of joint tissues to increased biomechanical stress resulting in thickening of the subchondral bone, osteophyte formation, inflammation in the synovium, hypertrophy of the joint capsule, and degenerative changes of articular cartilage, ligaments, and menisci.[2] The degradation of articular cartilage has

been attributed to increased proteolytic activity of matrix-degrading enzymes, including matrix metalloproteinases (MMPs), collagenases, aggrecanases, and proteinases belonging to serine and cysteine families.[2] While previous ex vivo models of OA using cartilage[3–7] have provided insights about the disease, studying the degenerative changes in the extracellular matrix of human menisci may contribute with valuable new knowledge about the degenerative role of cytokines in a complex disease affecting the whole joint. The network of cytokines as key mediators of inflammation and their involvement in catabolic processes receive increasing attention. It has been reported that synthesis and mechanisms of cytokines may vary during the disease process of OA.[8] The disruption of homeostasis that cytokines cause is, particularly in tissues often subjected to high mechanical load, of key interest when studying OA disease progression. The progressive degeneration involves processes of inflammation, degradation, and synthesis, which, together often result in the loss of joint function and pain.[9,10]

In vitro and ex vivo models have been extensively used to study pathological changes, molecular pathways, and the effect and role of cytokines in certain conditions, and proinflammatory stimulation of human meniscus has previously been studied using cell cultures.[11] However, no inflammatory disease model has been established for human meniscus using an explant tissue model.[12] Human OA tissue samples are often collected at the end-stages of disease, making it difficult to study early changes and factors that are involved in the disease progression.[13] In vitro and ex vivo models overcome these limitations by offering a controlled environment in which initial disease mechanisms may be simulated and studied.

In the current proof-of-concept study, we establish a human meniscus ex vivo model to explore the meniscal response of healthy human meniscus to cytokine treatment using a proteomics approach to study and follow the release of matrix proteins.

2 | MATERIALS AND METHODS

2.1 | Explant harvest and treatment

Lateral donor menisci ($N = 5$, non-OA, four male and one female, age range 69–80) were harvested within 40 h postmortem without known chronic joint disease. The procedure was approved by the Lund University ethics committee. The meniscus was visually inspected to be macroscopically intact.

Vertical slices (1 mm wide) were cut radially from the meniscal body and were divided into an inner (inner 1/3rd) and outer region (outer 2/3rd). The inner region represents the white/white avascular region while the outer region is vascularized. Each slice is weighed ("wet weight," later used for weight correction in the quantitative analysis) and placed in medium in a 24-well plate. Incubation at 37°C, 5% CO_2 for 24 h to let the explant slices equilibrate. Medium for culture with treatment: 500 mL Dulbecco's Modified Eagle Medium (DMEM)/F12 HEPES (SH30023.01), PEST 5 mL (1:100), L-Proline 0.4 mM, insulin, transferrin supplemented media (ITS) 5 mL, vitamin C 50 µg/mL + Fungizone (first two time points).

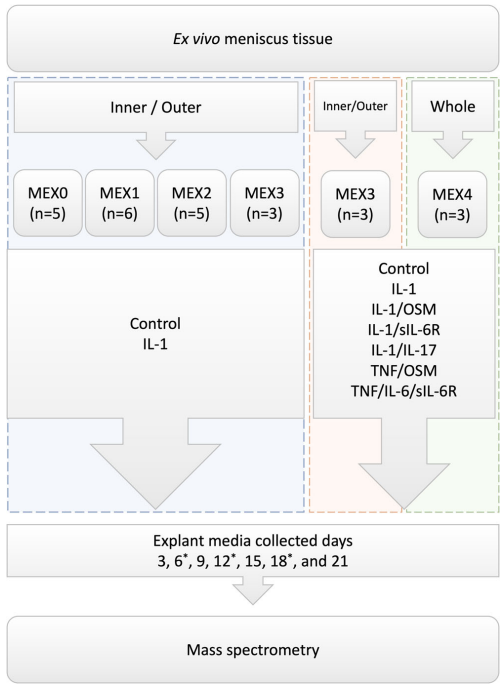

FIGURE 1 Schematic overview of the study. *Mass spectrometry analysis was not conducted for time points 6, 12, or 18 for MEX3 and MEX4.

Explants were untreated (controls) or treated with cytokines (Figure 1, Supporting Information: Table S7) and cultured in serum-free DMEM media (supplemented with ITS) for 21 days. Medium changes were carried out every 3 days and used medium was collected (96-deepwell plate) and stored at −20°C until analysis. Media was then added to the explants (1 mL for outer, 0.5 mL for inner explant wells). We first evaluated treatment with IL1 in four samples in total (MEX0-3). Then, we evaluated seven treatments in sample MEX3. Sample MEX4 was included as a validation of results in MEX3, without separating inner/outer zones.

2.2 | Metabolic activity and glycosaminoglycan (GAG) release

GAG-release was evaluated in all samples using a 1,9-dimethylmethylene blue (DMMB) assay. DMMB solution was prepared according to Farndale et al.[14] A 20 µL of sample was mixed with 200 µL DMMB-solution and absorbance was measured at

520 nm. Shark chondroitin-sulfate (Sigma-Aldrich) was used as standard. Wells were washed three times.

Fold-changes (treatment vs. control) were estimated using mixed linear regression models in R. Base-2 log-transformed GAG-release was used as response. Treatment, time point and zone with interactions between all terms were used as fixed effects, and subject and zone were used as nested random effects terms for MEX0-3, which had multiple biological replicates with repeated measurements. In MEX3, only zone was used as random effect.

The metabolic activity was evaluated at for time points d0, d6, d12, and d21, where 400 µL Alamar mix (10% AlamarBlue [BioRad] in medium without Fungizone and vitamin C) was added to explants after harvesting the media and incubated for 3 h at 37°C. Samples were transferred to a 96-well plate and the absorbance (570/600 nm) was measured to determine the metabolic activity of the explants.

2.3 | Mass spectrometry (MS) preparation and identification

Culture medium (50 µL) was prepared for MS analysis as previously described[15]: explant culture media (50 µL) was reduced by 4 mM dithiothreitol for 30 min at 56°C, alkylated by 16 mM iodoacetamide for 60 min in the dark at room temperature, ethanol precipitated (9:1) and then digested by 0.25 µg trypsin gold (Promega) in 0.1 M ammonium bicarbonate (AMBIC) pH 7.8 for 16 h on a shaker at 37°C. After drying, samples were resuspended in 100 µL AMBIC with 0.5 M NaCl, run through 30 kDa filter (PALL Life Sciences) and desalted with reversed-phase C18 cartridges (AssayMAP, Agilent Technologies) using a Bravo robot. Discovery MS was performed using a quadrupole Orbitrap benchtop mass spectrometer (Q-Exactive HFX, Thermo Scientific) with prior separation of peptides using a liquid chromatography system (EASY-nLC 1000, Thermo Scientific) on an analytical column (PepMap RSLC C18, 75 µm × 25 cm, Thermo Scientific) coupled on-line using a nano-electrospray ion source with a column temperature at +45°C (EASY-Spray, Thermo Scientific) using a flow rate of 300 nL/min and a 1 h binary gradient. Protein identification was performed in Proteome Discoverer 2.5 (Thermo Scientific) using two search engines in parallel: a tryptic search against the UniProt human (UP000005640 from January 2021) sequence database combined with an MSPep spectral search against the NIST_human_Orbitrap_HCD_20160923 library (mass tolerance: 10 and 20 ppm in MS1, MS2 respectively. Other Sequest search settings were modifications: carbamidomethylation (fixed: C), oxidation (variable: M, P) missed cleavages (max 2), mass tolerance (MS1-10ppm, MS2-0.02 Da). Label-free protein abundance quantification was obtained by averaging peak area intensities from the top three unique peptides for each protein. To determine individual peptide abundances, we performed a semitryptic database search to enable identification of nontryptic cleavages within the data set. This was performed using the same combined searches as above but in series. The protein false discovery rate (FDR) was 0.01 for both searches.

2.4 | Statistical analysis

Proteins with maximum five missing values per treatment group in MEX0-3, and maximum three missing values per treatment group in MEX3-4, were considered to have sufficient data points for statistical analysis. Based on this criterium, MEX0-3 included 804 proteins, MEX3 included 822 proteins, and MEX4 included 508 proteins for statistical analysis.

Statistical analysis was performed using mixed linear regression models in R, using the lme4 package.[16] Three linear regression models were used for the different data sets. In each model, base-2 log-transformed intensity was used as response. Treatment, time point, and zone with interactions between all terms were used as fixed effects, and subject and zone were used as nested random effects terms for MEX0-3, which had multiple biological replicates with repeated measurements. Treatment, time point, and zone (inner/outer) with interactions between all terms were used as fixed effects, and zone as random effect term were used for MEX3. Treatment and time point with interactions were used as fixed effect terms for the MEX4 data set. Contrasts between treatment and control were specified using the emmeans package[17] and are reported with 95% confidence intervals based on restricted maximum likelihood estimates using Kenward–Rogers method for estimation of degrees of freedom. Estimates (base 2 log fold changes) were extracted as means across time points. Model diagnostics was conducted to validate model fit. Proteins that had a 95% confidence interval not spanning zero were considered differentially abundant. Given the exploratory nature of the study and use of mixed models that minimize the multiplicity problem,[18] we did not apply any further corrections for multiplicity, but rather we report all derived estimates to inform future studies and meta-analyses.

2.5 | Principal component analysis (PCA)

PCA was conducted to examine whether treatment effect and time effect contributed to clustering of samples, and whether similarity in release profiles between treatments could be observed. We used the PCA function in the mixOmics R package,[19] which uses multilevel decomposition for repeated measurements on data which excluded proteins with any missing values, the outcome was log transformed (base 2), then scaled and centered using the IMIFA R package.[20] PCA was visualized using the plotly R package.[21]

2.6 | Analysis of semitryptic peptides

As a possible indicator of degradation and catabolic effect, we studied the number of semitryptic peptides identified by MS/MS in the different treatment groups. Tryptic peptides were filtered out and only peptides identified at FDR < 0.01 in at least two files using Sequest HT search engine were counted.

2.7 | Time series clustering

Unsupervised clustering was performed to identify groups of proteins with similar treatment response profiles over time. We excluded proteins with any missing values. Protein abundances were standardized to have mean = 0 and SD = 1. Then, for each treatment and time point, we took the mean value of each protein's replicates. We calculated pairwise slope distances, and clusters were obtained using the getClusters function in the R package tscR.[22] The number of clusters for each treatment was decided by visually interpreting cluster dendrogram of the slope distance. The trajectories were plotted using the R package ggplot2.[23]

2.8 | Protein classification and interaction analysis

Differentially upregulated proteins with a log2 fold change >1.5 was illustrated as interaction networks using the R-package igraph.[24] Interaction data was collected from STRINGdb.[25] Community detection was conducted using the cluster_louvain algorithm to emphasize dense subgraphs. Nodes were labeled in accordance with the protein's classification in pantherdb.[26] Communities were colored according to the most common protein class in each community.

2.9 | Complementary analysis (Western blot, proximity extension assays [PEAs], and quantitative polymerase chain reaction [qPCR])

Explant media from outer menisci Days 12 and 21 (MEX3) were thawed and triplicates from the different days and treatments were pooled (15 µL each). Samples were run nonreduced on NuPage Bis/tris gels 4%–12% in MOPS buffer. Western blots (against a COMP neoepitope fragment[27]) were run in Tris-Glycine buffer +10% methanol and proteins transferred to polyvinylidene difluoride membranes. Filters were blocked in 3% bovine serum albumin (BSA) T-TBS and all antibody incubations were done in 3% BSA T-TBS. As substrate Super signal West Dura from Pierce were used.

We used the Olink® Explore 384 inflammation panel (Olink proteomics AB) to obtain complementary data. Explant culture media from all time points were pooled for each replicate ($n = 3$) and region in the MEX3 experiment using treatments: control, IL1, (oncostatin M [OSM] + tumor necrosis factor [TNF]) and (TNF + IL6 + sIL6R). Samples were randomized in order and 40 µL was used for the analysis and relative quantification performed at Olink. Proteins with linear normalized protein expression (NPX) greater than or equal to two times limit of detection was selected for differential abundance analysis, including 102 proteins. Fold-changes (treatment vs. control) were estimated using mixed linear regression models in R. Base-2 log-transformed NPX was used as response. Treatment and zone with interactions were used as fixed effects, and zone were used as random effect.

As an additional complementary analysis, we used qPCR to study expression changes between control and IL1, (OSM + TNF), or (TNF + IL6 + sIL6R) treatment groups in explant tissue from Day 21 of MEX4. Briefly, the explant tissue samples were pulverized in liquid N_2 and total RNA extracted and purified using the RNAqueous kit (Invitrogen, #AM1912). RNA concentration and purity were determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific) and cDNA was synthesized from 60 ng of total RNA using the Maxima First Strand cDNA Synthesis kit (Thermo Fisher Scientific, #K1641). qPCR was performed with 3 ng cDNA per reaction, and TaqMan Fast Advanced Master Mix (Applied Biosystems, #4444556) on an Applied Biosystems StepOnePlus Real-Time PCR System. The following TaqMan Gene Expression Assays from Applied Biosystems were used; ACAN (Hs00153936_m1), CHI3L1 (Hs01072228_m1), COL1A1 (Hs00164004_m1), COL3A1 (Hs00943809_m1), MMP1 (Hs00899658_m1), MMP13 (Hs00942584_m1), TIMP2 (Hs00234278_m1), and IEF3I (Hs01116184_m1). Each tissue explant sample cDNA was analyzed in duplicate qPCR reactions and the relative expression levels of each gene were calculated using the ΔΔCt method[28] with IEF3I as endogenous reference,[29] and data presented as base 2 log Fold Change (i.e., −ΔΔCt).

3 | RESULTS

3.1 | Metabolic activity and GAG release

The reduction of AlamarBlue was consistently higher in explant plugs compared to negative control (Figure 2), suggesting explants were healthy and viable throughout the experiment. GAG-analysis by DMMB assay revealed IL1 had a slightly increased effect on the release of GAGs than control. Log2 fold-changes was greater for IL1 Days 6–12 in the inner zone (Supporting Information: Figure S1). The GAG release for the multigroup comparison is shown in Supporting Information: Figure S2 showing small differences with various treatments.

3.2 | Identification and quantification

Across all data sets, a total of 2248 proteins were identified and quantified in the explant media (1528 in MEX0-3 inner zone, 1090 in MEX0-3 outer zone, 1303 in MEX3 inner zone, 1434 in MEX3 outer zone, and 1307 in MEX4).

3.3 | PCA

PCA revealed cytokine treatment, and to a lesser extent, release over time, as the main contributing factors to the observed clustering effect exhibited in Figure 3.

FIGURE 2 Mean reduction of AlamarBlue (metabolic activity) of explant culture in MEX0-3; treatments interleukin-1 (IL1) and control. Blue line: metabolic activity for meniscus plugs. Red line: negative control.

3.4 | Differentially abundant proteins

Overall, a majority of proteins were upregulated in treatment versus control. While treatment with IL1 led to upregulation of many relevant proteins, ex growth-regulated alpha protein (gene symbol: CXCL1), interleukin-8 (CXCL8), interleukin-6 (IL6), neutrophil gelatinase-associated lipocalin (LCN2), C-X-C motif chemokine 6 (CXCL6) (Supporting Information: Table ST1), the effect was not as strong as expected. Thus, in the follow-up experiment, we expanded the number of treatments. In this multigroup treatment comparison, the most notable were comparisons of OSM + TNF versus control (MEX3), where 215 proteins were upregulated in the inner zone and 261 proteins were upregulated in the outer zone (Supporting Information: Table ST5). Corresponding comparison of TNF + IL6s + ILR versus control (MEX3) had 191 upregulated proteins in the inner zone and 95 upregulated proteins in the outer zone (Supporting

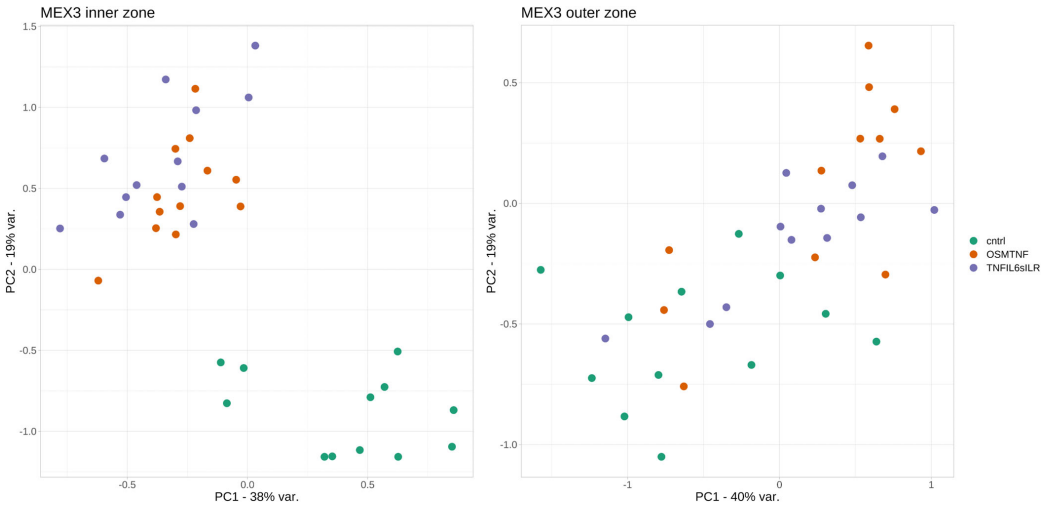

FIGURE 3 Principal component analysis of media proteome (MEX3) in treatments (oncostatin M [OSM] + tumor necrosis factor [TNF] and TNF + interleukin-6 [IL6] + sIL6R) for inner and outer zones. There is a clear treatment effect for cytokines versus control particularly in the inner zone.

Information: Table ST5). A majority of highly upregulated (log2 fold change >3, Figure 4) proteins in OSM + TNF versus control and TNF + IL6s + ILR (MEX3 inner and outer zones) had the pantherdb classification "extracellular matrix protein." Contrarily, none of the highly upregulated proteins in IL1 versus control (MEX0-3 inner and outer zones) were annotated as ECM proteins according to pantherdb classification.

4 | ZONAL DIFFERENCES

We selected a set of 33 extensively researched proteins based on their functional classifications as ECM proteins, MMPs, proteases, and protease inhibitors. We performed hierarchical clustering on the mean values of standardized abundances (Z-score) of these proteins, which revealed a similar clustering pattern between the inner and outer zones (Figure 5). In both zones, we found the clusters consisting of proteins with the highest Z-score contained the proteins COL3A1, ACAN, MMP1, and MMP3.

In Supporting Information: Figure S6, a Venn diagram of the inner and outer meniscal zones illustrates the number of unique and shared differentially abundant proteins that were identified for OSM + TNF and TNF + IL6 + sIL6R. The greatest number of unique differentially abundant proteins were identified in the OSM + TNF treatment in the outer zone, with 66 unique proteins. We identified 63 proteins which were differentially abundant in both zones and treatments.

Comparing protein–protein interactions for IL1 versus Control of the inner and outer zones (Figure 6), the main similarities observed between the zones were that IL1 induces the release of inflammatory mediators such as IL6, CXCL1, CXCL6, and CXCL8. Specifically for the outer zone, our analysis revealed the interaction between IL6 and superoxide dismutase (SOD2) as well as C-C motif chemokine 2 (CCL2), along with the interactions between cystatin B (CSTB) and tumor necrosis factor-inducible gene 6 protein (TNFAIP6) and CXCL1.

5 | MENISCAL RESPONSE TO PROINFLAMMATORY STIMULI

5.1 | Immune system response

The cytokine treatment increased the release of proteins involved in inflammation, such as interleukins, chemokines, and MMPs. For example, some of the most upregulated proteins in both inner and outer zones of the MEX0-3 replicates were growth-regulated alpha protein (CXCL1), interleukin-8 (CXCL8), IL6, neutrophil gelatinase-associated lipocalin (LCN2) and CXCL6 (Supporting Information: Table ST1).

5.2 | Catabolic effect

While the immune system response was most prominent in the contrast IL1 versus control (Figure 4), we observed upregulation of multiple ECM proteins in the contrast OSM + TNF (MEX3 inner zone) (Figure 6). In both the contrasts OSM + TNF versus control and TNF + IL6 + sIL6R versus control we found upregulated ECM proteins

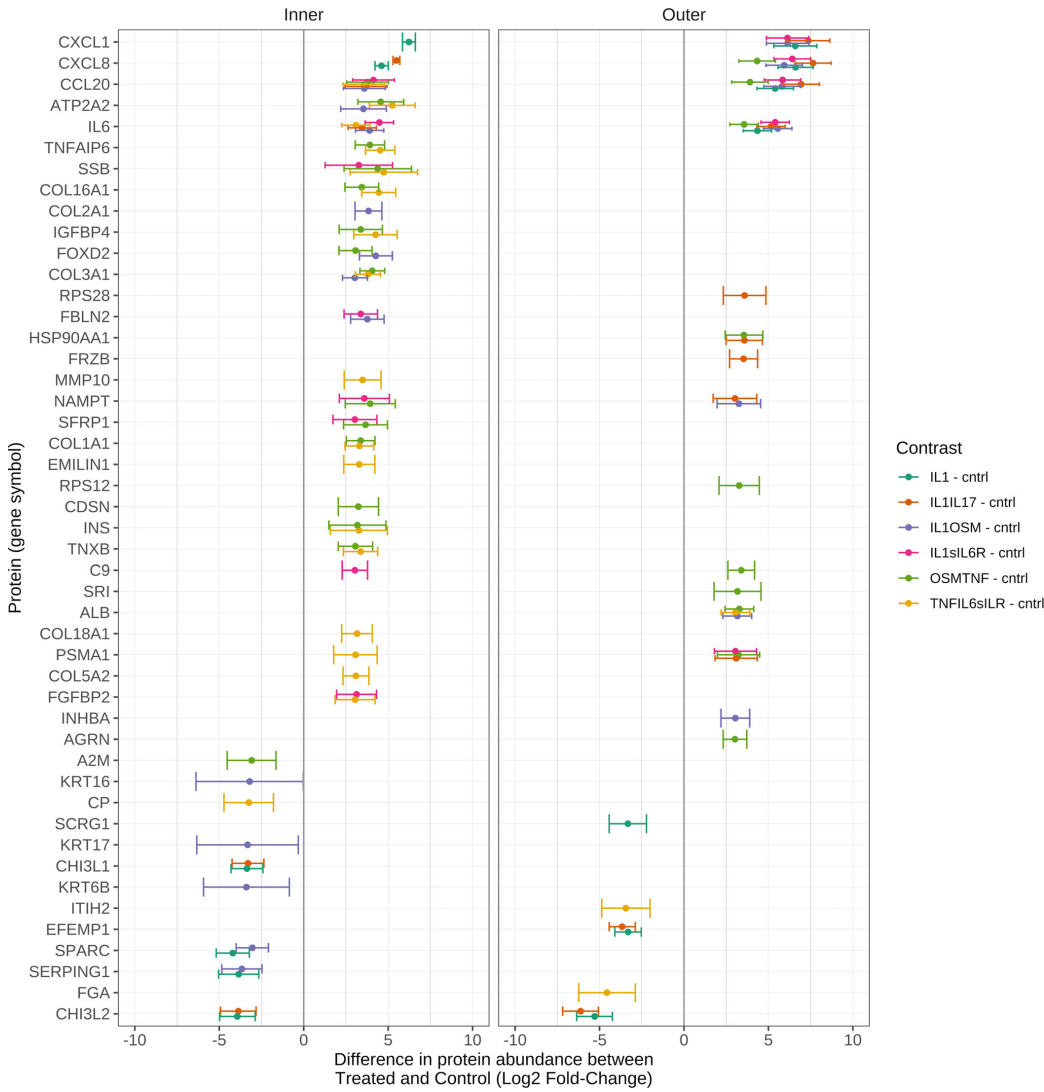

FIGURE 4 Differentially abundant proteins in MEX3 inner and outer zones. Proteins with a $\log_2$ fold change >3 were included.

von Willebrand factor and versican interacting with upregulated cytokines and protease inhibitors. In both the inner and outer zone, we identified ECM "communities" of primarily collagens with interactions to proteases such as MMPs and cathepsins. Upregulation of ECM proteins were also prominent in IL1 + OSM versus control (MEX3 inner zone). However, besides interleukins, few cytokines were estimated to be upregulated in this treatment group (Supporting Information: Table ST2, Supporting Information: Figure S3).

5.3 | Time-dependent clustering

The time-dependent effect of treatments was studied by clustering analysis, performed on standardized release at each time point (Figure 7). Each set of proteins annotated to a cluster for a specific treatment was also plotted for control group as reference. For most treatments, we observed three clusters: one with initial increase from Day 3, one which displayed peak in release on Day 9, and one where

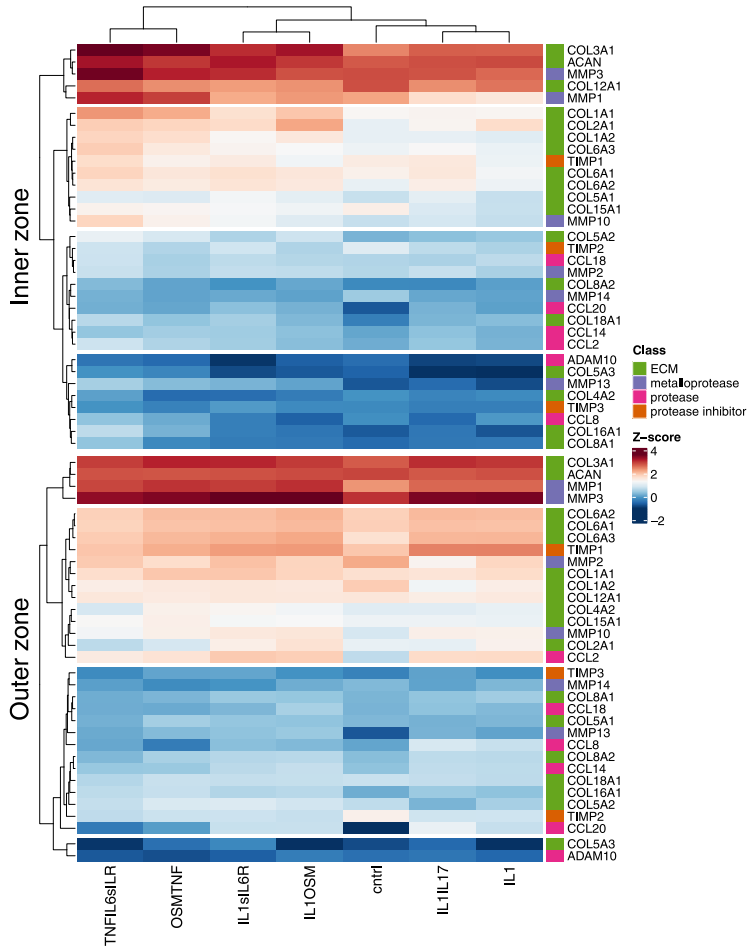

FIGURE 5 Standardized (Z-score) proteomic abundances of selected proteins in inner and outer meniscal zones (MEX3) across treatments, color annotated by functional classification. Heatmap shows the standardized proteomic abundances of selected proteins across treatments and the two meniscal zones (inner and outer). The color scale represents the relative Z-score (red: high, blue: low). Proteins are annotated by functional classification.

no clear increase occurred. The patterns were not seen to the same extent in the control clusters.

5.4 | Complementary analysis of catabolic activity

5.4.1 | Release of COMP fragment

Western blot analysis of the COMP neoepitope (QQS^{77})[27] (Supporting Information: Figure S5) supported the findings of increased catabolic effect, compared to the control that showed no fragment release, particularly by treatments OSM + TNF and TNF + IL6 + sIL6R at Day 12 while the release was almost absent at Day 21.

5.4.2 | Analysis of semitryptic peptides

Semitryptic were studied as a possible indicator of induced catabolism. In Supporting Information: Table ST6, the number of semitryptic peptides identified in each experiment and treatment group, as well as the total number of identified peptides in each experiment. In MEX3 (inner and outer zones), the highest number of

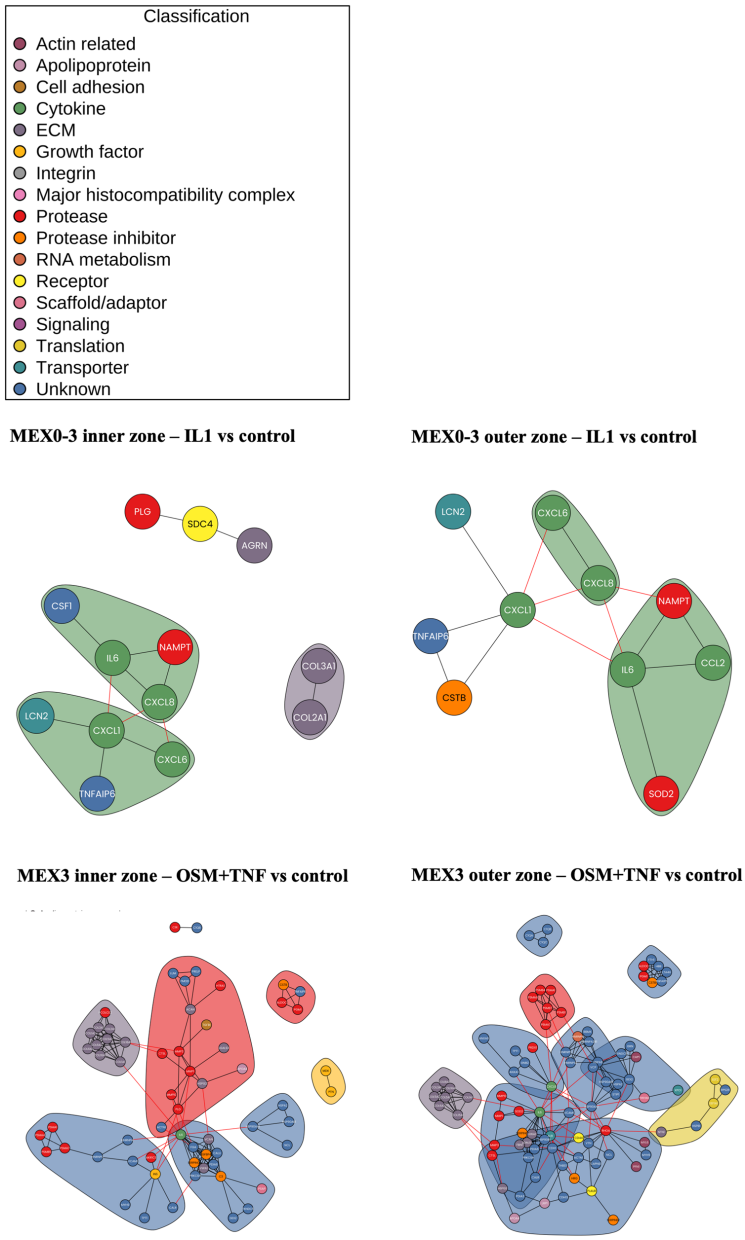

FIGURE 6 Protein–protein interaction network for treatment versus control in inner (left) and outer (right) zones. Proteins which had a log2 fold-change >1.5, annotated by classification. Red lines denote between-community interactions and black lines denote within-community interactions. See additional network figures in Supporting Information: Figure S4.

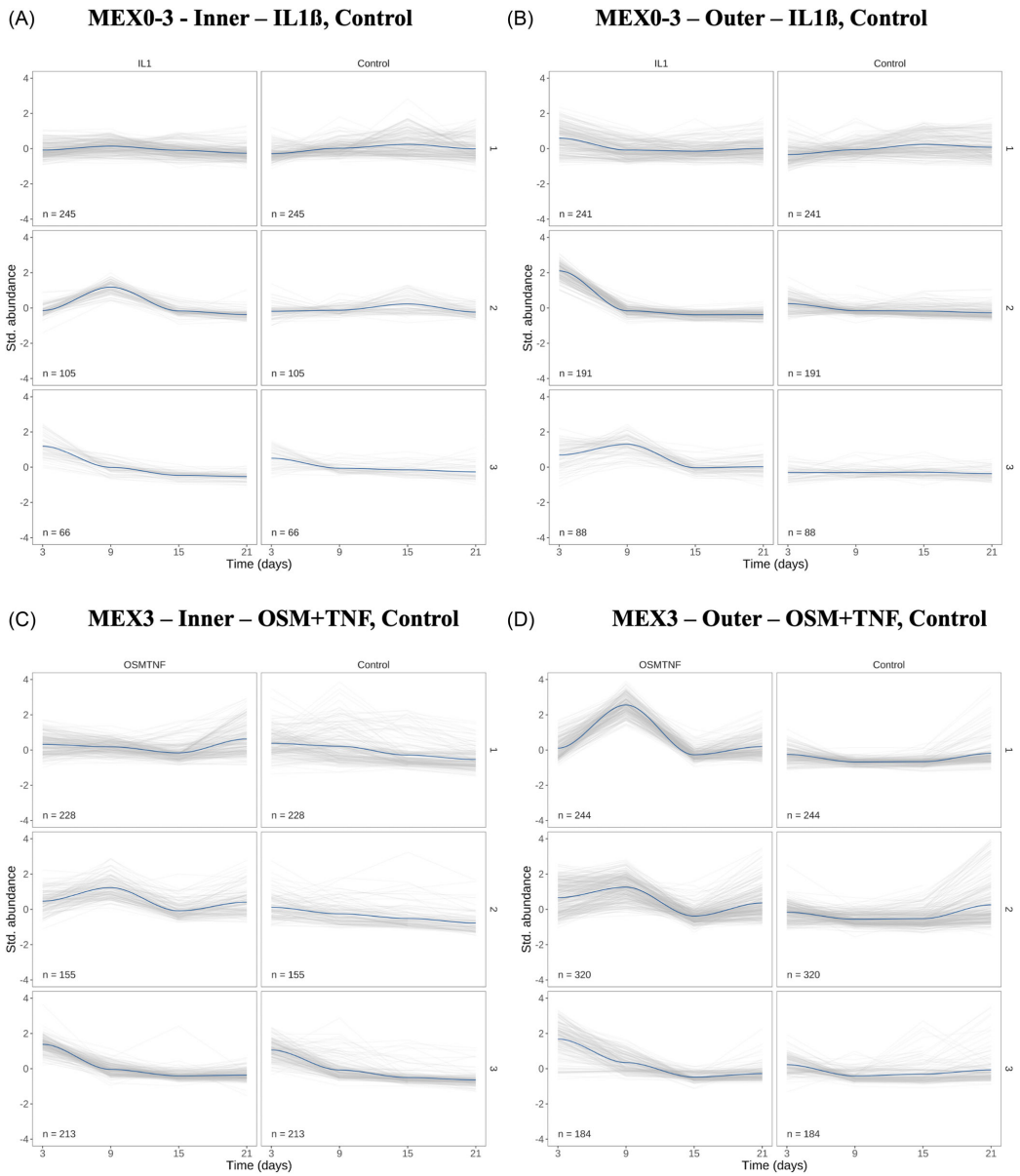

FIGURE 7 Release profiles for upregulated proteins by cluster for treatments interleukin-1 (IL1) and oncostatin M (OSM) + tumor necrosis factor (TNF) and the respective release patterns for control group in MEX3. The blue line indicates the mean standardized abundance for proteins within a cluster. Additional cluster figures are found in Supporting Information: Figure S4.

identified semitryptic peptides was for OSM + TNF. The highest number of identified semitryptic peptides in MEX3 inner zone was identified in the group treated with TNF + IL6 + sIL6R, while in both MEX3 outer zone and MEX4 the largest fractions were identified in the OSM + TNF treatment.

5.4.3 | Olink PEA

Differential abundance analysis of Olink was conducted as a complement to and comparison with the MS results. Differential abundance of Olink protein data revealed that most upregulated proteins were in the outer zone. In both the inner and outer zone, OSM + TNF versus control and TNF + IL6 + sILR versus control, we estimated more upregulated proteins than in IL1 versus control. In the inner zone, the same nine proteins were upregulated in OSM + TNF and TNF + IL6 + sILR. Twenty of the proteins in the outer zone overlapped between OSM + TNF and TNF + IL6 + sILR. Eight proteins upregulated in OSM + TNF were not upregulated in TNF + IL6 + sILR. These were interleukin-1 beta, placenta growth factor, matrilin-2, growth-regulated alpha protein (CXCL1), SPARC-related modular calcium-binding protein, endothelial cell-specific molecule, serpin B8, and toll-like receptor 3.

Comparing differentially abundant proteins in Olink (Supporting Information: Table ST4) and MS data (MEX3, Supporting Information: Table ST2), we found that in the inner zone, the number of upregulated proteins overlapping between Olink and MS methods were less than or equal to seven. Eight proteins in the contrast OSM + TNF versus control (outer zone) were not identified in any MS experiment, and included growth factors, cytokines, and signal receptors (Supporting Information: Table ST1–ST4).

5.4.4 | qPCR

The gene expression results revealed upregulation for MMP1 and MMP13 in meniscus explant tissue in all three treatments, while the ACAN, COL1A1, COL3A1, CHI3L1, and TIMP2 genes were down-regulated in all three treatments (Supporting Information: Figure S7). The results found by qPCR were consistent with the expression patterns and differential abundance results of the MS-proteomics data for MMPs in the culture medium, while the reverse was found for most of the ECM proteins. These findings support that the cytokine treatment have a catabolic effect on the meniscus, resulting in decreased expression of ECM components and increased expression of proteolytic enzymes.

6 | DISCUSSION

We have developed a human meniscus degeneration ex vivo model and studied protein release over time. While treatment with IL1 leads to increased release of GAG and upregulation of several ECM proteins, treatment with OSM + TNF or TNF + IL6 + sIL6R appear to induce the strongest catabolic effect.

Dysregulation of protein homeostasis mediated by mechanical injury, oxidative stress, and inflammation contribute to ECM catabolism in OA.[30] Damage to the ECM and its association with joint destruction has been well studied in cartilage,[31–33] and while multiple studies have confirmed the clinical importance of the meniscus in OA disease,[34–37] few have investigated the meniscal response to cytokine treatment in human tissue. In this study we used a proteomics approach to study early events of OA in an ex vivo model, and provide evidence that a catabolic response can be induced in such explant model.

Type I collagen is the most abundant in meniscus and constitutes 98% of the total collagen content.[38] The protein is susceptible to degradation by multiple collagenases, such as metalloproteinases. Our data showed upregulation of both MMP1 and MMP2 and several ECM proteins, particularly in OSM + TNF and TNF + IL6 + sIL6R treatments (Supporting Information: Table ST2). In MEX4, we found strong upregulation (log2 fold-change >3) of MMP1, MMP3, MMP10, and MMP14. We found the treatments OSM + TNF and TNF + IL6 + sIL6R to release matrix proteins and proteases involved in ECM breakdown (Supporting Information: Table ST2–ST3), suggesting these treatments can induce catabolic processes in meniscus. The most highly upregulated proteins released from tissue stimulated with IL1ß were primarily cytokines, chemotactic factors, and inflammatory regulators (Supporting Information: Table ST1). The low catabolic response to IL1 in this study compared to observations in previous publications[39–41] may be attributed to multiple factors, particularly donor age, or due to dissimilarities between the studies species.

We studied the time-dependent effect of treatments using time-series cluster analysis and found a reoccurring pattern of peak release on Day 9 in multiple treatments, but not to the same extent in controls (Figure 7). Clusters with this pattern was particularly noticeable for treatments IL1 (MEX0-3 inner and outer zone), IL1 + IL17 (MEX3 outer zone), and OSM + TNF (MEX3 inner and outer zone). While further research would be needed to confirm these results, these initial findings suggest these treatments may have a time-dependent effect on the release of proteins as previously shown for articular cartilage.[42] Clusters of proteins with high release at the earliest time point both in treatment group and control may be an artifact of the meniscus being cut upon harvest or cut into slices.

The inner zone of the meniscus is more cartilage-like, while the outer zone, located further to the edge of the joint, is more fibrous in structure.[39,43] The difference in composition of the tissue is reflected in observed differences in differentially abundant proteins between the zones. In both zones, IL1 induces the release of inflammatory mediators such as IL6, CXCL1, CXCL6, and CXCL8, which may play a role in recruiting immune cells to the site of inflammation.[44] Specifically for the outer zone, our analysis revealed interactions between IL6 and SOD2, a protein involved in oxidative stress response, and the chemokine CCL2. Additionally, the interactions between CSTB and TNFAIP6 and CXCL1 were identified for the

outer zone. CSTB is involved in protease inhibition, which could play a role in regulating ECM degradation.[45,46] While overall there was many overlapping upregulated proteins between the zones, some treatments, such as TNF + IL6 + sILR, is seemingly preferably for inducing release of ECM proteins and proteases particularly in the inner zone (Figure 6, Supporting Information: Table ST5). However, the distinction between inner and outer zone is not clear cut and could contribute to unwanted sample variability. Given the overall similar response in both zones, studying the whole meniscus (and thus studying the average response across meniscal zones) may be the preferable approach.

Semitryptic peptides have been digested by trypsin at one end and the other terminal by an unknown protease. These peptides can be used as markers of catabolic processes, or the breakdown of cellular components.[47] The highest ratio of semitryptic peptides to all identified peptides were found in treatments TNF + IL6 + sILR (MEX3 inner zone) and OSM + TNF (MEX3 outer zone and MEX4).

Olink PEA was used both as a way to expand the MS results using a highly sensitive and targeted proteomics technique, and to confirm the presence and relative abundance of proteins identified by MS, in particularly to assess the catabolic effect of treatments IL1, OSM + TNF, and TNFIL6sILR compared to controls. One notable limitation of this method is that only proteins specifically targeted to proteins in the Olink inflammation panel can be identified and will thus not provide knowledge about other important proteins. The Olink technology is optimized for plasma and serum samples while the performance in explant media is not much explored. In agreements with the MS results, the highest number of upregulated proteins were estimated in OSM + TNF outer zone (Supporting Information: Table ST4). Seven of these proteins were upregulated in MEX3 outer zone by MS and eight proteins were only found using the Olink biomarker panel. These proteins included growth factors, cytokines, and signal receptor which may contribute to a complex interplay of proteins contributing to dysregulated proteostasis in a catabolic state.

In comparison to the proteomics results, the qPCR analysis displayed similar expression patterns for MMP proteins, while the reverse pattern was found for most of the ECM proteins. While the proteomics data showed upregulation of ACAN, COL1A1, COL3A1, and MMP1 (MMP13 was excluded due to extensive missing values in control), and downregulation of CHI3L1 in two out of the three treatments, the qPCR analysis showed downregulation of five genes including ACAN, CHI3L1, COL1A1, COL3A1, and TIMP2. Additionally, while TIMP2 was not differentially abundant in the proteomics data, the qPCR analysis showed downregulation of TIMP2 in all three treatments. The downregulation of ECM genes observed in the qPCR analysis suggests that cytokine treatment may disrupt the synthesis of ECM components as previously shown in articular cartilage at protein level[15] and at mRNA level for articular cartilage as well as in menisci.[39] Upregulation of the corresponding proteins in the proteomics analysis is attributed to the cytokine treatment leading to breakdown of ECM and release into explant culture media.

The experimental design of ex vivo meniscus explant models is influenced by existing models using articular cartilage. Cytokine treatment on nonhuman cartilage explants have been demonstrated to have an effect of increased release of ECM-proteins and proteins involved in catabolic response of cartilage degradation into the explant culture media when compared to controls.[48] Combined mechanical loading and cytokine treatment on porcine and bovine meniscus was recently reported by McNulty et al. in which two mechanical treatments were used; cell stretching and dynamic loading, using transcriptomics.[49] The inflammatory response was highly modulated by mechanical loading. In our study of menisci, we find release of ECM- and catabolic proteins, although the specific proteins released are to a large degree different than previous studies using a cartilage model. For example, in the proteomic study on bovine knee articular cartilage, MMP13 was selected as a "representative protein" of the response group "cytokines versus control."[48] By contrast, in the current study MMP13 had a seemingly negligible effect, upregulated only in IL1 versus control (MEX0-3 inner zone) and with a log2 estimate of 0.69 (lower CL 0.03, upper CL 1.35) (Supporting Information: Table ST1).

GAG-analysis may also be less informative, as GAG content in the meniscus is much lower than in cartilage.[39] Thus, it is not surprising that GAG release was much more modest in our meniscus model than previously reported for cartilage.[48]

An important strength of the current report is use of human tissues. Many previous studies on tissue breakdown were conducted in bovine cartilage[4,5,7,48,50,51] and thus generalizability of results to humans could be limited, especially given typical ages used but also different loading patterns in animal versus human joints.[52,53] We examine the effect of multiple cytokine treatments, two of which we propose as possible triggering agents of catabolism in human meniscus.

While our model allowed us to examine the effects of cytokines on the meniscus in a controlled environment, there are important limitations to consider. Mechanical loading plays a critical role in the regulatory mechanisms of the function of the meniscus and ECM remodeling. The absence of mechanical stimulation in our model limits our ability to fully capture the complex interplay between cytokine signaling and mechanical loading. Our sample size is small, but given the novel nature of the model, we think it still provides important insights into feasibility of the approach. The meniscus is a complex and delicate structure, and it can be difficult to cut slices that are consistently the same size and thickness. This potential inconsistency may lead to variability in the samples being studied. To minimize this variability, we randomized the cut-out slices into treatments. By doing so, any differences between slices were assumed to be random rather than systematic. Moreover, the meniscus is composed of different zones that gradually change

from peripheral to inner regions which can contribute to variability in the samples being studied.

7 | CONCLUSIONS

Our analysis substantiates the hypothesis that a catabolic effect of cytokine treatments on human menisci can be induced in addition to already well-documented observations of inflammatory response by conventional inflammatory cytokines in articular cartilage. Results from our time-dependent cluster analysis suggest patterns of increased meniscus degeneration and release of matrix proteins peak on Day 9 after applied cytokine treatment. Differential abundance analysis implies increased release of cytokines from IL1 treatment while further activation of proteases and inhibitors which may contribute to dysregulated proteostasis was more strongly induced by OSM + TNF and TNF + IL6 + sIL6R. These findings may be relevant in the path to clinical therapy for OA through discovery of new biomarkers for early disease progression.

AUTHOR CONTRIBUTIONS

All were involved in the conception and design of the study as well as interpretation of the data. Karin Lindblom, Viveka Tillgren, Aida Yifter-Lindgren, and Patrik Önnerfjord conducted the experimental work and data analysis. Anders Aspberg and Karin Lindblom performed the qPCR experiment. Patrik Önnerfjord conducted the MS work and data analysis. Bioinformatics and data analysis was performed by Martin Rydén supported by Aleksandra Turkiewicz. Martin Rydén and Patrik Önnerfjord drafted the manuscript while all authors critically revised the manuscript and gave final approval of the article.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement #771121), The Swedish Research Council, the ALF agreement between the Swedish government and the county, the Swedish Rheumatism Association, IngaBritt and Arne Lundberg's Research Foundation, the Alfred Österlund Foundation, the Olle Engkvist Foundation, the Greta & Johan Kock Foundation, The Foundation for Movement Disabilities in Skåne, the Anna-Greta Crafoord and the Crafoord Foundations.

ORCID

Martin Rydén http://orcid.org/0000-0001-6968-4314
Patrik Önnerfjord http://orcid.org/0000-0002-2345-2937

REFERENCES

1. Glyn-Jones S, Palmer AJR, Agricola R, et al. Osteoarthritis. *Lancet*. 2015;386(9991):376-387.
2. Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum*. 2012;64(6):1697-1707.
3. Anderson JR, Phelan MM, Foddy L, Clegg PD, Peffers MJ. Ex vivo equine cartilage explant osteoarthritis model: a metabolomics and proteomics study. *J Proteome Res*. 2020;19(9):3652-3667.
4. Sinkeviciute D, Aspberg A, He Y, Bay-Jensen AC, Önnerfjord P. Characterization of the interleukin-17 effect on articular cartilage in a translational model: an explorative study. *Bmc Rheumatology*. 2020;4(1):30.
5. Mehta S, Akhtar S, Porter RM, Önnerfjord P, Bajpayee AG. Interleukin-1 receptor antagonist (IL-1Ra) is more effective in suppressing cytokine-induced catabolism in cartilage-synovium co-culture than in cartilage monoculture. *Arthritis Res Ther*. 2019;21(1):238.
6. Swan AL, Hillier KL, Smith JR, et al. Analysis of mass spectrometry data from the secretome of an explant model of articular cartilage exposed to pro-inflammatory and anti-inflammatory stimuli using machine learning. *BMC Musculoskelet Disord*. 2013;14(1):349.
7. Ravindra KC, Ahrens CC, Wang Y, et al. Chemoproteomics of matrix metalloproteases in a model of cartilage degeneration suggests functional biomarkers associated with posttraumatic osteoarthritis. *J Biol Chem*. 2018;293(29):11459-11469.
8. Vangsness Jr. CT, Burke WS, Narvy SJ, MacPhee RD, Fedenko AN. Human knee synovial fluid cytokines correlated with grade of knee osteoarthritis--a pilot study. *Bull NYU Hosp Jt Dis*. 2011;69(2):122-127.
9. Wojdasiewicz P, Poniatowski ŁA, Szukiewicz D. The role of inflammatory and anti-inflammatory cytokines in the pathogenesis of osteoarthritis. *Mediat Inflamm*. 2014;2014:561459.
10. Goldring MB. Osteoarthritis and cartilage: the role of cytokines. *Curr Rheumatol Rep*. 2000;2(6):459-465.
11. Stone AV, Loeser RF, Vanderman KS, Long DL, Clark SC, Ferguson CM. Pro-inflammatory stimulation of meniscus cells increases production of matrix metalloproteinases and additional catabolic factors involved in osteoarthritis pathogenesis. *Osteoarthritis Cartilage*. 2014;22(2):264-274.
12. Tarafder S, Park G, Lee CH. Explant models for meniscus metabolism, injury, repair, and healing. *Connect Tissue Res*. 2019;61(3-4):1-12.
13. Cope PJ, Ourradi K, Li Y, Sharif M. Models of osteoarthritis: the good, the bad and the promising. *Osteoarthritis Cartilage*. 2019;27(2):230-239.
14. Farndale R, Buttle D, Barrett A. Improved quantitation and discrimination of sulphated glycosaminoglycans by use of dimethylmethylene blue. *Biochim Biophys Acta*. 1986;883(2):173-177.
15. Wang Y, Li Y, Khabut A, Chubinskaya S, Grodzinsky AJ, Önnerfjord P. Quantitative proteomics analysis of cartilage response to mechanical injury and cytokine treatment. *Matrix Biol*. 2017;63:11-22.
16. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1-48.
17. Lenth RV. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means. Available from: https://CRAN.R-project.org/package=emmeans
18. Ji H, Liu XS. Analyzing'omics data using hierarchical models. *Nat Biotechnol*. 2010;28(4):337-340.
19. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13(11):e1005752. Available from http://www.mixOmics.org
20. Murphy MP, Koepke LS, Lopez MT, et al. Articular cartilage regeneration by activated skeletal stem cells. *Nat Med*. 2020;26(10):1583-1592.
21. Inc PT. 2015. Collaborative data science. Available from: https://plot.ly
22. Pérez-Sanz F, Riquelme-Pérez M 2022. tscR: a time series clustering package combining slope and Frechet distances.
23. Wickham H. 2009. ggplot2, Elegant Graphics for Data Analysis.

24. Csárdi G, Nepusz T [date unknown]. The igraph software package for complex network research.
25. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2014;4(D1):D447-D452.
26. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8-22.
27. Åhrman E, Lorenzo P, Holmgren K, et al. Novel cartilage oligomeric matrix protein (COMP) neoepitopes identified in synovial fluids from patients with joint diseases using affinity chromatography and mass spectrometry*. *J Biol Chem*. 2014;289(30):20908-20916.
28. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔC T method. *Methods*. 2001;25(4):402-408.
29. Reker D, Siebuhr AS, Thudium CS, et al. Sprifermin (rhFGF18) versus vehicle induces a biphasic process of extracellular matrix remodeling in human knee OA articular cartilage ex vivo. *Sci Rep*. 2020;10(1):6011.
30. Liu-Bryan R, Terkeltaub R. Emerging regulators of the inflammatory process in osteoarthritis. *Nat Rev Rheumatol*. 2015;11(1):35-44.
31. Nagase H, Kashiwagi M. Aggrecanases and cartilage matrix degradation. *Arthritis Res Ther*. 2003;5(2):94.
32. Peng Z, Sun H, Bunpetch V, et al. The regulation of cartilage extracellular matrix homeostasis in joint cartilage degeneration and regeneration. *Biomaterials*. 2021;268:120555.
33. Maldonado M, Nam J. The role of changes in extracellular matrix of cartilage in the presence of inflammation on the pathology of osteoarthritis. *BioMed Res Int*. 2013;2013:1-10.
34. Englund M, Haugen IK, Guermazi A, et al. Evidence that meniscus damage may be a component of osteoarthritis: the Framingham study. *Osteoarthritis Cartilage*. 2016;24(2):270-273.
35. Englund M, Roemer FW, Hayashi D, Crema MD, Guermazi A. Meniscus pathology, osteoarthritis and the treatment controversy. *Nat Rev Rheumatol*. 2012;8(7):412-419.
36. Anderson DD, Chubinskaya S, Guilak F, et al. Post-traumatic osteoarthritis: improved understanding and opportunities for early intervention. *J Orthop Res*. 2011;29(6):802-809.
37. Englund M, Guermazi A, Lohmander LS. The meniscus in knee osteoarthritis. *Rheum Dis Clin North Am*. 2009;35(3):579-590.
38. Mcdevitt CA, Webber RJ. The ultrastructure and biochemistry of meniscal cartilage. *Clin Orthop Relat Res*. 1990;252:8-18.
39. Fuller ES, Smith MM, Little CB, Melrose J. Zonal differences in meniscus matrix turnover and cytokine response. *Osteoarthritis Cartilage*. 2012;20(1):49-59.
40. Riera KM, Rothfusz NE, Wilusz RE, Weinberg JB, Guilak F, McNulty AL. Interleukin-1, tumor necrosis factor-alpha, and transforming growth factor-beta 1 and integrative meniscal repair: influences on meniscal cell proliferation and migration. *Arthritis Res Ther*. 2011;13(6):R187.
41. Lemke AK, Sandy JD, Voigt H, et al. Interleukin-1α treatment of meniscal explants stimulates the production and release of aggrecanase-generated, GAG-substituted aggrecan products and also the release of pre-formed, aggrecanase-generated G1 and m-calpain-generated G1-G2. *Cell Tissue Res*. 2010;340(1):179-188.
42. Black RM, Wang Y, Struglics A, et al. Proteomic clustering reveals the kinetics of disease biomarkers in bovine and human models of post-traumatic osteoarthritis. *Osteoarthritis and Cartilage Open*. 2021;3:100191.
43. Folkesson E, Turkiewicz A, Rydén M, et al. Proteomic characterization of the normal human medial meniscus body using data-independent acquisition mass spectrometry. *J Orthop Res*. 2020;38:1735-1745.
44. Russo RC, Garcia CC, Teixeira MM, Amaral FA. The CXCL8/IL-8 chemokine family and its receptors in inflammatory diseases. *Expert Rev Clin Immunol*. 2014;10(5):593-619.
45. Park J, Lee H-S, Go E-B, et al. Proteomic analysis of the meniscus cartilage in osteoarthritis. *Int J Mol Sci*. 2021;22(15):8181.
46. Moali C, Hulmes DJ. Extracellular and cell surface proteases in wound healing: new players are still emerging. *Eur J Dermatol*. 2009;19(6):552-564.
47. Kleifeld O, Doucet A, auf dem Keller U, et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol*. 2010;28(3):281-288.
48. Black RM, Wang Y, Struglics A, et al. Proteomic analysis reveals dexamethasone rescues matrix breakdown but not anabolic dysregulation in a cartilage injury model. *Osteoarthritis and Cartilage Open*. 2020;2(4):100099.
49. Andress BD, Irwin RM, Puranam I, Hoffman BD, McNulty AL. A tale of two loads: modulation of IL-1 induced inflammatory responses of meniscal cells in two models of dynamic physiologic loading. *Front Bioeng Biotechnol*. 2022;10:837619.
50. Stevens AL, Wishnok JS, Chai DH, Grodzinsky AJ, Tannenbaum SR. A sodium dodecyl sulfate-polyacrylamide gel electrophoresis-liquid chromatography tandem mass spectrometry analysis of bovine cartilage tissue response to mechanical compression injury and the inflammatory cytokines tumor necrosis factor alpha and interleukin-1beta. *Arthritis Rheum*. 2008;58(2):489-500.
51. Stevens AL, Wishnok JS, White FM, Grodzinsky AJ, Tannenbaum SR. Mechanical injury and cytokines cause loss of cartilage integrity and upregulate proteins associated with catabolism, immunity, inflammation, and repair*. *Mol Cell Proteomics*. 2009;8(7):1475-1489.
52. Zevenbergen L, Gsell W, Cai L, et al. Cartilage-on-cartilage contact: effect of compressive loading on tissue deformations and structural integrity of bovine articular cartilage. *Osteoarthritis Cartilage*. 2018;26(12):1699-1709.
53. Temple DK, Cederlund AA, Lawless BM, Aspden RM, Espino DM. Viscoelastic properties of human and bovine articular cartilage: a comparison of frequency-dependent trends. *BMC Musculoskelet Disord*. 2016;17(1):419.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rydén M, Lindblom K, Yifter-Lindgren A, et al. A human meniscus explant model for studying early events in osteoarthritis development by proteomics. *J Orthop Res*. 2023;1-14. doi:10.1002/jor.25633

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Sciences

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2023:120
ISBN 978-91-8021-461-2
ISSN 1652-8220