Deconvoluting synovial fluid molecular endotypes in knee osteoarthritis: primary results from the STEpUP OA Consortium

Authors: T.A.Perry^{*}", Y.Deng^{*"}, P.Hulley², R.A.Maciewicz², J.Mitchelmore⁹, S.Larsson⁻, J.Gogain⁹
S.Brachat³, A.Struglics⁴, C.T.Appleton⁶, S.Kluzek^{2,7}, N.K.Arden^{2,8}, A.J.Price², D.Felson⁹, L.Bond ,
,
, S.Brachat", A.Struglics", C.T.Appleton", S.Kluzek^{2,}', N.K.Arden²'', A.J.Price², D.Felson", L.Bondi²⁰,
M.Kapoor¹¹, L. S.Lohmander⁴, T.J.Welting¹², D.A.Walsh^{13,14}, A.M.Valdes¹³, the STEpUP OA
Consortium, L.

M.Kapoor¹¹, L. S.Lohmander", T.J.Welting¹¹, D.A.Walsh^{13,14}, A.M.Valdes¹⁹, the STEpUP OA
Consortium, L.Jostins-Dean¹⁸, F.E.Watt^{15, 18}, B.D.M.Tom¹⁰⁸, T.L.Vincent^{18*}.
¹ Centre for Osteoarthritis Pathogenesis Consortium, L.Jostins-Dean²⁶, F.E.Watt^{25, 24}, B.D.M.Tom²⁰⁴, T.L.Vincent²⁶.
¹ Centre for Osteoarthritis Pathogenesis Versus Arthritis, Kennedy In
NDORMS, University of Oxford, Oxford, UK.
² Nuffield Department o 1
|
(

² Centre for Osteoarthritis Pathogenesis Versus Arthritis, Kennedy Institute of Rheumatology,
NDORMS, University of Oxford, Oxford, UK.
² Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, Northeld Department of Orthopaedics, Rhendrik Protocolar
Noxford, Oxford, UK.
Novartis Biomedical Research, Basel, Switz
4 Faculty of Medicine, Department of Clini 1

Nuffield Department of Orthopaedical, Switzerland.
Nuffield Department of Clinical Sciences Lund, Orthopaedics, Lund University, Lund, Weden. ³ Novartis Biomedica
⁴ Faculty of Medicin
Sweden.
⁵ SomaLogic, Boulde ³ Novartis Biomedical Research, Basel, Switzerland.
⁴ Faculty of Medicine, Department of Clinical Sciel
Sweden.
⁵ SomaLogic, Boulder, Colorado, USA.
⁶ Bone and Joint Institute, University of Western On^t T, Faculty of Medicine, Department of Clinical Sciences Lundy, Orthopaedics, Lund University, Lund, 2008.

SomaLogic, Boulder, Colorado, USA.

Bone and Joint Institute, University of Western Ontario, London, Ontario, Canada.

⁵ SomaLo
⁶ Bone an
⁷ NIHR
Osteoartl ⁵ SomaLogic, Boulder, Colorado, USA.
⁶ Bone and Joint Institute, University c
⁷ NIHR Nottingham Biomedical R
Osteoarthritis Centre, University of No
⁸ Centre for Sport, Exercise and Os ⁸ Bone and Joint Institute, University of Western Ontario, London, Ontario, Canada.
⁷ NIHR Nottingham Biomedical Research Centre and Versus Arthritis Spo
Osteoarthritis Centre, University of Nottingham, Nottingham, UK. \overline{a}

Deteoarthritis Centre, University of Nottingham, Nottingham, UK.

Centre for Sport, Exercise and Osteoarthritis Research Versus Arthritis, University of Oxford,

Dxford, UK.

Section of Rheumatology, Boston University Scho ⁸ Centre for Sport, Exercise and Osteoarthritis Research Versu
Oxford, UK.
⁹ Section of Rheumatology, Boston University School of Medicine,
¹⁰ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.

0xford, UK.
⁹ Section of Rheumatology, Boston University School of Medicine, Boston, Massachusetts, USA.
¹⁰ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
¹¹ Schroeder Arthritis Institute. University

⁹ Section of P
⁹ Section of P
¹⁰ MRC Biost
¹¹ Schroeder
¹² Laborato ⁹ Section of Rheumatology, Boston University School of Medicine, Boston, Massachusetts, USA.
¹⁰ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
¹¹ Schroeder Arthritis Institute, University Health Netw ¹¹ Schroeder Arthritis Institute, University of Cambridge, Cambridge, UK.
¹¹ Schroeder Arthritis Institute, University Health Network, Toronto
¹² Laboratory for Experimental Orthopedics, Department of
University, Maa ¹¹ Schroeder Arthritis Institute, University Health Network, Toronto, Ontario, Canada.
¹² Laboratory for Experimental Orthopedics, Department of Orthopedic Surge
University, Maastricht, Netherlands.
¹³ Pain Centre Ve

¹¹ Laboratory for Experimental Orthopedics, Department of Orthopedic Surgery, Maastricht
University, Maastricht, Netherlands.
¹³ Pain Centre Versus Arthritis, Advanced Pain Discovery Platform, and the NIHR Nottingham
B ¹³ Pain Centre Versus Arthritis, Ac
Biomedical Research Centre, Univers
¹⁴ Sherwood Forest Hospitals NHS For
¹⁵ Department of Immunology and In ²³ Pain Centre Versus Arthritis, Advanced Pain Discovery Platform, and the NIHR Nottingham
Biomedical Research Centre, University of Nottingham, Nottingham, UK.
¹⁴ Sherwood Forest Hospitals NHS Foundation Trust, Sutton

Biomedical Research Centre, University of Nottingham, Nottingham, UK.
¹⁴ Sherwood Forest Hospitals NHS Foundation Trust, Sutton in Ashfield, UK.
¹⁵ Department of Immunology and Inflammation, Imperial College London, I
 ¹⁷ Sherwood Forest Hospitals NHS Foundation Trust, Sutton in Ashfield, UK.
¹⁵ Department of Immunology and Inflammation, Imperial College London, I
[#] Joint first authors
[&] Joint senior authors ¹⁵ Department of Immunology and Inflammation, Imperial College London, London, UK.
Joint first authors
[&] Joint senior authors

#
8

" Joint first authors
[&] Joint senior autho
TE: This preprint reports [&] Joint senior authors

*Corresponding author: tonia.vincent@kennedy.ox.ac.uk (TLV).

—
|
| Keywords: biomarker, proteomic, synovial fluid, osteoarthritis, knee injury, endotype, stratification,
SomaLogic. Keywords: biomarker, proteomic, synovial fluid, osteoarthritis, knee injury, endotype, stratification,
SomaLogic.
^Membership of the STEpUP OA Consortium is provided in the *STEpUP OA Consortium author block*

Running title: Endotype discovery in OA

Membership of the STEpUP OA Consori

section of the manuscript. Running title: Endotype discovery in OA

AMembership of the STEpUP OA Consort

section of the manuscript. Nembership of the STEPUP OA Consortium is provided in the STEPUP OA Consortium author block
section of the manuscript. section of the manuscript.

ABSTRACT

Background

Variability suggests that it could be more than one disease. Synovial fluid To detect Endotypes by
Unbiased Proteomics in OA (STEpUP OA) was established to test the hypothesis that there are
detectable distinct molecular e Unbiased Proteomics in OA (STEpUP OA) was established to test the hypothesis that there are
detectable distinct molecular endotypes in knee OA.
 $\frac{1}{2}$ <u>Unbiased Proteomics in OA (STEPUP OA)</u> was established to test the hypothesis that there are
detectable distinct molecular endotypes in knee OA.
Methods

Methods

detectable distinct molecular endotypes in knee OA.
Methods
OA knee synovial fluid (SF) samples (N=1361) were from pre-existing OA cohorts with cross-sectional |
|
| Clinical (radiographic and pain) data. Samples were divided into Discovery (N = 708) and Replication
(N=653) datasets. Proteomic analysis was performed using SomaScan V4.1 assay (6596 proteins).
Unsupervised clustering was (N=653) datasets. Proteomic analysis was performed using SomaScan V4.1 assay (6596 proteins).
Unsupervised clustering was performed using k-means, assessed using the f(k) metric, with and without adjustments for potential Unsupervised clustering was performed using k-means, assessed using the f(k) metric, with and
without adjustments for potential confounders. Regression analyses were used to assess protein
associations with radiographic (K without adjustments for potential confounders. Regression analyses were used to assess protein associations with radiographic (Kellgren and Lawrence) and knee pain (WOMAC pain), with and
without stratification by body mass index (BMI) or biological sex. Adjustments were made for cohort
(random intercept) or intracel without stratification by body mass index (BMI) or biological sex. Adjustments were made for cohort (random intercept) or intracellular protein, using an intracellular protein score (IPS). Analyses were carried out in R ac random intercept) or intracellular protein, using an intracellular protein score (IPS). Analyses were carried out in R according to a pre-published plan. (random intervals) or intervals or intervals or intervals are published plan.
Results
Results

Results

carried out in R according to a pre-published plan.
Results
No distinct SF molecular endotypes were identified in OA but two indistinct clusters were defined in |
|
| non-IPS regressed data which were stable across subgroup analyses. Clustering was lost after IPS
regression adjustment. Strong, replicable protein associations were observed with radiographic
disease severity, which were r non-IPS regression adjustment. Strong, replicable protein associations were observed with radiographic
disease severity, which were retained after adjustment for cohort or IPS. Pathway analysis identified
a strong "epithel regression angularities in the gregority protein adjustment for cohort or IPS. Pathway analysis identified
a strong "epithelial to mesenchymal transition (EMT)" pathway, and weaker associations with
"angiogenesis", "comple a strong "epithelial to mesenchymal transition (EMT)" pathway, and weaker associations with
"angiogenesis", "complement" and "coagulation". The latter were variably lost after adjustment for
BMI or biological sex. Associat angiogenesis", "complement" and "coagulation". The latter were variably lost after adjustment for
"BMI or biological sex. Associations with patient reported pain were weaker. angiogenesis", "complement" and "coagulation". The latter were veaker.
BMI or biological sex. Associations with patient reported pain were weaker.
Conclusion BMI or biological sex. Associations with patient reported patient relations with the weaker.

\overline{a} Conclusion

These data support knee OA as a biologically continuous disease in which disease severity is associated with a strong, robust, tissue remodelling signature. Subtle differences were found in pathways after stratification by pathways after stratification by BMI or sex.
 $\frac{d}{dt}$ or sex. pathways after stratification by BMI or sex.

BACKGROUND

Osteoarthritis (OA) of the knee is common, affecting up to a third of adults aged 60 years or older¹⁴¹.
Characterised by failure of the synovial joint, OA is a major contributor to healthcare costs and is a
leading cause Eading cause of disability, manifesting as a spectrum of symptoms including chronic pain and
limitations in function. Age and obesity are important risk factors, both of which have contributed to
increasing disease burden limitations in function. Age and obesity are important risk factors, both of which have contributed to
increasing disease burden across global populations^[2-4]. There are currently no approved treatments
for knee OA tha Increasing disease burden across global populations^[2-4]. There are currently no approved treatments
for knee OA that effectively target structural disease and those that target symptomatic disease
have modest efficacy a for knee OA that effectively target structural disease and those that target symptomatic disease
have modest efficacy and are associated with adverse events^[5, 6]. There remains, therefore, a major
unmet clinical need.
 have modest efficacy and are associated with adverse events^[5, 6]. There remains, therefore, a major

have modest efficacy and are associated with adverse events^{15, 6}. There remains, therefore, a major
unmet clinical need.
Limited understanding of disease pathogenesis coupled with a failure to translate findings from bas Limited understandir
research to clinical s
is the broad clinical s
whether it is driven l The search to clinical settings has hampered clinical translation in $OA^{[7, 8]}$. Another significant challenge
is the broad clinical spectrum of disease that has led many to question whether OA is one disease, or
whether research to clinical settings has hampered clinical translation in OA^[7, 8]. Another significant challenge
is the broad clinical spectrum of disease that has led many to question whether OA is one disease, or
whether it is the broad clinical spectrum of disease that has led many to question matrice or or the disease)^[9, 10].
Multiple clinical phenotypes have been suggested in the literature^[11-13], but these have not been validated as whether it is driven by multiple different pathways that converge on a common joint pathology^{e, 20}.
Multiple clinical phenotypes have been suggested in the literature^[11-13], but these have not been
validated as clinic Multiple clinical phenotypes have been suggested in the literature¹²², but these have not been
validated as clinically useful stratification tools either when testing treatment responses or as
predictors of disease prog predictors of disease progression^[14-16]. Endotypes, defined by distinct molecular signatures, may
have higher value, and could in part explain observable characteristics of a phenotype^[17].
Recent advances in underst

predictors of disease progression¹⁴⁴⁻⁴⁵. Endotypes, defined by distinct molecular signatures, may
have higher value, and could in part explain observable characteristics of a phenotype^[17].
Recent advances in understan have higher value, and could in part explain observable characteristics of a phenotype¹²⁷¹.
Recent advances in understanding complex disease have been greatly enhanced by the
of multi-omic approaches to disease relevant Ferricant advances in the set of multi-omic approaches to disease relevant tissues^[11, 18]. The strengths of these approaches are the focus on human disease cohorts at scale (hundreds to tens of thousands of participatin of multi-omic approaches to disease relevant tissues^{[22, 23}]. The strengths of these approaches are the focus on human disease cohorts at scale (hundreds to tens of thousands of participating individuals), the unbiased a for the unbiased and systematic nature of molecular identification, the ability to map molecules to a shared pathway, and the ability to replicate results across independent cohorts. Technological advances in genomics, tra shared pathway, and the ability to replicate results across independent cohorts. Technological
advances in genomics, transcriptomics and proteomics have enabled such studies to be carried out
with low tissue volumes and at shared pathway, and the ability of studies to the carried out
with low tissue volumes and at an affordable cost.
To date, the majority of studies that have attempted to identify molecular subgroups in OA have
used systemic

and any sensure in genomics, transcriptomics and protections that is that to be carried to the carried such studies
To date, the majority of studies that have attempted to identify molecular subgroups in OA have
used syste To date, the majority of studies that have attem
used systemic samples derived from blood (ser
contrast, offers a promising alternative discovery t
the joint and is enriched with locally derived bi In the systemic samples derived from blood (serum or plasma)^[19-21]. The synovial fluid (SF), in contrast, offers a promising alternative discovery tissue, as it has proximity to the diseased tissues of the joint and is contrast, offers a promising alternative discovery tissue, as it has proximity to the diseased tissues of
the joint and is enriched with locally derived biomolecules. Thus, it is likely to represent more
accurately the sev the joint and is enriched with locally derived biomolecules. Thus, it is likely to represent more
accurately the severity of disease in that given joint. We have also previously shown that proteins
regulated in knee OA or accurately the severity of disease in that given joint. We have also previously shown that proteins
regulated in knee OA or after knee injury, compared with healthy controls, are readily detected in
the SF but correlate po regulated in knee OA or after knee injury, compared with healthy controls, are readily detected in
the SF but correlate poorly in paired blood^[22-25]. Furthermore, we have previously confirmed the
utility of high scale p the SF but correlate poorly in paired blood^[22-25]. Furthermore, we have previously confirmed the utility of high scale protein measurements in SF using the SomaScan® platform (SomaLogic, Inc, the SF but correlate poorly in paired blood⁽²²⁻²⁵⁾. Furthermore, we have previously confirmed the
utility of high scale protein measurements in SF using the SomaScan® platform (SomaLogic, Inc,
inc., Inc., \mathcal{L}_{max} scale protein measurements in SF using the Soma \mathcal{L}_{max}

Boulder, Colorado), an aptamer-based assay^(26, 27). The SomaScan® platform V4.1 measures over
6596 distinct human proteins.
The <u>S</u>ynovial fluid <u>T</u>o detect <u>E</u>ndotypes by <u>U</u>nbiased <u>P</u>roteomics in OA (STEpUP OA) Consor The Synovial fluid To detect Erestablished to test the primary
Knee OA. We set out to perform
With established OA where established to test the primary hypothesis that there are detectable distinct molecular endotypes in
knee OA. We set out to perform an unsupervised analysis of a single SF sample from 1361 individuals
with established OA w knee OA. We set out to perform an unsupervised analysis of a single SF sample from 1361 individuals
with established OA where cross-sectional clinical data were also available. The standardised
protocol, which describes th with established OA where cross-sectional clinical data were also available. The standardised
protocol, which describes the cohorts in detail, and includes how we adjusted for pre-defined
technical and other confounding fa protocol, which describes the cohorts in detail, and includes how we adjusted for pre-defined
technical and other confounding factors is available elsewhere^[27]. Here we present the primary
analysis of STEpUP OA, in whic protocol, which describes the confounding factors is available elsewhere^[27]. Here we present the primary analysis of STEpUP OA, in which we determine whether protein molecular endotypes exist in the SF of participants w analysis of STEpUP OA, in which we determine whether protein molecular endotypes exist in the SF
of participants with established knee OA, and further explore the relationship between proteomic
signatures and structural an of participants with established knee OA, and further explore the relationship between proteomic of participants with established knee OA, and structural and symptomatic disease.
 METHODS

METHODS

symptomatic and structural and symptomatic disease.
METHODS
Study Design principles Study Design principles

 $\frac{1}{2}$ STEpUP OA is an international Consortium, set up to search for molecular endotypes in knee OA
utilising existing demographic factors including age, biological sex (verified through assessing the
correlation between clinici correlation between clinician-reported sex and four established sex biomarkers (PSA, FSH, LH and
beta HCG)), body mass index (BMI) and clinical data (harmonised patient reported knee pain
measures and radiographic scores) beta HCG)), body mass index (BMI) and clinical data (harmonised patient reported knee pain
measures and radiographic scores) as well as matched knee SF samples (Supplementary Table 1).
STEpUP OA utilised data and samples f measures and radiographic scores) as well as matched knee SF samples (Supplementary Table 1).
STEpUP OA utilised data and samples from 17 cohorts, including N = 1780 SF samples from 1676
individuals with established knee O measure and radiographic server, as well as matched mixed of samiples (supprementary) radio after the stablished knee OA (by x-ray or knee joint symptoms), at risk of knee OA (following acute knee injury), or from control STEPUP OF AMMEED AND EMPTED MOMET CENTER, MEDDING AT THE COMMITTED MOMET IN 2019
individuals with established knee OA (by x-ray or knee joint symptoms), at risk of knee OA (following
acute knee injury), or from control sam individuals with established knee $\frac{1}{2}$ (by x-ray or knee yingle yingle kny) at risk of the extractions.
acute knee injury), or from control samples (disease-free or inflammatory arthritis participants). All
participa participants gave written informed consent with local (institution specific) ethical approvals in place.
Following the QC procedure, 1361 samples were identified from unique participants with established
OA^[27]. Individ Following the QC procedure, 1361 samples were identified from unique participants with established OA^[27]. Individual cohorts were assigned, *a priori*, into Discovery (N = 708) and Replication (N = 653) datasets (Suppl $OA^{[27]}$. Individual cohorts were assigned, *a priori*, into Discovery (N = 708) and Replication (N = 653) datasets (Supplementary Table 1). Most samples were spun after joint aspiration but appropriate correction was app OA⁽²⁷⁾. Individual cohorts were assigned, *a priori*, into Discovery (N = 708) and Replication (N = 653)
datasets (Supplementary Table 1). Most samples were spun after joint aspiration but appropriate
correction was appl correction was applied when unspun samples were included in analyses. Full details of the cohorts
and their associated metadata, how SF was collected and processed prior to SomaScan analysis, as
well as how we corrected fo and their associated metadata, how SF was collected and processed prior to SomaScan analysis, as
well as how we corrected for predefined technical and other confounders can be found in Deng *et*
al. 2023^[27]. The primary and their confounders can be found in Deng et
al. 2023^[27]. The primary Discovery statistical analysis was pre-specified and cross-sectional (Data
Analysis Plan, see link below). al. 2023^[27]. The primary Discovery statistical analysis was pre-specified and cross-sectional (Data Analysis Plan, see link below). al. 2023¹²⁷¹. The primary Discovery statistical analysis was pre-specified and cross-sectional (Data
Analysis Plan, see link below).
. Analysis Plan, see link below).

Sample numbers and SOMAmers®(28) in the presented experiments varied according to data
availability and analysis performed.
Analysis platform availability and analysis performed.
Analysis platform
All SF samples were analysed on the Discovery Plex V4.1 (SomaLogic, Inc, Boulder, Colorado); a high-

Analysis platform

 $\frac{1}{t}$ $\mathcal{L}_{\mathcal{A}}$ synthetic DNA slow off-rate modified aptamers (SOMAmers®) (7289 unique human targets)^[29]. All SF
samples were randomized and analysed as a single batch at SomaLogic's laboratory in Boulder, Co,
USA. synthetic DNA slow off-rate modified aptamers (SOMAmers®) (7289 unique human targets)⁽²⁹⁾. All SF
samples were randomized and analysed as a single batch at SomaLogic's laboratory in Boulder, Co,
USA. samples were randomized and analysis were randomized as a single batch at Soma
Statistical Analysis

Statistical Analysis

Statis
Quali

All proteomic data received frontiers as previously reported
SomaLogic's normalization pipeline
aptamers of insufficient quality to procedures as previously reported^[27]. Briefly, raw data was standardised using a modified version of SomaLogic's normalization pipeline and batch-effect correction, followed by removal of samples and aptamers of insuffi SomaLogic's normalization pipeline and batch-effect correction, followed by removal of samples and
aptamers of insufficient quality to produce our initial downstream dataset for future analyses. All
statistical analyses we aptamers of insufficient quality to produce our initial downstream dataset for future analyses. All
statistical analyses were pre-specified and outlined in our data analysis plans (see below).
Unsupervised clustering for e aptainstands of the intervalstand outlined in our data analysis plans (see below).

Statistical analyses were pre-specified and outlined in our data analysis plans (see below).

Unsupervised clustering for endotype detecti

Dimension reduction on batch-corrected, log-transformed proteomic data was performed using <u>์เ</u>
[University
Dimension reduction on batch-corrected, log-t
unscaled Principal Component Analysis (PCA),
variation. Unsupervised clustering was perforr
clustering with 10 sets of random starting value unscaled Principal Component Analysis (PCA), with the top principal components explaining 80%
variation. Unsupervised clustering was performed on the reduced feature space using k-means
clustering with 10 sets of random st variation. Unsupervised clustering was performed on the reduced feature space using k-means
clustering with 10 sets of random starting values. We tested for the presence of significant clusters
using the $f(K)$ statistic[[] clustering with 10 sets of random starting values. We tested for the presence of significant clusters
using the $f(K)$ statistic^[30]; with the $f(K)$ statistic visualised across cluster numbers. Data were
determined to be using the $f(K)$ statistic^[30]; with the $f(K)$ statistic visualised across cluster numbers. Data were determined to be significantly clustered if, for any number of clusters K, $f(K) < 0.85$ (*a priori* specified).
Elbow p using the f(K) statistic¹³⁰; with the f(K) statistic visualised across cluster numbers. Data were
determined to be significantly clustered if, for any number of clusters K, f(K)<0.85 (*a priori* specified).
Elbow plots w Elbow plots were constructed to test the robustness of our findings. If the data were significantly clustered, we picked the optimal cluster number by majority vote across different clustering metrics
(as implemented in the R package *NbClust^[31], version*: 3.0.1) for downstream analyses.
Data visualisation and present (as implemented in the R package NbClust^[31], version: 3.0.1) for downstream analyses. (as implemented in the R package *NbClust^{s31}),* version: 3.0.1) for downstream analyses.
Data visualisation and presentation
Data visualisation and presentation

\mathbf{I} Data visualisation and presentation

Approximation and Projection for Dimension Reduction $(UMAP)^{[32]}$ plots.
Protein–clinical feature association testing

Approximation and Projection for Dimension Reduction (UMAP)⁽⁵²⁾ plots.
Protein—clinical feature association testing
Associations between protein expression and clinical outcomes were m_' $\frac{1}{2}$ Associations between protein expression a
models for each SOMAmer separately, wit
expression for each protein set as the ind
ordinal regression models were fitted f models for each SOMAmer separately, with clinical features set as the dependent variable and log-
expression for each protein set as the independent variable. Linear, logistic or proportional odds
ordinal regression models meing the models, protein expression values were transformed dang nutural logalitimis and were expression models were fitted for continuous, binary or ordered categorical variable
outcomes respectively. Residual diagnostics confirmed adequacy of model assumptions. Before
fitting the models, protein expression values outcomes respectively. Residual diagnostics confirmed adequacy of model assumptions. Before
fitting the models, protein expression values were transformed using natural logarithms and were
standardized on a per protein bas outcomes respectively. Residual angulation committed analytics, or model adequation and were
fitting the models, protein expression values were transformed using natural logarithms and were
standardized on a per protein ba standardized on a per protein basis (within Discovery, Replication and Combined datasets) by
subtracting mean log protein abundance and then dividing by its standard deviation, to make the
slopes comparable between models. subtracting mean log protein abundance and then dividing by its standard deviation, to make the
slopes comparable between models. The resulting beta estimates (from linear regression models) or
log odds ratios (from logist slopes comparable between models. The resulting beta estimates (from linear regression models) or
log odds ratios (from logistic and ordinal models) can be interpreted respectively as either mean
outcome change or log odd slopes comparable between models and ordinal models) can be interpreted respectively as either mean
outcome change or log odds ratio per standard deviation change in the log protein abundance.
Replication was defined as pr outcome change or log odds ratio per standard deviation change in the log protein abundance.
Replication was defined as proteins that were significant at Benjamini-Hochberg adjusted p-value \le
0.05 in both Discovery and Replication was defined as proteins that were significant at Benjamini-Hochberg adjusted p-value \leq 0.05 in both Discovery and Replication datasets and with effects in the same direction.
All primary regression models

biological sex-stratified analyses that were adjusted for age only, and regression models exploring
associations with BMI, which were adjusted for biological sex and radiographic disease status). All All primary regression models were adjusted for age and biological sex (with the
biological sex-stratified analyses that were adjusted for age only, and regression mo
associations with BMI, which were adjusted for biologi biological sex-stratified analyses that were adjusted for age only, and regression models exploring
associations with BMI, which were adjusted for biological sex and radiographic disease status). All
analyses were batch c associations with BMI, which were adjusted for biological sex and radiographic disease status). All
analyses were batch corrected for spin-status (using the R function $ComBat^{[33, 34]}$) and run in
duplicate using either prot analyses were batch corrected for spin-status (using the R function $ComBat^{[33, 34]}$) and run in
duplicate using either proteomic data that had undergone further regression adjustment for
intracellular protein score (IPS)^{[2} analyses were batch corrected for spin-status (using the R function *ComBat^{103, 341}*) and run in
duplicate using either proteomic data that had undergone further regression adjustment for
intracellular protein score (IPS intracellular protein score (IPS)^[27] ('IPS regressed' analyses) or without ('non-IPS regressed').
Association testing between IPS, that had been transformed using natural logarithms, and
demographic, clinical and techni Association testing between IPS, that had been transformed using natural logarithms, and
demographic, clinical and technical features was performed using regression modelling, with all
analyses either non-adjusted or adjus demographic, clinical and technical features was performed using regression modelling, with all
analyses either non-adjusted or adjusted for cohort (as a random intercept). Volcano plots were
generated to display associate analyses either non-adjusted or adjusted for cohort (as a random intercept). Volcano plots were
generated to display associated proteins from the regression analyses, with the most strongly
positively and negatively associ analyses either non-adjusted or adjusted or a random intercept). There are presented to display associated proteins from the regression analyses, with the most strongly positively and negatively associating proteins labell generatively and negatively associating proteins labelled by their given SomaLogic protein target name.
The most significantly associated proteins, ordered by their adjusted p-value, were labelled. A small
number of protei positively and negatively and negatively their adjusted p-value, were labelled. A small number of proteins (non-IPS & COMBAT corrected for spin-status filtered ist: N = 383, IPS & COMBAT corrected for spin-status filtered The most significant (non-IPS & COMBAT corrected for spin-status filtered list: N = 383, IPS & COMBAT corrected for spin-status filtered list: N = 375), had more than one detection SOMAmer on the platform. Where this was t COMBAT corrected for spin-status filtered list: $N = 375$), had more than one detection SOMAmer on
the platform. Where this was the case, only one SOMAmer was labelled on the volcano plot (i.e. the
most significant based o The platform. Where this was the case, only one SOMAmer was labelled on the volcano plot (i.e. the
most significant based on ranked adjusted p-value). We also conducted interaction testing for
most significant based on ran most significant based on ranked adjusted p-value). We also conducted interaction testing for most symmetric anti-value). We also conducted p-value). We also conducted interaction testing for
a conducted p-value of the system of the
conducted p-v associations between protein abundance and clinical features of disease. A protein abundance-bybiological sex interaction term was included to test explicitly whether biological sex modified the
association between protein abundance and WOMAC knee pain. Similarly, a protein abundance-by-
obesity status (a dichotomo obesity status (a dimensionless candidate), BMI 2 33) interaction term was included to examine it
associations with advanced radiographic status were modulated by protein abundance differences
above and below this clinical associations with advanced radiographic status were modulated by protein abundance differences above and below this clinically relevant BMI threshold. Pre-specified clinical outcomes used in association testing are listed in (Supplementary Table 1). association testing are listed in (Supplementary Table 1).
association testing are listed in (Supplementary Table 1).
Pathway enrichment analysis association testing are listed in (Supplementary Table 1).

association testing are listed in (Supplementary Table 2).
Pathway enrichment analysis
We tested for enrichment of associated proteins within |
|
| Mathway Entertainment of
Molecular Signatures Databa
Hallmark, Gene Ontology (GO
All proteins were mapped to Molecular Signatures Database (MSigDB, <u>https://www.gsea-msigdb.org/gsea/msigdb</u>); specifically,
Hallmark, Gene Ontology (GO), Reactome, and Kyoto Encyclopaedia of Genes and Genomes (KEGG).
All proteins were mapped to the Mallmark, Gene Ontology (GO), Reactome, and Kyoto Encyclopaedia of Genes and Genomes (KEGG).
All proteins were mapped to the corresponding gene set based on 'EntrezGeneSymbol', 'Target' or
'EntrezGenelD' variables provided All proteins were mapped to the corresponding gene set based on 'EntrezGeneSymbol', 'Target' or
'EntrezGenelD' variables provided by SomaLogic. Protein set enrichment testing was performed
using the *fgsed*^[35] package i The corresponding gene set and to the corresponding the corresponding the fase of 'EntrezGeneID' variables provided by SomaLogic. Protein set enrichment testing was performed using the fased^[35] package in R (version: 1. using the *fgsea*^[35] package in R (version: 1.28.0) to identify pathways whose genes were enriched for association with a given outcome. All proteins featured in the respective regression models were ranked by a 'rank using the *fgsea*⁽³⁵⁾ package in R (version: 1.28.0) to identify pathways whose genes were enriched for
association with a given outcome. All proteins featured in the respective regression models were
ranked by a 'rank m ranked by a 'rank metric' calculated as; rank metric = $-\log(p \text{-values}) * \text{sign}(\text{beta estimate or log odds})$
ratio per standard deviation). The sign function returns +1 if the estimate is positive, -1 if it is
negative, and 0 if it is zero thereby c ratio per standard deviation). The sign function returns +1 if the estimate is positive, -1 if it is
negative, and 0 if it is zero thereby capturing the direction of effect (whether the feature is
upregulated or downregula regative, and 0 if it is zero thereby capturing the direction of effect (whether the feature is
upregulated or downregulated). Enrichment scores were calculated as the maximum value of the
running sum and normalized relati upregulated or downregulated). Enrichment scores were calculated as the maximum value of the
running sum and normalized relative to pathway size, resulting in Normalized Enrichment Scores
(NES). Direction and magnitude of running sum and normalized relative to pathway size, resulting in Normalized Enrichment Scores
(NES). Direction and magnitude of pathway enrichment for a given outcome (i.e. differential
regulation of the pathway) was dete (NES). Direction and magnitude of pathway enrichment for a given outcome (i.e. differential
regulation of the pathway) was determined using the NES score; with positive values representing
positively associated pathways w regulation of the pathway) was determined using the NES score; with positive values representing
positively associated pathways whilst negative values represented negatively associated pathways.
The *gaplot2*^[36] R packa regulatively associated pathways whilst negative values represented negatively associated pathways.
The *gaplot* 2^[36] R package (version: 3.5.0) was used to draw bubble plots and visualise results.
Protein-protein inter

The *ggplot* 2^[36] R package (version: 3.5.0) was used to draw bubble plots and visualise results.
Protein-protein interaction (PPI) networks were constructed using the Search Tool for the Retrieval
of Interacting Genes/ The *ggplot2*¹³⁶³ R package (version: 3.5.0) was used to draw bubble plots and visualise results.
Protein-protein interaction (PPI) networks were constructed using the Search Tool for the R
of Interacting Genes/Proteins of Interacting Genes/Proteins database (STRING version 11.5, $\frac{https://string-db.org/}{https://string-db.org/})$. The filter condition was set as follows: network type selected; "full-STRING network"; confidence ≥ 0.2 -0.4. condition was set as follows: network type selected; "full-STRING network"; confidence ≥ 0.2 -0.4.
Statistical Significance

condition was set as follows: network the statistical Significance
Fearson correlation and relevant p-values are given for both correlation testing and regress <u>י</u>
|
ו Statistical Controllering
Statistical Significan
Significance
Significances
Significances modelling. All analyses were carried out in R (version 4.3.2), unless otherwise stated (R Core Team. modelling. All analyses were carried out in R (version 4.3.2), unless otherwise stated (R Core Team.

(2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/). Statistical significance was defined using Benjamini-Hochberg^[37] corrected p-values adjusted https://www.R-project.org/). Statistical significance was defined using Benjamini-Hochberg^{es,}
corrected p-values adjusted for multiple testing, at a false discovery rate (FDR) of 5% (padj ≤ 0.05).
Data Analysis Plan: <u>ht</u>

correctivity practice adjusted for multiple testing, at a false discovery rate (FDR) of 5% (padj ≤ 0.05).
Data Analysis Plan: <u>https://www.kennedy.ox.ac.uk/oacentre/stepup-oa</u>
RESULTS
Endotype Detection in OA SF |
|
| RESULTS

Endotype Detection in OA SF

To search for molecular endotypes in OA using SF protein profiles, the f(K) cluster metric was
employed. We had previously reported that a large contributor of variance in the initial processed To state for molecular endotypes in OA and got protein promas, the r(a) state metric metric interesting employed. We had previously reported that a large contributor of variance in the initial processed data (principal com data (principal component 1, accounting for 48% of variance), was due to intracellular proteins^[27].
Appreciating that the intracellular protein signature could obscure subtle clustering patterns within
the data, we perf data (principal component 1, accounting for 48% of variance), was due to intracellular proteins¹²⁷¹.
Appreciating that the intracellular protein signature could obscure subtle clustering patterns within
the data, we perf the data, we performed cluster analyses with and without regression adjustment for intracellular
protein^[27], using an intracellular protein score (IPS) that correlated highly with principal component 1
(r = 0.94)^[27]. protein^[27], using an intracellular protein score (IPS) that correlated highly with principal component 1
(r = 0.94)^[27]. Cluster analysis revealed 2 clusters that were evident within Discovery, Replication and
Combine protein⁽²⁷⁾, using an intracellular protein score (IPS) that correlated highly with principal component 1
(r = 0.94)^[27]. Cluster analysis revealed 2 clusters that were evident within Discovery, Replication and
Combine protein^[27], using an intracellular protein score (IPS) that correlated highly with principal component 1 (r = 0.94)⁽²⁷⁾. Cluster analysis revealed 2 clusters that were evident within Discovery, Replication and
Combined datasets for the non-IPS regressed analysis (Figure 1A, left panel). In contrast, no clusters
were detecte Were detected in the IPS-regressed dataset (Figure 1A, right panel). Visualisation of the proteomic
data structure in two-dimensional space showed that the two clusters were indistinct and could be
defined by dichotomising data structure in two-dimensional space showed that the two clusters were indistinct and could be
defined by dichotomising the continuous IPS, a feature that was lost after IPS regression (Figure 1B).
Association testing o

defined by dichotomising the continuous IPS, a feature that was lost after IPS regression (Figure 1B).
Association testing of IPS with pre-defined clinical and technical features (N = 1134, spun OA
samples only) demonstrat Association testing of IPS with pre-defined clinical and technical features (N = 1134, spun OA
samples only) demonstrated that IPS was significantly, but modestly, greater in females, greater in
advanced radiographic dise samples only) demonstrated that IPS was significantly, but modestly, greater in females, greater in
advanced radiographic disease (KL grade \geq 3), and was greater in SF samples with visual blood
staining scores \geq 2 (advanced radiographic disease (KL grade \geq 3), and was greater in SF samples with visual blood staining scores \geq 2 (Table 1). We therefore repeated the cluster analysis, using IPS and non-IPS regressed datasets, but advanced radiographic disease (KL grade ≥3), and was greater in SF samples with visual blood
staining scores ≥2 (Table 1). We therefore repeated the cluster analysis, using IPS and non-IPS
regressed datasets, but stratifi stating scores ≥2 (Table 2). We therefore repeated the cluster analysis, analysis is an increase if the clusters repressed datasets, but stratified by biological sex (Figure 1C), radiographic disease severity (Figure 1D) a regressed datasets, but stratified analysis, clusters (Figure 1D) and presence of blood staining (Figure 1E). As with our non-stratified analyses, clusters (again indistinct) were only identified in non-IPS regressed data. Indistinct) were only identified in non-IPS regressed data. Collectively these data suggest that there
are two potential endotypes in the non-IPS corrected data, but they are on a continuum, defined by
the IPS, and are not individual endotypes in the non-IPS corrected data, but they are on a continuum, defined by
the IPS, and are not distinct. Furthermore, the cluster structure is independent of stage of disease,
biological sex and visible b the IPS, and are not distinct. Furthermore, the cluster structure is independent of stage of disease,
biological sex and visible blood staining. the IPS, and are not distinct the IPS, and are not distinct the cluster structure is independent of the cluster structure is in the control of the

Synovial Fluid protein associations with radiographic OA

biological sex and visible blood statining.
Synovial Fluid protein associations with
We next examined which SF proteins we $\frac{1}{2}$ We next examined which SF proteins were associated with radiographic disease severity. Over 1000
proteins were significantly associated with radiographic disease severity in each of the Discovery (N proteins were significantly associated with radiographic distance, in each of the Discovery (New York)
The Discovery (New York)

= 1021, 126.0 proteins replicating across both datasets. Figure 2A shows the Combined dataset where

3815 proteins were associated with radiographic disease severity. Top associated proteins that

replicated (across Discov (24.1%) proteins were associated with radiographic disease severity. Top associated proteins that
replicated (across Discovery and Replication cohorts) and that remained significant in the Combined
dataset after cohort adj 381418 proteins were associated with radiographic disease. Protein abundance profiles for a selection of the labelled proteins were also significantly associated with ordinal KL grade, either significantly decreasing with dataset after cohort adjustment, are labelled in orange. Protein abundance profiles for a selection of
the labelled proteins were also significantly associated with ordinal KL grade, either significantly
decreasing with wo dataset after contready, are labeled in orange. The labelled proteins were also significantly associated with ordinal KL grade, either significantly decreasing with worsening radiographic disease severity (LYVE1, IGFPB-6, the labelled proteins were also significantly assessment with sharen in egalic discussing with worsening radiographic disease severity (LYVE1, IGFPB-6, FGFP1, sFRP-3) or increasing (TSG-6, sTREM-1, Activin A, VEGF121) (Fig increasing (TSG-6, sTREM-1, Activin A, VEGF121) (Figure 2B). Two additional proteins, associated
with OA, MMP-13^[38] and COL2^[39], followed this latter pattern. Using the Hallmark gene set
repository, nine differentia with OA, MMP-13^[38] and COL2^[39], followed this latter pattern. Using the Hallmark gene set
repository, nine differentially expressed pathways were significantly enriched across at least one of
the three datasets (Figu repository, nine differentially expressed pathways were significantly enriched across at least one of
the three datasets (Figure 2C). Of these, "Epithelial Mesenchymal Transition (EMT)", "Complement"
and "Angiogenesis" wer repository, nine and standing angles to pathways were significantly entered with three differential
the three datasets (Figure 2C). Of these, "Epithelial Mesenchymal Transition (EMT)", "Complement"
and "Angiogenesis" were and "Angiogenesis" were significantly associated with advanced radiographic OA across all datasets.
Protein-protein interactions within each of the enriched pathways are shown in Figures 2D-F. "EMT"
contained a number of Protein-protein interactions within each of the enriched pathways are shown in Figures 2D-F. "EMT"
contained a number of molecules previously associated with matrix remodelling in OA^[40] including,
but not limited to, T contained a number of molecules previously associated with matrix remodelling in $OA^{[40]}$ including,
but not limited to, TIMP1, TIMP3, MMP-2, TGF β 1, VEGFA and Fibronectin 1 (FN1). The correlation
between protein associ contained a number of molecules previously associated with matrix remodelling in OA^[40] including,
but not limited to, TIMP1, TIMP3, MMP-2, TGF β 1, VEGFA and Fibronectin 1 (FN1). The correlation
between protein associ but not limited to, TIMP1, TIMP3, MMP-2, TGFβ1, VEGFA and Fibronectin 1 (FN1). The correlation
between protein associations within Discovery and Replication datasets was r = 0.49 (p<2.2 x 10⁻¹⁶)
(Figure 2G).
We also per

between protein associations within Discovery and Replication datasets was r = 0.49 (p<2.2 x 10^{-v})
(Figure 2G).
We also performed similar analyses after correction for cohort (as a random intercept) or after IPS
regressi (Figure 2G).
We also per
regression.
high (r=0.88
disease seve regression. Correlation of corresponding protein effects before and after cohort adjustment was
high (r=0.88, p<2.2 x 10⁻¹⁶)(Supplementary Figure 1A), irrespective of differences in radiographic
disease severity across c high (r=0.88, p<2.2 x 10^{-16})(Supplementary Figure 1A), irrespective of differences in radiographic
disease severity across cohorts (Supplementary Figure 1B). Pathway analysis showed a robust "EMT"
signature across all significantly enriched (Supplementary Figure 1C). For IPS regressed data, the volcano plot of proteins
associated with radiographic disease severity is shown in Supplementary Figure 2A. Correlation of disease severity and the complementary Figure 12). Pathway analysis showing signature across all datasets, although "complement" and "angiogenesis" pathways were no longer significantly enriched (Supplementary Figure 1C). significantly enriched (Supplementary Figure 1C). For IPS regressed data, the volcano plot of proteins
associated with radiographic disease severity is shown in Supplementary Figure 2A. Correlation of
corresponding protein associated with radiographic disease severity is shown in Supplementary Figure 2A. Correlation of
corresponding protein effects was also high (r=0.82, p<2.2 x 10^{-16})(Supplementary Figure 2B) and
pathway associations fo corresponding protein effects was also high (r=0.82, p<2.2 x 10^{-16})(Supplementary Figure 2B) and
pathway associations for "EMT", "complement" and "angiogenesis" remained robust, but also
included "coagulation" (Supplem pathway associations for "EMT", "complement" and "angiogenesis" remained robust, but also
included "coagulation" (Supplementary Figure 2C). Data associated with these analyses can be found
in Supplementary Data files 1 & 2 included "coagulation" (Supplementary Figure 2C). Data associated with these analyses can be found included "coagulation" (Supplementary Figure 2C). Data association with these analyses can be found
in Supplementary Data files 1 & 2.
Synovial Fluid protein associations with advanced radiographic OA after stratification

in Supplementary Data files 1 & 2.
Synovial Fluid protein association
biological sex $\begin{array}{c} \frac{1}{2} \end{array}$ Synovial Fluid protein associations with advanced radiographic OA after stratification by BMI or biological sex

As "Metabolic OA", driven largely by BMI, has been suggested as a potential OA phenotype⁽⁴⁴⁾, we
used STEpUP OA data to examine the proteins associated with radiographic disease severity after
terminal of the proteins as used STEpUP OA data to examine the proteins associated with radiographic disease severity after

stratification by participant BMI (≥30 indicating obesity, N = 587 and <30, N = 649). We first looked
at proteins in the SF that were associated with BMI, irrespective of disease status. Reassuringly, a
number of proteins mumber of proteins known to be associated with BMI, including the appetite suppressing hormone,
leptin (LEP) insulin (INS), growth hormone receptor (GHR) and C-reactive protein (CRP) were
identified (N = 248, 66.9% upregu heptin (LEP) insulin (INS), growth hormone receptor (GHR) and C-reactive protein (CRP) were
identified (N = 248, 66.9% upregulated) (Supplementary Figure 3A; Supplementary Data file 3).
Leptin's SF levels correlated close ldentified (N = 248, 66.9% upregulated) (Supplementary Figure 3A; Supplementary Data file 3).
Leptin's SF levels correlated closely with BMI (r=0.58, p<2.2 x 10⁻¹⁶)(Supplementary Figure 3B) and
associations of obesity-as Leptin's SF levels correlated closely with BMI (r=0.58, p<2.2 x 10⁻¹⁶)(Supplementary Figure 3B) and associations of obesity-associated proteins appeared robust across datasets, and after cohort adjustment (Supplementary associations of obesity-associated proteins appeared robust across datasets, and after cohort
adjustment (Supplementary Figures 3C-E). When stratified by obesity status, over 1800 proteins
were significantly associated wit adjustment (Supplementary Figures 3C-E). When stratified by obesity status, over 1800 proteins
were significantly associated with advanced radiographic OA in each of the obese and non-obese
groups (Figure 3A, B), with a c were significantly associated with advanced radiographic OA in each of the obese and non-obese
groups (Figure 3A, B), with a correlation between the corresponding protein effects in the obese and
non-obese groups of $r = 0$ groups (Figure 3A, B), with a correlation between the corresponding protein effects in the obese and
non-obese groups of $r = 0.72$ (p <2.2 x 10^{-16})(Figure 3C). No significant interaction terms with obesity
status were non-obese groups of $r = 0.72$ ($p < 2.2 \times 10^{-16}$)(Figure 3C). No significant interaction terms with obesity
status were identified by formal interaction testing (at padj <0.05). Interestingly, Hallmark pathway
analysis sh non-obese groups of r = 0.72 (p <2.2 x 10²⁰)(Figure 3C). No significant interaction terms with obesity
status were identified by formal interaction testing (at padj <0.05). Interestingly, Hallmark pathway
analysis showed standly is showed a strong consistent "EMT" pathway signature in both groups, but only samples from
obese participants retained significant associations with "coagulation" and "complement" (Figure
3D) (Supplementary Data f analysis showed a strong consistent "EMT" pathway signature in both groups, but only samples from
obese participants retained significant associations with "coagulation" and "complement" (Figure
3D) (Supplementary Data fil

3D) (Supplementary Data file 4).
To explore the influence of other participant factors on radiographic disease-protein associations,
we also stratified samples by biological sex (Figure 4A, B). Protein associations with ra To explore the influence of other
we also stratified samples by bidisease severity, after stratifica
 $(r=0.69, p < 2.2 \times 10^{-16}$, Figure 40 To explore the influence of other participant factors on radiagnipate instact positions with radiographic disease-severity, after stratification by biological sex, also had a strong cross-strata correlation (r=0.69, p <2. disease severity, after stratification by biological sex, also had a strong cross-strata correlation
(r=0.69, p <2.2 x 10¹⁶, Figure 4C), with 1437 significantly associated proteins common to the two
groups. No significa ($r=0.69$, $p < 2.2 \times 10^{-16}$, Figure 4C), with 1437 significantly associated proteins common to the two groups. No significant interaction terms with biological sex were identified by formal interaction testing (at padj < groups. No significant interaction terms with biological sex were identified by formal interaction
testing (at padj <0.05). Hallmark pathway analysis also showed a strong "EMT" pathway signature in
both sexes, but only mal testing (at padj <0.05). Hallmark pathway analysis also showed a strong "EMT" pathway signature in both sexes, but only males showed significant associations with "angiogenesis" and "coagulation" both sexes, but only males showed significant associations and the coagulations.
(Figure 4D) (Supplementary Data file 5).
Synovial Fluid protein associations with WOMAC pain in OA

Synovial Fluid protein associations with WOMAC pain in OA

Finally, we explored the association of SF proteins with patient reported pain. We identified 797 SF
proteins that were significantly associated with WOMAC knee pain in the Combined non-IPS י)
ו
ו proteins that were significantly associated with WOMAC knee pain in the Combined non-IPS
regressed dataset. However, none of these proteins replicated across Discovery and Replication
datasets and the cross-dataset correl proteins that were significantly associated with weak (mesos piscovery and Replication
datasets and the cross-dataset correlation was weak (r=0.36, p <2.2 x 10⁻¹⁶)(Figure 5A, B). Noelin-2
(NOE2) and ecto-ADP-ribosyltran datasets and the cross-dataset correlation was weak (r=0.36, p <2.2 x 10^{-16})(Figure 5A, B). Noelin-2 (NOE2) and ecto-ADP-ribosyltransferase 3 (NAR3) were the only significantly associated proteins in the Combined datas datasets and the cross-dataset correlation was weak (r=0.36, p <2.2 x 10^{-c})(Figure 5A, B). Noelin-2
(NOE2) and ecto-ADP-ribosyltransferase 3 (NAR3) were the only significantly associated proteins in
the Combined dataset (NOE2) and the Combined dataset after cohort adjustment (Supplementary Figure 4A and labelled green in Figure 5A). The relationships between NOE2 and NAR3 protein abundance with WOMAC pain subscores are shown in Figure 5C Figure 5A). The relationships between NOE2 and NAR3 protein abundance with WOMAC pain subscores are shown in Figure 5C (Pearson correlation). The pathway analysis did not identify subscores are shown in Figure 5C (Pearson correlation). The pathway analysis did not identify consistent associations across Discovery, Replication and Combined datasets (Figure 5D) and no
significant pathways were identified within the Discovery dataset alone (at padj <0.05). Lack of
replication may have been infl signification may have been influenced by unevenly distributed knee pain subscores across Discovery
and Replication cohorts (Supplementary Figure 4B). The number of proteins associated with pain
was also reduced in the Com replication cohorts (Supplementary Figure 4B). The number of proteins associated with pain
was also reduced in the Combined dataset after adjustment for radiographic disease severity
(Supplementary Figure 4C). NOE2 and NAR and Replication content (Supplementary) Replication constrained to proteins attenuate and paint
was also reduced in the Combined dataset after adjustment for radiographic disease severity
(Supplementary Figure 4C). NOE2 an (Supplementary Figure 4C). NOE2 and NAR3 remained significantly associated with WOMAC pain
after adjustment, and their levels were not independently associated with radiographic grade (by
ordinal regression) (Supplementar after adjustment, and their levels were not independently associated with radiographic grade (by ordinal regression) (Supplementary Figure 4D). The correlation between pain-associated protein effects from non-IPS and IPS and independent and the correlation between pain-associated protein
effects from non-IPS and IPS regressed analyses using the Combined datasets was r=0.97 (p <2.2 x
 10^{-16}) (Supplementary Figure 4E, Supplementary Data f effects from non-IPS and IPS regressed analyses using the Combined datasets was r=0.97 (p <2.2 x 10⁻¹⁶) (Supplementary Figure 4E, Supplementary Data files 6 & 7). Further analyses on patient reported pain e.g. following effects from and 10⁻¹⁶) (Supplementary Figure 4E, Supplementary Data files 6 & 7). Further analyses on patient
reported pain e.g. following stratification were not performed.
Discussion 10⁻⁰) (Supplementary Figure 4E, Supplementary Data files 6 & 7). Further analyses on patient
reported pain e.g. following stratification were not performed.
Discussion

Discussion

reported pain e.g. following stratification were not performed.

Discussion

In this manuscript we describe the primary results of STEpUP OA, the largest unbiased, replicated,

cross-sectional synovial fluid proteomics ana |
|
| In the cross-sectional synovial fluid proteomics analysis in knee OA ever performed. We uncover the balance of biological pathways in disease and how they change with structural and symptomatic disease severity. This datas balance of biological pathways in disease and how they change with structural and symptomatic
disease severity. This dataset provides an unprecedented data resource from which to interrogate
OA biology, address specific mo balance of biological pathways in disease and how they change multipled and opaque disease severity. This dataset provides an unprecedented data resource from which to interrogate OA biology, address specific molecular que disease specific molecular questions and consider the influence of important patient-
related factors, such as BMI and biological sex.
The data presented here do not reveal evidence for distinct molecular endotypes in knee

were identified by cluster analysis, which were defined by the IPS gradient. We still do not fully The data presented here do not reveal evidently and a presented here do not reveal evidently radiographic distribution when considering early radiographic distribution when considering early radiographic distribution when The data presented here de instituted a business in additional endotyped when considering early radiographic disease separately. Rather, two continuous endotypes were identified by cluster analysis, which were defined by t ever identified by cluster analysis, which were defined by the IPS gradient. We still do not fully
understand the importance or origin of intracellular protein in spun SF. Importantly, correcting for
this signal using the understand the importance or origin of intracellular protein in spun SF. Importantly, correcting for
this signal using the IPS did not substantially change proteins or pathways associated with clinical
features, suggesting understand the importance or origin of intrictional proteins or pathways associated with clinical
features, suggesting that it is a minor influence on clinically relevant OA biology. It is therefore
possible that the IPS-d features, suggesting that it is a minor influence on clinically relevant OA biology. It is therefore
possible that the IPS-driven clustering is due to technical confounding during sample collection and
processing. Taken to possible that the IPS-driven clustering is due to technical confounding during sample collection and
processing. Taken together, the results support OA being a single heterogenous disease rather than
multiple conditions ea clusters have been described in the transcriptome of OA cartilage and synovium^[42-45], in SF using multiple conditions each driven by a distinct pathway. This may appear at odds with studies
suggesting discernible molecular clusters in tissues from participants with OA. Indeed, patient
clusters have been described in th suggesting discernible molecular clusters in tissues from participants with OA. Indeed, patient clusters have been described in the transcriptome of OA cartilage and synovium^[42-45], in SF using mass spectrometry^{[46, 4} clusters have been described in the transcriptome of OA cartilage and synovium^[42-45], in SF using
mass spectrometry^[46, 47], and in plasma^[17, 19, 21]. However, these studies are smaller than STEpUP OA,
and only a f clusters have been described in the transcriptome of OA cartilage and synovium¹⁴²⁻⁴⁵¹, in SF using
mass spectrometry^[46, 47], and in plasma^[17, 19, 21]. However, these studies are smaller than STEpUP OA,
and only a f mass spectrometry(",",",", and in plasma^{127, 23, 22}). However, these studies are smaller than STEpUP OA,
and only a few included replication. Some of the studies examined prospective outcomes associated
with clusters, ra and only a few included replication. Some of the statute manning prospective outcomes associated with clusters, rather than the cross-sectional analysis that we present here. with clusters, rather than the cross-sectional analysis than the present here.
The cross-sectional analysis than the cross-sectional analysis that we present here.

Synovial fluid is an ultrafiltrate of the plasma but also reflects joint-specific processes such as active
secretion from cells^[48], including in extracellular vesicles, release from damaged or short-lived cells,
and she secretion from cells⁽⁴⁸⁾, including in extracellular vesicles, release from damaged or short-lived cells, and shedding from cell surfaces. Pathway analysis of knee OA SF proteins associated with radiographic disease seve radiographic disease severity indicates a robust activation of "EMT", indicative of tissue remodelling,
presumably part of the joint tissue injury response^[49]. The "EMT" signature was consistent across all
groups, irres radiographic disease severity, indicated a robust activation of the "CMT" signature was consistent across all
groups, irrespective of stratification and correction by cohort or IPS, or factors such as BMI and sex,
suggesti groups, irrespective of stratification and correction by cohort or IPS, or factors such as BMI and sex,
suggesting that this is the common pathway in OA pathogenesis. Activation of complement, groups, in the transmitted and transmitted by the correction by the transmitted and and complement, coagulation and angiogenesis was also evident, although was variable across subgroups. Whether these protein signatures id suggestion and angiogenesis was also evident, although was variable across subgroups. Whether
these protein signatures identify groups of patients who display distinct treatment responses
remains to be seen.
Replication ac these protein signatures identify groups of patients who display distinct treatment responses
remains to be seen.
Replication across Discovery and Replication cohorts was robust for associations with structural

disease but less so for pain. Patient reported outcome measures, such as knee pain, are known to be
influenced by external factors beyond molecular drivers made by the joint e.g. psychological factors Replication across I
disease but less so fo
influenced by exterr
^[50], making cross-se disease but less so for pain. Patient reported outcome measures, such as knee pain, are known to be
influenced by external factors beyond molecular drivers made by the joint e.g. psychological factors
^[50], making crossinfluenced by external factors beyond molecular drivers made by the joint e.g. psychological factors
[50], making cross-sectional analyses of this sort challenging. Such extra-articular factors are complex
and were not con ^[50], making cross-sectional analyses of this sort challenging. Such extra-articular factors are complex
and were not consistently collected within STEpUP OA cohorts. Protein associations with pain may
also have been li [50], making cross-sectional analyses of this sort challenging. Such extra-articular factors are complex
and were not consistently collected within STEpUP OA cohorts. Protein associations with pain may
also have been limit also have been limited by the fact that WOMAC pain scores were only available on a subset within
STEpUP OA (N = 805) and most of these were within a relatively narrow range of pain severity.
Despite this being the largest

STEpUP OA (N = 805) and most of these were within a relatively narrow range of pain severity.
Despite this being the largest analysis of its kind in OA, we recognise a number of limitations: protein
detection using the Som STEPUP OF (N = 805) and most of these were with a relatively narrow range of pain severity.
Despite this being the largest analysis of its kind in OA, we recognise a number of limitations: p
detection using the SomaScan pl detection using the SomaScan platform, rather than mass spectrometry, is biased towards detection
of full-length proteins, thus potentially missing fragments of proteins that could be biologically
informative; our samples of full-length proteins, thus potentially missing fragments of proteins that could be biologically
informative; our samples were generated from a diverse set of, largely, pre-existing cohorts and
adjustment for cohort did of fundative; our samples were generated from a diverse set of, largely, pre-existing cohorts and adjustment for cohort did reduce the number of significantly associated proteins; finally, by only focusing on proteins foun informative; our samples were generated from a largely set of, largely, pre-enoung senters and
adjustment for cohort did reduce the number of significantly associated proteins; finally, by only
focusing on proteins found i adjustment for contribute the number of significantly, accessive proteins, many, by only
focusing on proteins found in the synovial fluid, it is possible that key disease molecules or pathways
were unintentionally excluded

Fourier unintentionally excluded.
The cross-sectional analysis presented in this manuscript provides strong proof of concept that knee
OA synovial fluid, provides an informative window into disease-relevant biology. Future were unintentionally entertainty
The cross-sectional analysis pre
OA synovial fluid provides an i
STEpUP OA are now planned to
whether they are driven in part The cross-sectional fluid provides an informative window into disease-relevant biology. Future studies in
STEpUP OA are now planned to ask whether SF signatures predict prospective clinical outcomes and
whether they are dr OF STEPUP OA are now planned to ask whether SF signatures predict prospective clinical outcomes and
Whether they are driven in part by genetic variants associated with OA risk. Ultimately, we hope that
SF analyses of this STEPUP OF A are now planned to anti-modular or againstic prospective clinical outcomes and
Whether they are driven in part by genetic variants associated with OA risk. Ultimately, we hope that
SF analyses of this sort will SF analyses of this sort will assist in experimental medicine studies to test treatment responsiveness,
helping to de-risk subsequent clinical trials of new interventions. The publication of this manuscript
also marks the SE also marks the opportunity to welcome external parties to apply for access to STEpUP OA data for
research purposes in accordance with our Consortium Agreement. helping to de-risk subsequent clinical trials strict interventions the publication of new internations.
also marks the opportunity to welcome external parties to apply for access to STEpUP OA data for
research purposes in research purposes in accordance with our Consortium Agreement.

The original parties to apply for a property of a data for a data for any formula for any state for an research purposes in accordance with our Consortium Agreement.

Funding Statement:
The study was supported by Kennedy Trust for Rheumatology Research (grant number: 171806),
Versus Arthritis (grant number: 22473), Centre for Osteoarthritis Pathogenesis Versus Arthritis (grant
numbers: The study was supported by the study of the study derivants of the study of the study dersus Arthritis (grant
numbers: 21621, 20205), Galapagos, Biosplice, Novartis, Fidia, UCB, Pfizer (non-consortium member)
and Somalogic Final Arthritis (grammation 2211), Samiltic Correstational Cattigation Correstational numbers: 21621, 20205), Galapagos, Biosplice, Novartis, Fidia, UCB, Pfizer (non-consortium member)
and Somalogic (in kind contributions) numbers: 21212, 21211, 2014, 2021, 2021, 1021, 1021, 1012, 1012, 1012, 1012, 1011, 1011, 1011, 1011, 1011, 101
and Somalogic (in kind contributions). The funders Kennedy Trust for Rheumatology Research,
Versus Arthritis an and Somalogic (in hand contributions). The funders fields, there is incomparating, therein,
Versus Arthritis and Pfizer had no role in the study design, data collection and analysis, decision to
publish or preparation of t Protect Minima and Phila manufolial multiplem and paragoty and permetic in the analysis) dentity to
publish or preparation of the manuscript. The funders Galapagos, Biosplice, Novartis, Fidia, UCB and
SomaLogic were all ac SomaLogic were all active consortium members, attending consortium meetings. As such they made
contributions to the study design and support of data collection, decision to publish and review and
commenting on the manuscri Somalism to the study design and support of data collection, decision to publish and review and
Commenting on the manuscript. In addition, SomaLogic, UCB and Novartis were members of the
Data Analysis Group.
Additional rel commenting on the manuscript. In addition, SomaLogic, UCB and Novartis were members of the
Data Analysis Group.
Additional relevant funding sources: LJD is supported by a Wellcome Trust fellowship grant
208750/Z/17/Z and K

Commenting on the manuscript. In addition, Somalogicy 202 and Novartian Novartism in the manuscript.
Data Analysis Group.
208750/Z/17/Z and Kennedy Trust for Rheumatology Research for the present manuscript. FEW was
direct Mathematyce Strates
Additional relevant
208750/Z/17/Z and K
directly supported
(MR/S016538/1;MR/ Additional Preliant Protein and the Propport and the Preliant Proteining grant
208750/Z/17/Z and Kennedy Trust for Rheumatology Research for the present manuscript. FEW was
directly supported in this work by her UKRI Futur directly supported in this work by her UKRI Future Leaders Fellowship and its renewal
(MR/S016538/1;MR/S016538/2; MR/Y003470/1). FW, NKA and SK are members of the Centre for
Sport, Exercise and Osteoarthritis Research Vers (MR/S016538/1;MR/S016538/2; MR/Y003470/1). FW, NKA and SK are members of the Centre for
Sport, Exercise and Osteoarthritis Research Versus Arthritis (grant number 21595). MK is supported
by grants from CIHR, NSERC, The Art (MR/S016538/1;MR/S016538/2; MR/Y003470/1). FW, NKA and SK are members of the Centre for Sport, Exercise and Osteocration of the Community (gram number 2119), minister experience
by grants from CIHR, NSERC, The Arthritis Society Canada, Krembil Foundation, CFI, Canada Research
Chairs program, and has received Chairs program, and has received support from the University Health Network Foundation, Toronto
for the present manuscript. TJW is supported by grants from NWO-TTW Perspectief (#P15-23),
Stichting de Weijerhorst and ReumaN For the present manuscript. TJW is supported by grants from NWO-TTW Perspectief (#P15-23),
Stichting de Weijerhorst and ReumaNederland (LLP14) for the present manuscript. CTA is supported
by the Canadian Institutes of Heal For the present manuscript. The reppetition of grants from the term integrate (these say,
Stichting de Weijerhorst and ReumaNederland (LLP14) for the present manuscript. CTA is supported
by the Canadian Institutes of Healt by the Canadian Institutes of Health Research, Western University Bone and Joint Institute, and the
Academic Medical Organization of Southwestern Ontario for the present manuscript. BDMT is
supported through the United Kin Academic Medical Organization of Southwestern Ontario for the present manuscript. BDMT is
supported through the United Kingdom Medical Research Council programme (grant MC UU
00002/2) and theme (grant MC_UU_00040/02 – Prec Academic Medical Organization of Southwestern Studies (or the present manuscript Studies)
Supported through the United Kingdom Medical Research Council programme (grant MC UU
00002/2) and theme (grant MC UU_00040/02 – Prec $00002/2$ and theme (grant MC UU 00040/02 – Precision Medicine) funding. LB is supported by grants from Kennedy Trust for Rheumatology Research (grant number 171806) and UK Medical
Research Council (grant MC UU 00002/2). This work was supported by the NIHR Oxford Biomedical
Research Centre (BRC) and the NIHR Nott Research Council (grant MC UU 00002/2). This work was supported by the NIHR Oxford Biomedical Research Centre (BRC) and the NIHR Nottingham BRC. The views expressed are those of the authors
and not necessarily those of the NHS, the NIHR or the Department of Health.
Competing Interest Statement: and not necessarily those of the NHS, the NIHR or the Department of Health.
 Competing Interest Statement:

TAP, YD, PH, SL, AS, NKA, AJP, DF, MK, BM, AMV and SK declare no conflicts of interest. FW has

and noting the necessary of the Neumannian Competing Interest Statement:
TAP, YD, PH, SL, AS, NKA, AJP, DF, MK, BM, AMV and SK declare no confli
Teceived consultancy fees from Pfizer. LSL has received consultancy fees fro $\frac{1}{\pi}$ Competing interest Statement:
TAP, YD, PH, SL, AS, NKA, AJP,
received consultancy fees from
AB, and is an advisory board
Nightingale Health PLC. TLV ha received consultancy fees from Pfizer. LSL has received consultancy fees from Arthro Therapeutics
AB, and is an advisory board member of AstraZeneca. LJD has received consultancy fees from RB, and is an advisory board member of AstraZeneca. LID has received consultancy fees from Nightingale Health PLC. TLV has no conflicts to declare with the exception of grant income for $\frac{1}{2}$ Nightingale Health PLC. TLV has no conflicts to declare with the exception of grant income for
Nightingale Health PLC. TLV has no conflicts to declare with the exception of grant income for Nightingale Health PLC. TLV has no conflicts to declare with the exception of grant income for STEpUP OA from industry partners (see above). RAM is a shareholder of AstraZeneca. SB and JM are
employees and shareholders of Novartis. CTA has received consultancy fees from Novartis, and has
received honoraria for educa received honoraria for educational purposes also from Novartis. TJW is a shareholder of
Chondropeptix BV. DAW has received consultancy fees from GlaxoSmithKline plc, AKL Research &
Development Limited, Pfizer Ltd, Eli Lill Chondropeptix BV. DAW has received consultancy fees from GlaxoSmithKline plc, AKL Research &
Development Limited, Pfizer Ltd, Eli Lilly and Company, Contura International, and AbbVie Inc, has
received honoraria for educati Chommark Pepting BV. Data Theories Consultancy, 1999 with Charleman and AbbVie Inc, has received honoraria for educational purposes from Pfizer Ltd and AbbVie Inc, is a board member (Director) of UKRI and Versus Arthritis Development Limited, Pitter Lid, Eli Lin, And Company, Contain International, and Pittern any more
received honoraria for educational purposes from Pfizer Ltd and AbbVie Inc, is a board member
(Director) of UKRI and Versus received honoranceive substantine purposes from Piazer 212 and Piazers fixer as included
(Director) of UKRI and Versus Arthritis Advanced Pain Discovery Platform.
Author contributions:
Conception and Design: TLV, FEW, LJD,

()
Author contributions:
Conception and Design: TLV, FEW, LJD, PH, RAM, DP, SL, SB, LSL, AS, CTA,
AMV. Analysis and interpretation of data: TAP, YD, LJD, FEW, TLV, PH, |
|
| Author commutions:
Conception and Desigr
AMV. Analysis and int
Drafting Article: TLV, CONCEPTION AND DESIGN: TETY, TETY, TETY, TRY, ARMY, DETY, DESITION, DETY, TETY, ANY, TETY, THY, DETY, ANY, ANY
AMV. Analysis and interpretation of data: TAP, YD, LJD, FEW, TLV, PH, RAM, JM, SB, BDMT, LB.
Drafting Article: Drafting Article: TLV, TAP, YD, LJD, FEW. Critical revision of article: all authors. Final Approval: all
authors.
Acknowledgements:

We would like to express our gratitude and thanks to all cohort participants who contributed aamsee
Acknowl
We wou
samples |
|
| Acknowledgements:
We would like to ex
samples to STEpUP O
(Utrecht Medical Cer
thank Gretchen Brew Samples to STEpUP OA. We are grateful for the support from Floris Lafeber and Simon Mastbergen
(Utrecht Medical Centre) for provision of samples. We thank the Oxford Knee Surgery Team. We
thank Gretchen Brewer for her admi (Utrecht Medical Centre) for provision of samples. We thank the Oxford Knee Surgery Team. We thank Gretchen Brewer for her administrative support of the consortium.

(Centric Medical Centre) for provision of samples. We thank the Omsortium.

The STEpUP OA Consortium author block includes: University of Nottingham: Ana M. Valdes, David thank Gretchen Brewer for her administrative support of the consentants
The STEpUP OA Consortium author block includes: University of Notting
A. Walsh, Michael Doherty, Vasileios Georgopoulos; Lund University: $\frac{1}{2}$ The Stephen Consortium and Consortium and Consortium and the University of A. Walsh, Michael Doherty, Vasileios Georgopoulos; Lund University: Staffan Larsson, L. Stefan
Lohmander, André Struglics; University of Cambridge: Lohmander, André Struglics; University of Cambridge: Brian D.M. Tom, Laura Bondi; University of
Toronto: Mohit Kapoor, Rajiv Gandhi, Anthony Perruccio, Y. Raja Rampersaud, Kim Perry; University
of Manchester: Tim Hardingha Loronto: Mohit Kapoor, Rajiv Gandhi, Anthony Perruccio, Y. Raja Rampersaud, Kim Perry; University
Of Manchester: Tim Hardingham, David Felson; University of Oxford: Tonia L. Vincent, Thomas A.
Perry, Luke Jostins-Dean, Yun Toronto: Manchester: Tim Hardingham, David Felson; University of Oxford: Tonia L. Vincent, Thomas A.
Perry, Luke Jostins-Dean, Yun Deng, Vicky Batchelor, Jennifer Mackay-Alderson, Gretchen Brewer,
Rose M. Maciewicz, Brian offerry, Luke Jostins-Dean, Yun Deng, Vicky Batchelor, Jennifer Mackay-Alderson, Gretchen Brewer,
Rose M. Maciewicz, Brian Marsden, Nigel K. Arden, Philippa Hulley, Andrew Price, Stefan Kluzek,
Megan Goff, Vinod Kumar, Jam Perry, Land Frame Dram, Tan Drag, 1999, Jostins-Deanner, Framer Machin, Tharrists, Dremen Brewer,
Rose M. Maciewicz, Brian Marsden, Nigel K. Arden, Philippa Hulley, Andrew Price, Stefan Kluzek,
Megan Goff, Vinod Kumar, Jam Megan Goff, Vinod Kumar, James Tey, Tamas Szommer; Imperial College London: Fiona E. Watt,
Andrew Williams, Artemis Papadaki; University College Maastricht: Tim J. Welting, Pieter
Emans, Tim Boymans, Liesbeth Jutten, Marjo Megan Gor, Timod Milliams, Artemis Papadaki; University College Maastricht: Tim J. Welting, Pieter
Emans, Tim Boymans, Liesbeth Jutten, Marjolein Caron, Guus van den Akker; University of Western
Ontario: C. Thomas Appleton Emans, Tim Boymans, Liesbeth Jutten, Marjolein Caron, Guus van den Akker; University of Western
Ontario: C. Thomas Appleton, Trevor B. Birmingham, J. Daniel Klapak; Biosplice: Sarah Kennedy, Emans, Tim Boymans, Eccession Pinters, Marystein Pinters, Pintersity of Western, Protection,
Ontario: C. Thomas Appleton, Trevor B. Birmingham, J. Daniel Klapak; Biosplice: Sarah Kennedy,
Jeymi Tambiah; Fidia: Devis Galess Jeymi Tambiah; Fidia: Devis Galesso, Nicola Giordan; SomaLogic: Joe Gogain, Darryl Perry, Anna
Mitchel, Ela Zepko; Novartis: Sophie Brachat, Joanna Mitchelmore, Juerg Gasser, Lori Jennings; UCB:
Waqar Ali Mitchel, Ela Zepko; Novartis: Sophie Brachat, Joanna Mitchelmore, Juerg Gasser, Lori Jennings; UCB:
Waqar Ali. Maqar Ali.
Maqar Ali.
. Waqar Ali.

Data Access:
The minimal datasets upon which this data relies and all R code, including the html vignette, are
available at <u>https://github.com/ndorms-tperry/STEpUP-OA-Primary-Manuscript</u>. The full STEpUP OA
dataset may be available at <u>https://github.com/ndorms-tperry/STEpUP-OA-Primary-Manuscript</u>. The full STEpUP OA
dataset may be made available by application to the Data Access and Publication Group of STEpUP
OA (stepupoa@kennedy.ox.ac.uk ataset may be made available by application to the Data Access and Publication Group of STEpUP
OA (stepupoa@kennedy.ox.ac.uk) once the primary analysis manuscript is published, in accordance
with what is stipulated in our dataset may be made a manufacture of application to the Data Access manuscript is published, in accordance
OA (stepupoa@kennedy.ox.ac.uk) once the primary analysis manuscript is published, in accordance
with what is stipul of the published in our Consortium Agreement. This may attract an access fee to cover
administrative processing. Neither the minimal dataset nor the full STEpUP OA dataset include
patient identifiable data. administrative processing. Neither the minimal dataset nor the full STEpUP OA dataset include
patient identifiable data.
Supplementary Data: patient identifiable data.
Supplementary Data:
Supplementary files include:

r
Supplementary Data:
Supplementary files inclu
Supplementary data files Supplementary Data:
Supplementary files in
Supplementary data fi
<mark>Patient and Public Inv</mark>

Supplementary data files 1-7
Patient and Public Involvement Statement:

People with lived experience of osteoarthritis have been involved in the design of this project. A
patient research panel was involved in discussing and inputting on the STEpUP OA project in |
|
|
| People with lived experience of osteoarthripatient research panel was involved in diffeoruary 2020 (invited to the Centre for Osteoarthripation of its involvement activities). Aspects report of its involvement activities). patient research panel was involved in discussing and inputting on the STEpUP OA project in
February 2020 (invited to the Centre for Osteoarthritis Pathogenesis Versus Arthritis in Oxford, as
part of its involvement activi patient research patient rise interested in discussing and inputing on the STEpUP OC project in
February 2020 (invited to the Centre for Osteoarthritis Pathogenesis Versus Arthritis in Oxford, as
part of its involvement ac part of its involvement activities). Aspects relevant to the development of the project were further
discussed with the panel in July 2022. The working groups for the consortium included one focused
on patient involvement discussed with the panel in July 2022. The working groups for the consortium included one focused
on patient involvement and engagement. A lay summary is included in the appendix of our publicly
available analysis plan. A on patient involvement and engagement. A lay summary is included in the appendix of our publicly
available analysis plan. A short video about the project was produced and is available on our
website: https://www.kennedy.ox available analysis plan. A short video about the project was produced and is available on our website: https://www.kennedy.ox.ac.uk/oacentre/stepup-oa/stepup-oa. In addition, the various constituent cohorts contributing to website: https://www.kennedy.ox.ac.uk/oacentre/stepup-oa/stepup-oa. In addition, the various
constituent cohorts contributing to STEpUP OA also typically have lay or patient members on their
steering committees. website: https://www.kennedy.com/stepup-oa.uk/oachtre/stepup-oa.uk/oachtre/stepup-oa.uk/oachtre/stepup-oa.uk/o
constituent cohorts contributing to STEpUP-OA also typically have lay or patient members on their
steering comm constitution contribution cohorts contributing to STEPUP OA also typically have lay or patient members on their
steering committees. steering committees.

 \overline{a} References:

- $\frac{1}{2}$ $\frac{1}{2}$
- 1. Neo_l
2013
2. Safir
syste
p. 8. 1. Neogi, T., The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage,

2013. 21(9): p. 1145-53.

2. Safiri, S., et al., *Global, regional and national burden of osteoarthritis 1990-2017: a*
 syste 2013. 21(9): p. 1145-93.
Safiri, S., et al., *Globa
systematic analysis of th*
p. 819-828.
Morgan, O.J., et al., *Oste*
Hospital Episode Statistic
Swain, S., et al., *Trends* i
- Morgan, O.J., et al., Osteoarthritis in England: Incidence Trends From National Health Service
- systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis, 2020. **79**(6):
p. 819-828.
Morgan, O.J., et al., Osteoarthritis in England: Incidence Trends From National Health Service
Hospital Episode Stat Morgan, O.J
Hospital Epi
Swain, S., et
findings fro
2020. **28**(6):
Karsdal, M.J. 3. Morgan, O.J., et al., Osteoarthritis in England: Incidence Trends From National Health Service

Hospital Episode Statistics. ACR Open Rheumatol, 2019. 1(8): p. 493-498.

3. Swain, S., et al., Trends in incidence and pre Hospital Episode Statistics. ACR Open Rheumator, 2015. 1(6): p. 453-498.
Swain, S., et al., Trends in incidence and prevalence of osteoarthritis in the
findings from the Clinical Practice Research Datalink (CPRD). Osteoart
- 4. Swain, 5., et al., Trends in incidence and prevalence of osteoarthritis in the Olinea Kingdom:

findings from the Clinical Practice Research Datalink (CPRD). Osteoarthritis and Cartilage,

2020. **28**(6): p. 792-801.

5. finality from the Clinical Practice Research Datalink (CPRD). Osteoarthritis and Cartilage,
2020. 28(6): p. 792-801.
Karsdal, M.A., et al., *Disease-modifying treatments for osteoarthritis (DMOADs) of the knee*
and hip: le
- 2020. 28(0): p. 752-801.
Karsdal, M.A., et al., *Dis
and hip: lessons learne*
Cartilage, 2016. 24(12): p
Oo, W.M., et al., *The*
(DMOADs): The Evidence
2945. S. Karsdal, M.A., et al., Disease-modifying treatments for osteodirmitis (DMOADs) of the knee
and hip: lessons learned from failures and opportunities for the future. Osteoarthritis
Cartilage, 2016. **24**(12): p. 2013-2021. and mp. ressons reamed from failures and opportunities for the future. Osteoarthritis
Cartilage, 2016. 24(12): p. 2013-2021.
Oo, W.M., et al., The Development of Disease-Modifying Therapies for Osteoarthritis
(DMOADs): The Cartilage, 2010. 24(12): p. 2013-2021.

Oo, W.M., et al., *The Developmeni*

(DMOADs): *The Evidence to Date. Drug*

2945.

Makarczyk, M.J., et al., *Current Model.*

Drugs. Tissue Eng Part C Methods, 202

Cope, P.J., et a 6. Oo, W.M., et al., *The Development of Disease-Modifying Therapies for Osteoarthritis*
(DMOADs): The Evidence to Date. Drug Design Development and Therapy, 2021. **15**: p. 2921-
2945.
Makarczyk, M.J., et al., *Current Mod*
-
- (DMOADS). The Evidence to Date. Drug Design Development and Therapy, 2021. 15: p. 2921-
2945.
Makarczyk, M.J., et al., Current Models for Development of Disease-Modifying Osteoarthritis
Drugs. Tissue Eng Part C Methods, 20 Makar
Drugs.
Cope,
Osteo:
Devez.
Rheun 7. Makarczyk, M.J., et al., Carrent Models for Deveropment of Disease-Modifying Osteoarthritis

Drugs. Tissue Eng Part C Methods, 2021. 27(2): p. 124-138.

8. Cope, P.J., et al., *Models of osteoarthritis: the good, the ba Drugs.* Tissue Eng Part C Methods, 2021. 27(2): p. 124-138.
Cope, P.J., et al., *Models of osteoarthritis: the good*
Osteoarthritis Cartilage, 2019. 27(2): p. 230-239.
Deveza, L.A. and R.F. Loeser, *Is osteoarthritis one*
- 8. Cope, P.J., et al., Models of osteoarthritis. The good, the bad and the promising.

0steoarthritis Cartilage, 2019. 27(2): p. 230-239.

9. Deveza, L.A. and R.F. Loeser, *Is osteoarthritis one disease or a collection of*
-
- Osteoarthritis Cartilage, 2019. 27(2): p. 250-239.
Deveza, L.A. and R.F. Loeser, *Is osteoarthriti*
Rheumatology, 2018. 57: p. 34-42.
Hunter, D.J., *Pharmacologic therapy for osteo*
Nature Reviews Rheumatology, 2011. 7(1): 9. Beveza, L.A. and R.F. Loeser, Is osteomining one disease or a conection of many?

Rheumatology, 2018. 57: p. 34-42.

10. Hunter, D.J., *Pharmacologic therapy for osteoarthritis-the era of disease modification*.

Nature Rheumatology, 2018. 57: p. 34-42.
Hunter, D.J., *Pharmacologic ther*
Nature Reviews Rheumatology, 201
Mobasheri, A., et al., *The future
throughput omics technologies actaxonomy of the disease?* Osteoart
Mobasheri, A., et 10. Hunter, D.J., Pharmacologic therapy for osteoarthmis-the era of disease modification.

Nature Reviews Rheumatology, 2011. 7(1): p. 13-22.

11. Mobasheri, A., et al., The future of deep phenotyping in osteoarthritis: Ho Nature Reviews Rheumatology, 2011. $Y(1)$: p. 13-22.
Mobasheri, A., et al., *The future of deep phenot*
throughput omics technologies advance our under-
taxonomy of the disease? Osteoarthritis and Cartilage
Mobasheri, A., throughput omics technologies advance our understanding of the cellular and molecular
taxonomy of the disease? Osteoarthritis and Cartilage Open, 2021. **3**(4): p. 100144.
-
- 11. Mobasheri, A., et al., The future of deep phenotyping in osteoarthmis. Trow can ingit
throughput omics technologies advance our understanding of the cellular and molecular
taxonomy of the disease? Osteoarthritis and Ca taxonomy of the disease? Osteoarthritis and cartilage Open, 2021: 3(4): p. 100144.
Mobasheri, A., et al., Recent advances in understanding the phenotypes of oste
F1000Res, 2019. 8.
Mobasheri, A., et al., Molecular taxonomy 12. Mobasheri, A., et al., Recent advances in understanding the phenotypes of osteoarthmis.

13. Mobasheri, A., et al., Molecular taxonomy of osteoarthritis for patient stratification, disease

13. Mobasheri, A., et al., M management and drug development: biochemical markers associated with emerging clinical
phenotypes and molecular endotypes. Curr Opin Rheumatol, 2019. **31**(1): p. 80-89.
- F1000Res, 2019. 8.
Mobasheri, A., et a
management and a
phenotypes and mo
Deveza, L.A., A.E. I
future implications.
72. 13. Mobasheri, A., et al., Molecular taxonomy of osteoartmins for patient stratification, alsease

management and drug development: biochemical markers associated with emerging clinical

phenotypes and molecular endotypes. phenotypes and molecular endotypes. Curr Opin Rheumatol, 2015. 31(1): p. 80-85.
Deveza, L.A., A.E. Nelson, and R.F. Loeser, *Phenotypes of osteoarthritis: current*
future implications. Clinical and experimental rheumatolog 14. Deveza, L.A., A.E. Nelson, and R.T. Loeser, Phenotypes of osteoarthmis. current state and
future implications. Clinical and experimental rheumatology, 2019. **37 Suppl 120**(5): p. 64-
72.
Attur, M., et al., *Prognostic*
- future *implications*. Clinical and experimental rheamatology, 2019. 37 Suppl 120(5): p. 64-
72.
Attur, M., et al., *Prognostic biomarkers in osteoarthritis*. Curr Opin Rheumatol, 2013. **25**(1):
p. 136-44. Attı
p. 1 15. Attur, M., et al., Prognostic biomarkers in osteoarthritis. Curr Opin Rheumatol, 2013. 25(1):
p. 136-44. p. 136-44.

-
-
- 16. Rocha, F.A.C. and S.A. Ali, Soluble biomarkers in osteoartimis in 2022. year in review.

16. Luo, Y., et al., A low cartilage formation and repair endotype predicts radiographic

progression of symptomatic knee osteoar
- Luo, Y., et al., *A low cartilage formation as progression of symptomatic knee osteoarthritis.* J
Beier, F., *The impact of omics research on our treatments.* Current Opinion in Rheumatology, 20
Angelini, F., et al., *Ost* 17. Luo, Y., et al., A low cartingle formation and repair endotype predicts radiographic
progression of symptomatic knee osteoarthritis. J Orthop Traumatol, 2021. 22(1): p. 10.
18. Beier, F., The impact of omics research o
- progression of symptomatic kinee osteoartimitis. J Orthop Traumator, 2021. 22(1): p. 10.
Beier, F., The impact of omics research on our understanding of osteoarthritis and f
treatments. Current Opinion in Rheumatology, 202 18. Beier, F., The impact of omics research on our anderstanding of osteoarthins and fatale
treatments. Current Opinion in Rheumatology, 2023. **35**(1): p. 55-60.
19. Angelini, F., et al., *Osteoarthritis endotype discovery* treatments. Current Opinion in Rheumatology, 2023. 33(1). p. 35-60.
Angelini, F., et al., *Osteoarthritis endotype discovery via clustering*
data. Ann Rheum Dis, 2022. **81**(5): p. 666-675.
Luo, Y.Y., et al., *A low cartila* 19. Angelini, F., et al., *Osteoarthritis endotype discovery via clustering of biochemical marker*
 data. Ann Rheum Dis, 2022. **81**(5): p. 666-675.

20. Luo, Y.Y., et al., *A low cartilage formation and repair endotype p* data. Ann Rheam Dis, 2022. **81**(5): p. 666-675.
Luo, Y.Y., et al., *A low cartilage formation*
progression of symptomatic knee osteoarthriti
2021. **22**(1).
Werdyani, S., et al., *Endotypes of primary osteoardysis.* Rheumat
-
- 20. Luo, Y.Y., et al., A low cartinge formation and repair endotype predicts radiographic
progression of symptomatic knee osteoarthritis. Journal of Orthopaedics and Traumatology,
2021. 22(1).
21. Werdyani, S., et al., End progression of symptomatic kinee osteoartimitis. Southa of Orthopaedics and Tradinatology,
2021. 22(1).
Werdyani, S., et al., Endotypes of primary osteoarthritis identified by plasma metabolomics
analysis. Rheumatology (Ox 2021. 22(1).
Werdyani, S.
analysis. Rhe
Watt, F.E., e
28(3): p. 324
Watt, F.E., e 21. Werdyani, S., et al., *Endotypes of primary osteoarthritis identified by plasma metabolomics*

22. Watt, F.E., et al., *The molecular profile of synovial fluid changes upon joint distraction and is*

22. Watt, F.E., et analysis. Rheumatology (Oxford), 2021. **Oc**(6): p. 2733-2744.
Watt, F.E., et al., *The molecular profile of synovial fluid chan*
associated with clinical response in knee osteoarthritis. Ost
28(3): p. 324-333.
Watt, F.E.
-
- 22. Watt, F.E., et al., *The molecular profile of synovial fluid changes upon joint distraction and is*
associated with clinical response in knee osteoarthritis. Osteoarthritis and Cartilage, 2020.
28(3): p. 324-333.
Watt, associated with clinical response in kinee osteoarthritis. Osteoarthritis and Cartilage, 2020.
28(3): p. 324-333.
Watt, F.E., et al., *Acute Molecular Changes in Synovial Fluid Following Human Knee Injury:*
Association Wit 28(3): p. 324-333.
Watt, F.E., et al., *Association With E*
Struglics, A., et al.
and Serum and in
Crosslinking Telop
Ligament Rupture 23. Watt, F.E., et al., Acute Molecular Changes in Synovial Fridit Following Framan Knee Injury.

Association With Early Clinical Outcomes. Arthritis Rheumatol, 2016. **68**(9): p. 2129-40.

24. Struglics, A., et al., Change Association With Early Cinnear outcomes. Arthritis Rheumatol, 2010. 68(9): p. 2129-40.
Struglics, A., et al., Changes in Cytokines and Aggrecan ARGS Neoepitope in Synovial
and Serum and in C-Terminal Crosslinking Telopepti 24. Struglies, A., et al., Changes in Cytokines and Aggrecan ARGS Neoephope in Synovial Fiata
and Serum and in C-Terminal Crosslinking Telopeptide of Type II Collagen and N-Terminal
Crosslinking Telopeptide of Type I Colla and Serum and in C-Terminal Crossinking Telopeptide of Type II Collagen and N-Terminal Crosslinking Telopeptide of Type I Collagen in Urine Over Five Years After Anterior Cruciate
Ligament Rupture: An Exploratory Analysis in the Knee Anterior Cruciate Ligament, Nonsurgical Versus Surgical Treatment Trial. Arthritis & Rheumatology, 2015. 67(7): p. 1816-
- Nonsurgical Versus Surgical Treatment Trial. Arthritis & Rheumatology, 2015. 67(7): p. 1816
1825.
Garriga, C., et al., Clinical and molecular associations with outcomes at 2 years after acute
knee injury: a longitudinal st Garrig
knee i
Rheun
Broom
the hil
Deng,
-
- 25. Garriga, C., et al., Chincal and molecular associations with outcomes at 2 years after acute

knee injury: a longitudinal study in the Knee Injury Cohort at the Kennedy (KICK). Lancet

Rheumatology, 2021. 3(9): p. E64 knee injury: a longitudinal study in the knee injury cohort at the kennedy (KICK). Lancet
Rheumatology, 2021. 3(9): p. E648-E658.
Broomfield, J.A.J., *Using synovial fluid biomarkers to define a phenotype of osteoarthritis* Rheumatology, 2021. 3(9): p. E648-E658.
Broomfield, J.A.J., *Using synovial fluid bit*
the hip [PhD thesis]. 2020, University of O
Deng, Y., et al., *Methodological developn
fluid of individuals with knee osteoarthri*
2023 26. Broomfield, J.A.J., Using Synoviar final biomarkers to define a phenotype of osteoarthritis in
the hip [PhD thesis]. 2020, University of Oxford.
27. Deng, Y., et al., Methodological development of molecular endotype di the mp₁ mb thesisj. 2020, Oniversity of Oxford.
Deng, Y., et al., Methodological development of
fluid of individuals with knee osteoarthritis: the
2023.08.14.23294059.
Rohloff, J., et al., Nucleic Acid Ligands With Proi

- Err. Beng, Y., et al., Methodological development of molecular endotype discovery from synoviar
fluid of individuals with knee osteoarthritis: the STEpUP OA Consortium. medRxiv, 2023: p.
28. Rohloff, J., et al., Nucleic Ac fluid of individuals with knee osteoarthritis: the STEPOT OA Consortium. medically, 2023. p.
2023.08.14.23294059.
Rohloff, J., et al., Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and
Their Use as 2023.08.14.23294059.
Rohloff, J., et al., *Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and
Their Use as Diagnostic and Therapeutic Agents. Molecular therapy. Nucleic acids, 2014. 3: p.
e201.
Candi*
-
- 28. Rohloff, 3.7, et al., Nucleic Acid Engines With Protein-like Side Chains: Modified Aptamers and
Their Use as Diagnostic and Therapeutic Agents. Molecular therapy. Nucleic acids, 2014. 3: p.
29. Candia, J., et al., Asse Frien Ose as Diagnostic and Therapeutic Agents. Molecular therapy. Nucleic acids, 2014. 3: p.
e201.
Candia, J., et al., Assessment of variability in the plasma 7k SomaScan proteomics assay. Sci
Rep, 2022. 12(1): p. 17147.

Candia
Rep, 2
Pham,
the In
2005.
Charra 29. Candia, J., et al., Assessment of variability in the plasma 7k SomaScan proteomics assay. Sci
Rep, 2022. 12(1): p. 17147.
30. Pham, D.T., S.S. Dimov, and C.D. Nguyen, Selection of K in K-means clustering. Proceedings Rep, 2022. 12(1): p. 17147.
Pham, D.T., S.S. Dimov, and
the Institution of Mechania
2005. 219(1): p. 103-119.
Charrad, M., et al., *Nbclust:
a Data Set.* Journal of Statis
McInnes, L. and J. Healy. 30. Pham, B.H., 3.3. Billiot, and C.D. Nguyen, Selection of Mechanical Engineering Science,

30. 2005. 219(1): p. 103-119.

31. Charrad, M., et al., *Nbclust: An R Package for Determining the Relevant Number of Clusters in*
- 1005. 219(1): p. 103-119.
2005. 219(1): p. 103-119.
Charrad, M., et al., *Nbclust: An R Package for Determining the Relevant Number of Clusters in*
a Data Set. Journal of Statistical Software, 2014. **61**(6): p. 1-36.
McInn 2005. 219(1): p. 103-115.
Charrad, M., et al., *Nbclus
a Data Set.* Journal of Stat
Mclnnes, L. and J. Heal
Dimension Reduction. ArX
ComBat: Adjust for batch
2024]; Available from: htt
- 31. Charrad, M., et al., Nbclust: An R Puckage for Determining the Relevant Number of Clusters in

a Data Set. Journal of Statistical Software, 2014. 61(6): p. 1-36.

32. McInnes, L. and J. Healy, *UMAP: Uniform Manifold A* a Data Set. Journal of Statistical Software, 2014. 01(6): p. 1-36.
McInnes, L. and J. Healy, *UMAP: Uniform Manifold Appro*
Dimension Reduction. ArXiv, 2018. **abs/1802.03426.**
ComBat: Adjust for batch effects using an empi
- S2. McInnes, L. and J. Healy, OWAP: Omform Mamijola Approximation and Projection for
Dimension Reduction. ArXiv, 2018. **abs/1802.03426**.
33. ComBat: Adjust for batch effects using an empirical Bayes framework. 2022 [cited1 *Dimension Reduction. ATAN, 2016. abs/1802.03426.*
ComBat: Adjust for batch effects using an empirica
2024]; Available from: https://rdrr.io/bioc/sva/man/
Johnson, W.E., C. Li, and A. Rabinovic, Adjusting bat
using empiric
- 33. Combat: Adjust for batch effects using an empirical bayes framework. 2022 [cited10-02-2024]; Available from: https://rdrr.io/bioc/sva/man/ComBat.html.
34. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effect* Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in m*
Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in m*
using empirical Bayes methods. Biostatistics, 2007. **8**(1): p. 118-127
Korotkevi
-
- 34. Johnson, W.E., C. Li, and A. Rabinovic, Adjusting batch effects in microdirity expression data
using empirical Bayes methods. Biostatistics, 2007. 8(1): p. 118-127.
35. Valero-Mora, P.M., ggplot2: Elegant Graphics for using empirical Bayes methods. Biostatistics, 2007. **6**(1): p. 110-127.
Korotkevich, G., et al., *Fast gene set enrichment analysis.* bioRxiv, 20
Valero-Mora, P.M., *ggplot2: Elegant Graphics for Data Analysi.*
Software, B 35. Korotkevich, G., et al., Fast gene set emicimient analysis. bioliviv, 2021: p. 060012.
36. Valero-Mora, P.M., *ggplot2: Elegant Graphics for Data Analysis.* Journal of
Software, Book Reviews, 2010. **35**(1): p. 1 - 3. Sol. Valero-Mora, P.M., ggplotz: Elegant Graphics for Data Analysis. Journal of Statistical
Software, Book Reviews, 2010. 35(1): p. 1 - 3. Software, Book Reviews, 2010. $35(1)$: p. 1 - 3.

-
-
- 37. Benjamini, Y. and Y. Hochberg, Controlling the False Discovery Rate. A Practical and Powerful
Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B
(Methodological), 1995. 57(1): p. 289-300.
 Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B
(Methodological), 1995. 57(1): p. 289-300.
Wang, M., et al., *MMP13 is a critical target gene during the progression of osteoarthritis.*
Arth (Methodological), 1995. **37(1):** p. 289-300.

Wang, M., et al., *MMP13 is a critical targ*

Arthritis research & therapy, 2013. **15(1):** p.

Garnero, P., et al., *Cross sectional evaluatis*
 synovial tissue metabolism in 38. Wang, M., et al., *MMP13* is a critical target gene during the progression of osteoarthritis.
Arthritis research & therapy, 2013. 15(1): p. R5.
39. Garnero, P., et al., Cross sectional evaluation of biochemical markers Arthritis research & therapy, 2013. 15(1): p. R5.
Garnero, P., et al., *Cross sectional evaluation of*
synovial tissue metabolism in patients with knee
and joint damage. Annals of the rheumatic disea
Aigner, T., et al., *L* and joint damage. Annals of the rheumatic diseases, 2001. $60(6)$: p. 619-26.
- 39. Garnero, P., et al., Cross sectional evaluation of biochemical markers of bone, cartilage, and
synovial tissue metabolism in patients with knee osteoarthritis: relations with disease activity
and joint damage. Annals Aigner, T., et al., *Large-scale gene expression profiling reveals major pathogenetic pathways*
of cartilage degeneration in osteoarthritis. Arthritis and rheumatism, 2006. **54**(11): p. 3533-
2hang, W., et al., *Classifica* 40. Aigner, T., et al., Large-scale gene expression projumg reveals major pathogenetic pathways
of cartilage degeneration in osteoarthritis. Arthritis and rheumatism, 2006. 54(11): p. 3533-
44. Zhang, W., et al., Classifi
- of cartinage degeneration in osteoarthritis. Arthritis and medinatism, 2000. 54(11): p. 3333-44.
2hang, W., et al., *Classification of osteoarthritis phenotypes by metabolomics analysis.* BMJ
Open, 2014. 4(11): p. e006286. Zha
Ope
Sou
gen
Stei
phe
-
- 2H. Zhang, W., et al., Classification of osteoarthritis phenotypes by metabolomics analysis. BMJ
Open, 2014. 4(11): p. e006286.
42. Soul, J., et al., Stratification of knee osteoarthritis: two major patient subgroups ident
- Soul, J., et al., *Stratification of k*
Soul, J., et al., *Stratification of k*
genome-wide expression analys
Steinberg, J., et al., *Linking c*
phenotype in osteoarthritis. Anr
Wijesinghe, S.N., et al., Obesity
osteoarthr 42. Soul, J., et al., *Stratifiedton of knee osteoarthritis. two major patient subgroups identified by*
genome-wide expression analysis of articular cartilage. Ann Rheum Dis, 2018. **77**(3): p. 423.
43. Steinberg, J., et al genome-wide expression analysis of articular cartinge. Ann Kheum Dis, 2016. 77(3): p. 423.
Steinberg, J., et al., Linking chondrocyte and synovial transcriptional profile to clinico
phenotype in osteoarthritis. Ann Rheum D 43. Steinberg, J., et al., *Linking chondrocyte and synovial transcriptional projne to clinical*
phenotype in osteoarthritis. Ann Rheum Dis, 2021. **80**(8): p. 1070-1074.
44. Wijesinghe, S.N., et al., *Obesity defined molec* phenotype in osteoarthritis. Ann Kheum Dis, 2021. 80(8): p. 1070-1074.
Wijesinghe, S.N., et al., Obesity defined molecular endotypes in the syno
osteoarthritis provides a rationale for therapeutic targeting of fibroblas
Me
-
- Wijesinghe, S.N., et al., Obesity defined molecular endotypes in the synoviam of patients with
osteoarthritis provides a rationale for therapeutic targeting of fibroblast subsets. Clin Transl
Med, 2023. 13(4): p. e1232.
Fe osteoarthritis provides a rationale for therapeutic targeting of fibrookist subsets. Clin Transl
Med, 2023. 13(4): p. e1232.
Fernandez-Tajes, J., et al., *Genome-wide DNA methylation analysis of articular chondrocytes*
rev Med, 2023. 13(4): p. e1232.
Fernandez-Tajes, J., et al., *Genome-wide DNA methylation analysis of articular chondrocytes*
reveals a cluster of osteoarthritic patients. Ann Rheum Dis, 2014. 73(4): p. 668-77.
Ali, N., et al. 45. Fernandez-Tajes, J., et al., Genome-wide DNA methylation analysis of articular chondrocytes
reveals a cluster of osteoarthritic patients. Ann Rheum Dis, 2014. **73**(4): p. 668-77.
Ali, N., et al., Proteomics Profiling o
- reveals a cluster of osteoarthritic patients. Ann Rheam Dis, 2014. 73(4): p. 668-77.
Ali, N., et al., *Proteomics Profiling of Human Synovial Fluid Suggests Increas*
Interplay in Early-Osteoarthritis (OA) That Is Lost in L 46. Ani, N., et al., *Troteomics Trojining of Human Synovial Fluid Suggests Increased Trotem*
Interplay in Early-Osteoarthritis (OA) That Is Lost in Late-Stage OA. Mol Cell Proteomics,
2022. 21(3): p. 100200.
47. Carlson, Interplay in Early-Osteoarthritis (OA) That Is East in Eate Stage OA. Mol Cell Proteomics,
2022. 21(3): p. 100200.
Carlson, A.K., et al., Characterization of synovial fluid metabolomic phenotypes of cartilage
morphological 2022. 21(3): p. 100200.
Carlson, A.K., et al., *Chi*
morphological changes
p. 1174-1184.
Timur, U.T., et al., *Ide*
Osteoarthritis Cartilage,
Muthu, S., et al., *Fail* 47. Carlson, A.K., et al., *characterization of synovial fluid metabolomic phenotypes of cartilage*

morphological changes associated with osteoarthritis. Osteoarthritis Cartilage, 2019. **27**(8):

p. 1174-1184.

Timur, U.T
- r
Timur, U.T., e
Osteoarthritis
Muthu, S., et
therapeutic str
Bartley, E.J., S
Resilience. Cur
- morphological changes associated with osteoarthmis. Osteoarthmis cartilage, 2019. 27(0).

p. 1174-1184.

Timur, U.T., et al., *Identification of tissue-dependent proteins in knee OA synovial fluid.*

Osteoarthritis Cartila 48. Timur, U.T., et al., *Taentification of tissue dependent proteins in knee OA synovial fluid.*

0steoarthritis Cartilage, 2021. **29**(1): p. 124-133.

49. Muthu, S., et al., *Failure of cartilage regeneration: emerging h* Osteoarthritis Cartilage, 2021. 29(1): p. 124-133.
Muthu, S., et al., *Failure of cartilage regeneration: emerging hypotheses and related*
therapeutic strategies. Nat Rev Rheumatol, 2023. 19(7): p. 403-416.
Bartley, E.J.
- 49. Muthu, S., et al., Failure of cartilage regeneration. emerging hypotheses and related
therapeutic strategies. Nat Rev Rheumatol, 2023. 19(7): p. 403-416.
50. Bartley, E.J., S. Palit, and R. Staud, *Predictors of Osteo* therapeutic strategies. Nat Rev Rheumatol, 2023. 19(7): p. 403-416.
Bartley, E.J., S. Palit, and R. Staud, *Predictors of Osteoarthritis I*
Resilience. Current rheumatology reports, 2017. 19(9): p. 57. Bartley, E.J., S. Palit, and R. Staud, Predictors of Osteoarthritis Pain: the Importance of
Resilience. Current rheumatology reports, 2017. 19(9): p. 57. Resilience. Current rheumatology reports, 2017. 19(9): p. 57.

Figure 1: Endotype discovery by cluster analysis in Discovery, Replication and combined datasets.

(A) f(K) metric for non-IPS and IPS regressed analyses. Significant clustering was observed (f(k) <

0.85) within all thre (A) fix) metric for non-in-3 and in-3 regressed analyses. Significant clustering was observed (f(k) <
0.85) within all three datasets for non-IPS regressed analyses only (left panel). (B) Visualisation of
data structure a data structure and IPS on UMAP by dataset, stratified by non-IPS (top panel) and IPS regressed (bottom panel) analyses. $f(K)$ metric plots for Combined dataset stratified by (C) biological sex, (D) radiographic disease se (bottom panel) analyses. $f(K)$ metric plots for Combined dataset stratified by (C) biological sex, (D) radiographic disease severity (KL grades: 0-2 as 'non advanced OA' and ≥ 3 as 'advanced OA') or (E) blood staining (bottom panel) analyses. $f(x)$ metric plots for combined dataset stratified by (c) biological sex, (b) radiographic disease severity (KL grades: 0-2 as 'non advanced OA' and ≥ 3 as 'advanced OA') or (E) blood staining blood staining (visual blood staining: 1 as 'no blood staining' and ≥ 2 as 'with blood staining') for non-IPS and IPS regressed analyses. Abbreviations: osteoarthritis (OA), intracellular protein score (IPS), Uniform Mani non-IPS and IPS regressed analyses. Abbreviations: osteoarthritis (OA), intracellular protein score non-IPS and IPS regressed analyses. Abbreviations: osteoarthritis (OA), intracellular protein score
(IPS), Uniform Manifold Approximation and Projection (UMAP). (IPS), Uniform Manifold Approximation and Projection (UMAP).

Table 1: Baseline characteristics of participants, their SF samples and association of these factors with IPS. Association testing was carried out between IPS and core demographic, clinical and technical features in spun OA samples where relevant data were available. Linear regression models were constructed with log scaled IPS (i.e. IPS that were transformed using natural logarithms) as the outcome with each feature listed in the table used as a univariate exposure. Adjusted models where cohort was included as random intercept are also shown. Asterisks (bold) denote statistical significance at Benjamini-Hochberg cutoff (adjusted p-value ≤0.05). Abbreviations: osteoarthritis (OA), synovial fluid (SF), intracellular protein score (IPS), blood staining (BS), Kellgren Lawrence (KL), standard deviation (SD).

49 non-n 9 regressed data.
50 Protein abundance was
51 unspun (N=226)), correc

unspun (N=226)), corrected for spin-status by ComBat and then adjusted for age and biological sex. 51.1% unspin, (N=22)), corrected for spin-status by ComBat and then adjusted for age and biological sex.

(A) volcano plot showing log odds ratios against adjusted p-values (Benjamini-Hochberg corrected)

for proteins associated with advanced radiographic knee OA in the Combined dataset. Proteins in

red are positively associ 54 red are positively associated, those in blue negatively associated, with advanced radiographic status
55 (KL grades: 3-4) at an adjusted p-value ≤ 0.05 . Top associating proteins by adjusted p-value are
56 labelled (KL grades: 3-4) at an adjusted p-value \leq 0.05. Top associating proteins by adjusted p-value are labelled (top 30 positively and negatively associating proteins by adjusted p-value). Proteins that replicated (signific 56 labelled (top 30 positively and negatively associating proteins by adjusted p-value). Proteins that
57 replicated (significant at padj ≤0.05 and with effects in the same direction in Discovery & Replication
58 datasets 57 replicated (significant at padj \leq 0.05 and with effects in the same direction in Discovery & Replication
58 datasets), and that remained significant after the Combined dataset was adjusted for cohort
59 (random inte 57 datasets), and that remained significant after the Combined dataset was adjusted for cohort (random intercept) are shown in orange. Proteins that either did not replicate but remained significant after adjustment for co For the Combined attention of the Combined attention of the Combined Significant after adjustment for cohort, or did replicate but were not significant after cohort adjustment are shown in green. A single protein, GM-CSF n 59 Significant after adjustment for cohort, or did replicate but were not significant after cohort

51 adjustment are shown in green. A single protein, GM-CSF neither replicated nor was significant after

52 further adjus 61 adjustment are shown in green. A single protein, GM-CSF neither replicated nor was significant after
62 further adjustment for cohort (labelled white). (**B**) Select examples of protein expression values
63 (transformed 62 further adjustment for cohort (labelled white). (B) Select examples of protein expression values
63 (transformed by natural logarithms) by ordinal KL grade (N=766) from the Combined dataset.
64 Statistically significan Further adjustment for conort (labelled white). (B) Select examples of protein expression values

(transformed by natural logarithms) by ordinal KL grade (N=766) from the Combined dataset.

Statistically significant associ Statistically significant associations with ordinal KL grade were tested by ordinal regression analysis

(log odds ratio (OR) and unadjusted p-values are presented for each protein for models adjusted for

age and biologic (log odds ratio (OR) and unadjusted p-values are presented for each protein for models adjusted for

age and biological sex). Two additional OA-related proteins (MMP-13 & COL2) are included. Number

of samples in each gro For a geand biological sex). Two additional OA-related proteins (MMP-13 & COL2) are included. Number
67 of samples in each group are shown. (C) Bubble plot of significantly enriched pathways
68 (adjusted p-value <0.05) usi 67 of samples in each group are shown. (C) Bubble plot of significantly enriched pathways (adjusted p-value <0.05) using the Hallmark Gene set for advanced RKOA status for Discovery, Replication and Combined non-IPS, non-c 68 (adjusted p-value <0.05) using the Hallmark Gene set for advanced RKOA status for Discovery,
69 Replication and Combined non-IPS, non-cohort adjusted datasets. Protein-protein interaction
67 networks, using STRING, for Replication and Combined non-IPS, non-cohort adjusted datasets. Protein-protein interaction

70 networks, using STRING, for (D) Epithelial Mesenchymal Transition, (E) Complement and (F)

71 Angiogenesis pathways. (G) Scatt The metworks, using STRING, for (D) Epithelial Mesenchymal Transition, (E) Complement and (F)

71 Angiogenesis pathways. (G) Scatter plot of log odds ratio from logistic regression models of the

72 associations between pr The metworks, using STRING, for (D) Epithelial Mesenchymal Transition, (E) Complement and (F)

71 Angiogenesis pathways. (G) Scatter plot of log odds ratio from logistic regression models of the

72 associations between pr The Migogenesis pathways. (G) Scatter plot of log odds ratio from logistic regression models of the

27 associations between protein abundance and advanced radiographic disease status using either

27 Discovery or Replicat 73 Discovery or Replication datasets is shown with significantly associated proteins in different datasets

74 shown in different colours (see key). Pearson correlation coefficient and p-value (unadjusted) are

75 presente Frank of the correlation of the correlation of the data serves is always and the correlation between log odds ratio generated in Discovery and Replication analyses. Abbreviations: intracellular protein score (IPS), Kellgre Frame and the correlation between log odds ratio generated in Discovery and Replication

27 shown in different correlations: intracellular protein score (IPS), Kellgren-Lawrence (KL). Lymphatic vessel

27 endothelial hyalu 27 analyses. Abbreviations: intracellular protein score (IPS), Kellgren-Lawrence (KL). Lymphatic vessel

27 endothelial hyaluronan receptor 1 (LYVE1), Insulin-like growth factor-binding protein 6 (IGFBP-6),

28 Fibroblast endothelial hyaluronan receptor 1 (LYVE1), Insulin-like growth factor-binding protein 6 (IGFBP-6),

78 Fibroblast Growth Factor Binding Protein 1 (FGFP1), secreted frizzled-related protein 3 (sFRP-3),

79 tumour necrosis f Fibroblast Growth Factor Binding Protein 1 (FGFP1), secreted frizzled-related protein 3 (sFRP-3),

19 tumour necrosis factor-inducible gene 6 (TSG-6), soluble triggering receptor expressed on myeloid

180 cells-1 (sTREM-1) The sum our necrosis factor-inducible gene 6 (TSG-6), soluble triggering receptor expressed on myeloid

28 cells-1 (sTREM-1), activin A and vascular endothelial growth factor A-(isoform 121)(VEGF-121),

28 Matrix metallopr cells-1 (sTREM-1), activin A and vascular endothelial growth factor A-(isoform 121)(VEGF-121),
81 Matrix metalloproteinase-13 (MMP-13) and Collagen Type II (COL2), ON (osteonectin/SPARC). Full
82 list of proteins available 80 cells-1 (sTREM-1), activin A and vascular endothelial growth factor A-(isoform 121)(VEGF-121),

81 Matrix metalloproteinase-13 (MMP-13) and Collagen Type II (COL2), ON (osteonectin/SPARC). Full

82 list of proteins avai 181 Matrix metalloperature 21 (MMP-13) and Collagen Type II (COL2), ON (2000-0001), ON 13
82 list of proteins available in Supplementary Data file 1.
83
85
85
87 83
84
85
86
87
88
88 84
85
867
889
89

-
- 85
86
87
88
90
90 86
8788
8890
901
- 87
88890
9091
9200
-
- 90
- 91
- 92
- 88
8890
91
92
93 89
90
92
93
94
94 90
90
92
93
94
95
95 91
92
93
95
96
96 92
93
94
95
96
97 - 93
94 95 6
97 98 99 93
- 94
95
96
98
99
99 94 95
- 96
- 95
96 7
98 99
00 01 96
96
98
99
00
01
02 97 98
- 99
-
- 100
101
- 97
98
99
00
01
02 98
99
00
01
02 99
00
01
02 00
01
02 102

128
129
130
131

129
130
131
132
132 129 Figure 3: Association between protein abundance and advanced radiographic disease status

130 stratified by obese and non-obese OA participants using non-IPS regressed Combined data.

131 Protein abundance was measure 131 Protein abundance was measured in 1,236 patient samples where BMI was available (Co
132 dataset (N = 1,236, spun (N=1,045) and unspun (N = 191), corrected for spin-status are adjusted for age and biological sex. The gr dataset (N = 1,236, spun (N=1,045) and unspun (N = 191), corrected for spin-status and then
133 adjusted for age and biological sex. The groups were then stratified by BMI into obese, BMI ≥30 (N =
134 587, 504 spun sample and a division of the 1,236, spun (N = 1,236, spun ampendance in training the stratified by BMI into obese, BMI \geq 30 (N = 134 = 587, 504 spun samples) and non-obese, BMI <30 (N = 649, 541 spun samples) participants. Vol 134 587, 504 spun samples) and non-obese, BMI <30 (N = 649, 541 spun samples) participants. Volcano

135 plots show log odds ratio against adjusted p-values (Benjamini-Hochberg corrected) for proteins

136 associated with plots show log odds ratio against adjusted p-values (Benjamini-Hochberg corrected) for proteins
136 associated with radiographic disease severity in (A) obese and (B) non-obese groups using
137 Combined, non-IPS regressed plots show log odds ratio against adjusted p-values (Benjamini-Hochberg corrected) for proteins

136 associated with radiographic disease severity in (A) obese and (B) non-obese groups using

137 Combined, non-IPS regress 137 Combined, non-IPS regressed and age adjusted data. Proteins in red are positively associated, those

138 in blue negatively associated, with advanced radiographic disease at an adjusted p-value ≤ 0.05. Top

139 20 ass 138 in blue negatively associated, with advanced radiographic disease at an adjusted p-value ≤ 0.05 . Top
139 20 associated proteins by adjusted p-value are labelled. In orange are proteins that replicated
140 (signific 139 20 associated proteins by adjusted p-value are labelled. In orange are proteins that replicated

140 (significant at padj ≤0.05 and with effects in the same direction) across obese and non-obese groups.

141 (C) Pears Sum interest 140 (significant at padj \leq 0.05 and with effects in the same direction) across obese and non-obese groups.

141 (C) Pearson correlation of the log odds ratios comparing associations between protein expres 141 (C) Pearson correlation of the log odds ratios comparing associations between protein expression

142 and advanced radiographic disease status in obese and non-obese groups. (D) Bubble plot of

27 141 (C) Pearson correlation of the log odds ratios comparing associations between protein expression
142 and advanced radiographic disease status in obese and non-obese groups. (D) Bubble plot of
27 142 and advanced radiographic disease status in obese and non-obese groups. (D) Bubble plot of

- 143 significantly enriched pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins
144 associated with advanced radiographic disease status by obesity status. Abbreviations: intracellular
145 protein sco
-
- 145 protein score (IPS). Full list of proteins is available in Supplementary data file 4.
146
147 146
147
148
149
-
-
- 147
148
149 ---
148
149
150
151
- ---
149
150
151
152
- 150
151
152
153
-
-
- 151
152
153
154 153
154
- 152
153
154
155 ---
154
155
156
157
-
- 155
156
157
158
- 156
157
158
159 ---
157
158
159
160
-
-
-
- 158
159
160
161
- ---
159
160
161
162 160
161
162
163 161
162
163
164
- 162
163
164 163
164
165
166
-
-
-
- ---
164
165
166
167
- 165
166
167
168
- ---
166
167
168
169 ---
167
168
169
170 168
169
170
- 169
170
171
172
-
-
- 172
- ---
170
171
172
173 ---
171
172
173
174 172
-
- 173

195
195
196
198
199 195 Figure 4: Association between protein abundance and radiographic OA severity after stratifying for

196 biological sex using non-IPS regressed Combined data.

197 (N = 226)), corrected for spin-status and then adjuste Biological sex using non-in Stegressed combined data.
197 Protein abundance was measured in 1,322 samples (Co
198 (N = 226)), corrected for spin-status and then adjusted
199 against adjusted p-values (Benjamini-Hochberg co 197 Protein abundance was measured in 1,322 samples (Combined dataset, spun (N=1,056) and unspun

198 (N = 226)), corrected for spin-status and then adjusted for age. Volcano plots showing odds ratios

201 adiographic dis and a metally interest in the combined dataset stratified by (A) males (N = 623) and (B)
and adiographic disease severity in the Combined dataset stratified by (A) males (N = 623) and (B)
201 females (N = 699). Proteins in against adjusted p-values (Benjamini-Hochberg corrected) for proteins associated with advanced
200 radiographic disease severity in the Combined dataset stratified by (A) males (N = 623) and (B)
201 females (N = 699). Pro Frame and the severity in the combined dataset stratified by (A) males (N = 023) and (B)
201 females (N = 699). Proteins in red are positively associated, those in blue negatively associated, with
202 increased radiograph 202 increased radiographic disease severity at an adjusted p-value \leq 0.05. Top 20 associated proteins in each direction, by p-value, are labelled. Orange labelled proteins represent those that were significantly assoc each direction, by p-value, are labelled. Orange labelled proteins represent those that were
204 significantly associated in both males and females, whereas white labelled proteins were only
205 associated in the sex-speci 204 significantly associated in both males and females, whereas white labelled proteins were only associated in the sex-specific set. (C) A scatter plot of log odds ratio from logistic regression models of the association Experience in the sex-specific set. (C) A scatter plot of log odds ratio from logistic regression models
206 of the association between protein abundance and advanced radiographic disease status in males
207 and females is 205 associated in the sex-specific set. (C) A scatter plot of log odds ratio from logistic regression models
206 of the association between protein abundance and advanced radiographic disease status in males
207 and female 207 and females is shown with significantly associated proteins in different groups in different colours
208 (see key). Pearson correlation coefficient and p-value (unadjusted) are presented for the correlation
209 between 208 (see key). Pearson correlation coefficient and p-value (unadjusted) are presented for the correlation
209 between log odds ratio generated in male and female sex-specific analyses. (D) Bubble plot of
210 significantly 209 between log odds ratio generated in male and female sex-specific analyses. (D) Bubble plot of significantly enriched pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins 210 209 between log odds ratio generated in male and female sex-specific analyses. (D) bubble plot of
210 significantly enriched pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins
29 210 significantly enriched pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins

211

--
231
232
234
235
236 232
233
234
235
236
237 **Example 1:** Association between protein abundance and WOMAC pain in non-IPS regressed data.

233 Protein abundance was measured in 805 samples where WOMAC pain was available (Discovery and

234 Replication, spun (N=748) 233 Protein abundance was measured in 805 samples where WOMAC pain was available (Discovery and
234 Replication, spun (N=748) and unspun (N = 57)), corrected for spin-status and then adjusted for age
235 and biological se 235 and biological sex. (A) Volcano plot showing beta estimates against adjusted p-values (Benjamini-
236 Hochberg corrected) for proteins associated with WOMAC pain subscore in the Combined dataset.
237 Proteins in red a and biological sex. (A) volcano plot showing beta estimates against adjusted p-values (Benjamin-

236 Hochberg corrected) for proteins associated, those in blue negatively associated, with increasing

238 WOMAC pain subsc Proteins in red are positively associated, those in blue negatively associated, with increasing

238 WOMAC pain subscore at an adjusted p-value ≤ 0.05 . Top associated proteins, for each direction,

239 ordered by adju 238 WOMAC pain subscore at an adjusted p-value ≤ 0.05 . Top associated proteins, for each direction,
239 ordered by adjusted p-value are labelled. In green are two proteins that were significant in
240 Replication and 239 ordered by adjusted p-value are labelled. In green are two proteins that were significant in Replication and Combined datasets, including after cohort adjustment (random intercept), see
241 Supplementary Figure 4). (240 Replication and Combined datasets, including after cohort adjustment (random intercept), see
241 Supplementary Figure 4). (B) A scatter plot of beta estimates from linear regression models of the
242 association betwee 241 Supplementary Figure 4). (B) A scatter plot of beta estimates from linear regression models of the
242 association between protein abundance and WOMAC knee pain in non-IPS analyses is shown for
243 Discovery and Replic 241 Supplementary Figure 4). (B) A scatter plot of beta estimates from linear regression models of the
242 association between protein abundance and WOMAC knee pain in non-IPS analyses is shown for
243 Discovery and Replic Discovery and Replication datasets with significantly associated proteins in different groups shown in

244 different colours (see key). Pearson correlation coefficient and p-value (unadjusted) are presented

245 for the c different colours (see key). Pearson correlation coefficient and p-value (unadjusted) are presented

245 for the correlation between beta estimates generated in non-IPS regressed analyses using Discovery

246 and Replicati 245 for the correlation between beta estimates generated in non-IPS regressed analyses using Discovery
246 and Replication datasets. (C) Scatter plots of WOMAC pain subscore against NOE2 or NAR3
247 protein abundance (tran and Replication datasets. **C** Scatter plots of WOMAC pain subscore against NOE2 or NAR3
247 protein abundance (transformed by natural logarithms) in OA participants using Combined, spin-
248 status corrected, non-IPS regre 247 and Replication datasets. (C) Scatter plots of WomAc pain subscore against NOE2 of NARS
247 protein abundance (transformed by natural logarithms) in OA participants using Combined, spin-
248 status corrected, non-IPS r 249 for linear models adjusted for age and biological sex. (D) Bubble plot of significantly enriched
250 pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins associated with WOMAC
30 249 for linear models adjusted for age and biological sex. (D) bubble plot of significantly enriched
250 pathways (adjusted p-value <0.05) using the Hallmark Gene set for proteins associated with WOMAC
30 250 pathways (adjusted p-value $\frac{1}{2}$ using the Hallmark General proteins associated with WOMAC
30

251 knee pain by Replication and Combined datatsets not adjusted for IPS or cohort. No pathways were
252 significantly enriched at padj <0.05 in the Discovery dataset. Abbreviations: osteoarthritis (OA),
253 standard devia 253 standard deviation (SD), intracellular protein score (IPS), Western Ontario and McMaster
254 Universities Osteoarthritis Index (WOMAC, 0 = no pain, 100 = worst possible pain). Full list of
255 proteins available in Sup 254 Universities Osteoarthritis Index (WOMAC, 0 = no pain, 100 = worst possible pain). Full list of
255 proteins available in Supplementary data file 6.
256
257
258 255 Proteins available in Supplementary data file 6.
255 proteins available in Supplementary data file 6.
256
257
258
260 256 proteins available in Supplementary data file 6.
257 proteins available in Supplementary data file 6.
260 proteins available 6.
261 --
257
258
259
260
261 --
258
259
260
262
262 ---
259
260
261
262
263 ---
260
261
262
263 261
262
263
264 262
263
264
265 ---
263
264
265 264
265 --
265
266
267 ---
266
267
268 267 268 268

270