Osteoarthritis endotype discovery via clustering of biochemical marker data

Angelini, F.; Widera, P.; Mobasheri, A.; Blair, J.; Struglics, A.; Uebelhoer, M.; ... ; Bacardit, J.

EPIDEMIOLOGICAL SCIENCE

Osteoarthritis endotype discovery via clustering of biochemical marker data

Federico Angelini ,[1] Paweł Widera ,[1] Ali Mobasheri ,[2,3,4,5,6] Joseph Blair ,[7] André Struglics ,[8] Melanie Uebelhoer ,[9] Yves Henrotin ,[9,10] Anne CA Marijnissen,[4] Margreet Kloppenburg ,[11,12] Francisco J Blanco ,[13] Ida K Haugen ,[14] Francis Berenbaum ,[15] Christoph Ladel ,[16] Jonathan Larkin ,[17] Anne C Bay-Jensen ,[7] Jaume Bacardit [1]

Handling editor Josef S Smolen

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2021-221763).

For numbered affiliations see end of article.

Correspondence to
Jaume Bacardit, School of Computing, Newcastle University, Newcastle upon Tyne, UK;
jaume.bacardit@newcastle.ac.uk

Received 3 November 2021
Accepted 1 February 2022
Published Online First 4 March 2022

ABSTRACT

Objectives Osteoarthritis (OA) patient stratification is an important challenge to design tailored treatments and drive drug development. Biochemical markers reflecting joint tissue turnover were measured in the IMI-APPROACH cohort at baseline and analysed using a machine learning approach in order to study OA-dominant phenotypes driven by the endotype-related clusters and discover the driving features and their disease-context meaning.

Method Data quality assessment was performed to design appropriate data preprocessing techniques. The k-means clustering algorithm was used to find dominant subgroups of patients based on the biochemical markers data. Classification models were trained to predict cluster membership, and Explainable AI techniques were used to interpret these to reveal the driving factors behind each cluster and identify phenotypes. Statistical analysis was performed to compare differences between clusters with respect to other markers in the IMI-APPROACH cohort and the longitudinal disease progression.

Results Three dominant endotypes were found, associated with three phenotypes: C1) low tissue turnover (low repair and articular cartilage/subchondral bone turnover), C2) structural damage (high bone formation/resorption, cartilage degradation) and C3) systemic inflammation (joint tissue degradation, inflammation, cartilage degradation). The method achieved consistent results in the FNIH/OAI cohort. C1 had the highest proportion of non-progressors. C2 was mostly linked to longitudinal structural progression, and C3 was linked to sustained or progressive pain.

Conclusions This work supports the existence of differential phenotypes in OA. The biomarker approach could potentially drive stratification for OA clinical trials and contribute to precision medicine strategies for OA progression in the future.

Trial registration number NCT03883568.

Key messages

What is already known about this subject?

- There is an unmet need for new therapies that target the underlying pathology in osteoarthritis (OA).
- Computational methods based on unsupervised machine learning have the potential to stratify OA cohorts into subsets that correspond to distinct molecular endotypes.

What does this study add?

- By applying these methods to the IMI-APPROACH cohort, we identified three dominant clusters and characterised them as inflammatory, low-repair and subchondral bone/articular cartilage-driven phenotypes.
- Patients in the discovered clusters had statistically significant differences in clinical characteristics.

How might this impact on clinical practice or future developments?

- The biomarker-based endotype discovery approach could potentially drive stratification for OA clinical trials and contribute in the future to precision medicine strategies for OA care.

INTRODUCTION

Osteoarthritis (OA) is the most common form of arthritis among older people, affecting more than 500 million people (7% of the global population).[1] It is one of the most frequent causes of physical disability among older individuals and a major contributor to healthcare and societal costs globally.[2] The risk factors for the development of OA include age, sex, obesity, previous joint injuries, repeated stress on the joint, malalignment, genetics, bone shape (including deformities) and certain metabolic diseases.[3] According to studies on the global burden of disease, knee OA represents the greatest burden.[4 5] However, despite the ever-increasing rise in the incidence and burden of OA, there is an unmet need for new therapies that target the underlying pathophysiologies.[6] The currently available pharmacological treatments are only able to target the symptoms of OA, and they have adverse side effects, especially in older adults with common comorbidities.

The development of effective treatments and disease-modifying OA drugs (DMOADs) for this debilitating condition is extremely challenging.[7] Many of the approaches that have been tried thus far have either failed or produced unsatisfactory outcomes. One of the greatest

To cite: Angelini F, Widera P, Mobasheri A, *et al*. *Ann Rheum Dis* 2022;**81**:666–675.

Table 1 Biochemical markers analysed in the APPROACH cohort sampled from serum (S) and urine (U)

Name	Inter- and intra-CV	Detection range	Description
S_C3M	<15%	1–85 ng/mL	Matrix metalloproteinase (MMP)-mediated type III collagen degradation fragment. Type III collagen is a major collagen of connective tissues, including synovial membrane. C3M has been shown to be released from synovial membranes in the presence of proinflammatory cytokines which activate MMPs.[43]
S_CRPM	<15%	1–110 ng/mL	MMP-mediated C reactive protein (CRP) degradation fragment. CRP is an acute reactant elevated in chronic inflammatory diseases. CRPM is a metabolite of CRP.[44]
S_ARGS	<15%	0.01–0.40 pmol/mL	ADAMTS-mediated aggrecan degradation products. Aggrecan is the major proteoglycan of articular cartilage. Like MMPs, ADAMTS are expressed and activated in the presence of proinflammatory cytokines.[45]
S_C10C	<15%	500–7500 ng/mL	Cathepsin K-mediated type X collagen degradation fragment. Type X collagen is a minor collagen expressed by the cartilage cells called chondrocytes.[46]
S_C2M	<15%	0–10 ng/mL	MMP-mediated type II collagen degradation fragment. Type 2 collagen is the major fibrillar protein of cartilage and C2M is released on activation of MMPs.[47]
S_COLL2_1	<15%	200–2200 nM	Type II collagen degradation fragment similar, but from a different domain compared with C2M.[48]
S_COLL2_1NO2	<15%	150–6000 pg/mL	Inflammation-related (nitrated) type-II collagen degradation fragment. Nitrosylation is a post-translational modification induced by an increase in oxidative stress associated with inflammation.[48]
S_COMP	<15%	1–50 units/L	Cartilage oligomeric matrix protein. COMP is articular cartilage protein, which is released when cartilage is turned over.[49]
S_CTXI	<10%	0–3 ng/mL	Cross-linked, isomerised and cathepsin K-generated fragment of type I collagen C-terminal telopeptide. Type I collagen is the major fibrillar protein of bone and some connective tissues. Cathepsin K is mainly expressed by osteoclast, making CTX-I a marker of bone resorption.[50]
S_HA	<15%	10–800 ng/mL	Hyaluronic acid is a glycosaminoglycan distributed widely across connective, epithelial and neural tissues, including articular cartilage. It is released as part of tissue remodelling and turnover induced by, for example, inflammation.[50]
S_hsCRP	<10%	0–60 mg/L	High-sensitive C reactive protein (hsCRP) is an acute reactant elevated in chronic inflammatory diseases and used as a diagnostic marker in different rheumatic diseases.[51]
S_PRO_C2	<10%	5–1000 ng/mL	Type IIB collagen propeptide (synthesis). When new type II collagen is expressed by cartilage cells, PRO-C2 is released and is a reflection of cartilage formation.[12]
S_NMID	<10%	1–180 ng/mL	Bone gamma-carboxyglutamic acid-containing protein.[52]
S_RE_C1M	<15%	10–500 ng/mL	MMP-mediated type I collagen degradation. See S_C3M and S_CTX-I.[53]
U_CTXI_ALPHA	<15%	0–10 µg/mmol	Cathepsin K-generated fragment of type I collagen C-terminal telopeptide (corrected for creatinine) is a non-isomerised version of S_CTX-I and therefore believed to reflect degradation of young bone in contrast to the isomerised which measures old bone.[54]
U-CTXII	<15%	10–2500 ng/mmol	MMP- and cathepsin K-mediated type II collagen degradation fragment (corrected for creatinine). See CTX-I and C2M as well.[50]

Coefficient of variation (CV) and detection range are shown; for further assay validation, see references.

challenges in OA drug development is the heterogeneity of the disease.[8 9] However, despite being a multifaceted and heterogeneous syndrome, there is an opportunity to target different treatments to patients according to their disease drivers characterised by molecular endotypes (a description of a subset of patients with common molecular characteristics) and clinical phenotypes (an observable characteristic or trait of a disease).[9 10] OA may be amenable to tailored treatments that target specific phenotypes, including inflammatory, low repair, subchondral bone, metabolic or articular cartilage-driven phenotypes.[11–17]

Therefore, development of computational tools that includes objectively measured markers, such as biochemical markers, may facilitate OA drug development through patient subgrouping based on endotypic characteristics.[9 11] An example of an OA endotype could be a group of patients with elevated bone biochemical markers, as compared with the remaining of the OA population. Then, based on the link with clinical data, this subgroup could be annotated as having a bone-driven disease (ie, a OA disease phenotype), and hypothetically, this group of patients should be enriched for in clinical trials testing the efficacy of a bone-modulating drug.[18]

At present, defining the appropriate outcome measures that are needed for OA clinical trials and the objective assessment of new therapies is challenging.[19] Therefore, new computational methods based on machine learning (ML) and big data analytics can help advance this field of research by enabling protocols for patient classification into subtypes, using a combination of clinical, biochemical and/or imaging data.[20–22]

The aim of this study was to develop a methodology based on unsupervised ML (specifically, clustering) to identify/discover OA endotypes in the IMI-APPROACH cohort of patients with knee OA from a panel of 16 biochemical markers related to different joint tissue processes (eg, degradation, formation or inflammation), measured at the baseline of the study. The properties of the discovered clusters were thoroughly analysed using a combination of statistical and ML techniques, and the

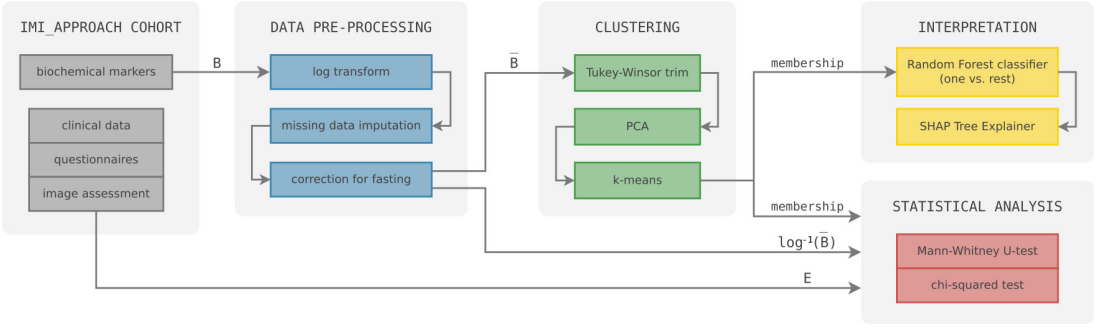

Figure 1 Overview of the data analysis pipeline. PCA, principal component analysis.

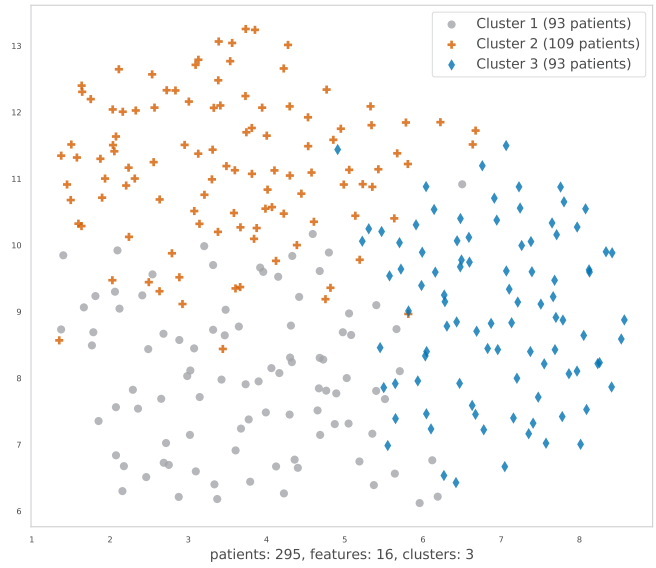

Figure 2 Clustering visualisation (k=3) obtained with UMAP (Uniform Manifold Approximation and Projection).

consistency of the discovered clusters was validated using data from an external cohort.

METHODS

Cohort description and data collection

Applied Public-Private Research enabling OsteoArthritis Clinical Headway funded by the Innovative Medicines Initiative (IMI-APPROACH, trial registration number: NCT03883568) is a prospective cohort study including 297 patients with tibiofemoral OA according to the American College of Rheumatology classification criteria. Patients were (pre)selected from existing cohorts using ML models, developed on data from the CHECK cohort, to display a high likelihood of radiographic joint space width (JSW) loss and/or knee pain progression.[23 24] The ultimate objective of APPROACH is to use real-world data to develop analysis methodologies to define disease subtypes and identify different knee OA clusters/phenotypes, to allow targeted treatment.

The IMI-APPROACH cohort screened 433 patients with OA (at five centres: Utrecht and Leiden, The Netherlands; A Coruña, Spain; Paris, France; Oslo, Norway) and enrolled 297 patients most likely to be pain and/or structural progressors at 2-year follow-up.[24] Enrolled patients were predominantly

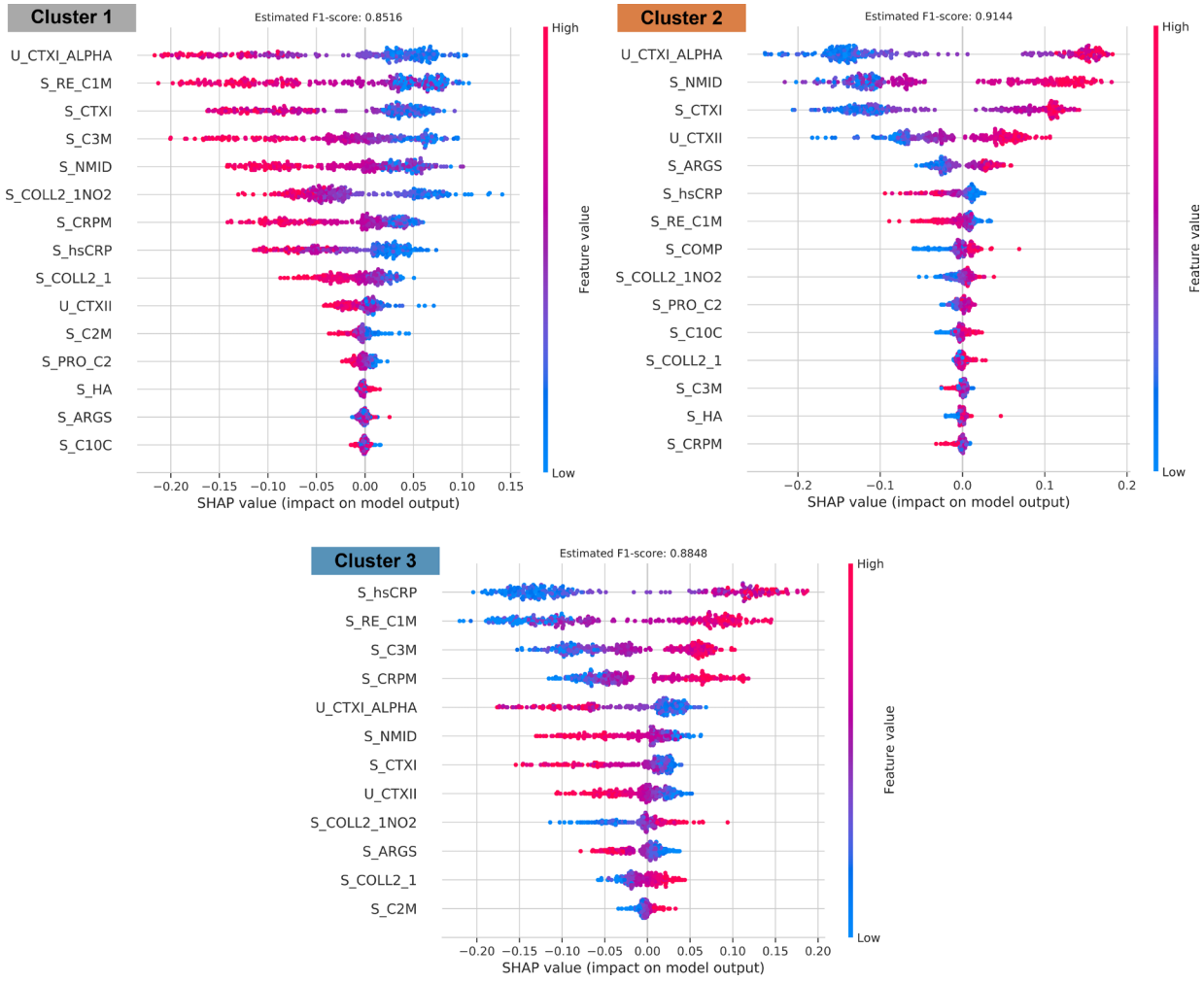

Figure 3 Impact of biomarker values on classification models decisions. Biomarkers are ordered by importance (most important on top). The SHAP values on the x-axis represent strength and direction of impact (positive value indicates increased probability of belonging to the cluster) for each patient. The colour represents the biomarker value (blue if low, red if high).

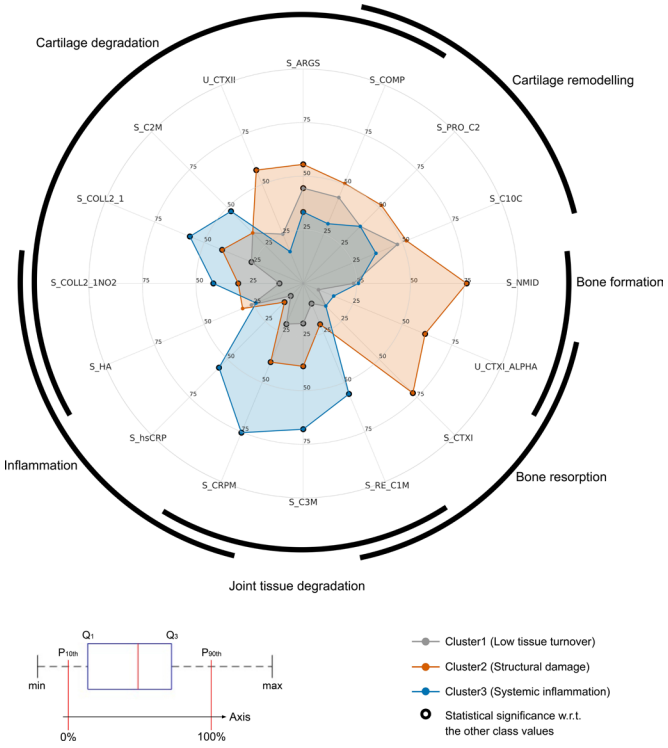

Figure 4 Radar plot showing the median biomarker concentrations for each cluster. When the difference between the medians is statistically different, it is marked with a circle (instead of a dot). The axes show values between the 10% and 90% quantile and are expressed as percentages. The black arcs on the outside show the pathology associated with each biomarker.

women (n=230), predominantly Caucasian/white (n=283), aged 44–82 years (median age: 67.5, IQR 62–71 years) and mostly overweight (median body mass index (BMI): 27 kg/m^2, IQR 24.4–31.6). At baseline, serum (S) and urine (U) samples were collected for analyses of 16 biochemical markers (table 1). The biomarkers were measured in International Organization for Standardization-certified laboratories at Nordic Bioscience (S_RE_C1M, S_C2M, S_C3M, S_C10C, S_CRPM, S_PRO_C2, U_CTXII, S_CTXI, U_CTXI_ALPHA, S_NMID, S_HA, S_COMP and S_hsCRP), Artialis (S_COLL2_1 and S_COLL2_1NO2) and Lund University (S_ARGS). The list of biomarkers was selected based on present knowledge of joint tissue turnover and OA.

In addition to the biochemical markers data (B), extra information (E) was collected as part of the IMI-APPROACH cohort.[23] These included assessment of radiographs of knees and hands, MRIs and CT scans of the knees, and outcomes of physical examinations and questionnaires: Function Index of Hand OA (FIHOA), Hip Disability and Osteoarthritis Outcome Score, Intermittent and Constant Osteoarthritis Pain Score, Knee-Injury and Osteoarthrosis Outcome Score (KOOS) and the 36-Item Short Form Health Survey. See online supplemental table B1. All data used in this paper were collected at the baseline visit of the study, except for the data on progression (Relation of clusters to progression section).

Data preprocessing

The biochemical markers data (B) were log transformed to account for long-tailed distributions. Missing data in B (<0.01% of values) were estimated (imputed) using either random Forest (RF) or k-nearest neighbour (KNN) regression models (see online supplemental appendix A, section 1.1).

As not all patients fasted before the sample collection, the fasting sensitivity of the biomarkers had to be assessed. The Spearman rank correlation with the patient's fasting status was found to be weak, except for U_CTXI (r=0.41). The values for this biomarker were corrected with an imputation approach (see online supplemental appendix A, section 1.2). We opted for a model-agnostic correction (ie, correcting the data rather than altering the analysis model) because it is more suitable for the downstream ML analysis we performed. $\bar{B}$ identifies the processed biomarkers data.

Clustering process

The extremes values of $\bar{B}$ were trimmed with a combination of Tukey and Winsor methods[25] to reduce the effect of outliers. Afterwards, principal component analysis was used to eliminate correlated biomarkers (see online supplemental appendix A, section 1.4). This resulted in 13 principal components which were found to explain 95% of data variance. These components were clustered using the k-means algorithm.[26] The optimum value for k (number of clusters) was identified from the consensus of silhouette score, the j-score and adjusted mutual information score. To obtain a robust estimate of these metrics, for each $k = 1, \ldots, 9$ the k-means algorithm was run 10 times with different random seeds (see online supplemental appendix A, section 1.5). The clustering with the highest quality was found for $k = 2, 3$. We chose $k = 3$ for the rest of the analysis in this paper as we aimed to investigate the highest number of meaningful clusters. The final cluster membership was taken from the algorithm run with the highest silhouette score for $k = 3$.

Cluster interpretation

Using data in $\bar{B}$, we trained three RF classification models predicting membership to each cluster (one cluster vs the rest) and then interpreted the model decisions using the SHAP (SHapley Additive exPlanations) TreeExplainer method,[27] to understand which variables determine the cluster membership. RF hyperparameters were tuned through a nested cross-validation procedure with recursive feature elimination (RFE-CV). See online supplemental appendix A, section 1.6 for more details.

Statistical analysis of cluster differences

To further describe the clusters, statistical tests were conducted for each feature in $\bar{B}$ and E, to assess whether the clusters had statistically different distributions for individual markers. The Mann-Whitney U test was used for continuous and ordinal features, and the $\chi 2$ test for categorical ones. The clusters were compared pairwise, and the null hypothesis was rejected following the Benjamini-Hochberg correction procedure for multiple comparisons applied across features.[28] The features in $\bar{B}$ were inverse log transformed to operate on actual biomarker concentrations (see online supplemental appendix A, section 1.3, for normality tests).

Figure 1 shows an overview of the entire data analysis pipeline described in this section, including data preprocessing, clustering, cluster's interpretation and the statistical analysis.

Validation on an external cohort

The proposed clustering pipeline was also applied to FNIH/OAI. The FNIH/OAI is the largest available OA cohort that was similar to IMI-APPROACH in terms of biomarkers.[29] The two cohorts had 11 biomarkers in common. Incurrent sample remeasurement for the adjustment of technical batch effects could not

be performed, as no samples were left and available from the FNIH/OAI cohort for this purpose. Therefore precise data alignment on the absolute mean concentrations and variance between the two cohorts was not possible to conduct.[30 31] As a result, the only possible type of external validation consisted in replicating the clustering pipeline for the two cohorts restricted to the common set of 11 biomarkers and evaluating the consistency of the identified clusters across cohorts.

Potential age and sex-based bias

To investigate the potential bias of age and gender in the clustering process, we statistically analysed the differences in age and sex across clusters, and we applied our clustering pipeline separately to the male and female subcohorts for both IMI-APPROACH and FNIH/OAI, to assess the consistency across clusters.

Relation of clusters to progression

To verify a relation between the clusters and disease progression, we used 2-year follow-up data to decide for each patient whether and how they have progressed, available only for a subset of 221 IMI-APPROACH participants. We defined one non-progressive category and three progressive categories related to pain, structure, and combined pain and structure.[23 24] Then we analysed the distribution of progressors in each cluster. See online supplemental appendix E for more details.

RESULTS

Cluster interpretation

Our clustering pipeline identified three clusters. These are shown in figure 2 as a two-dimensional projection obtained with UMAP (Uniform Manifold Approximation and Projection).[32] UMAP hyperparameters were optimised via grid search to maximise the two-dimensional silhouette score. The projection preserves the local neighbourhood structure and gives an idea of the strength of the global separation between the clusters in the original multidimensional space.

The classification models trained to predict patient's cluster membership achieved high F1 scores (C1 vs rest: 0.85, C2 vs rest: 0.91, C3 vs rest: 0.88). As a result, the subsequently performed model interpretation was expected to be meaningful. Figure 3 shows which biomarkers were predominantly used by each model to decide the cluster membership. Figure 4 compares the median biomarker concentrations for each cluster in a radar plot. Figure 5 shows the differences in biomarker value distributions across clusters. Bringing all these results together, the three clusters were interpreted as follows:

- *Cluster 1* represents a *low tissue turnover* phenotype: patients have all the inflammation and structural damage related biomarkers in the mid/low ranges.
- Cluster 2 represents a *structural damage* phenotype: patients have high values of the bone and cartilage markers: S_CTXI, U_CTXIALPHA, S_NMID and U_CTXII.
- *Cluster 3* represents a *systemic inflammation* phenotype: patients have high values of the inflammatory and MMP-driven markers: S_hsCRP, S_RE_C1M, S_CRPM and S_C3M. In contrast, these patients show low values of bone and cartilage related markers: U_CTXIALPHA, S_NMID, and S_CTXI.

Clustering stability

The clustering stability was investigated by comparing the results obtained for $k = 3$ with those obtained for $k = 4$ and $k = 5$. We found that clusters and interpretation were reasonably preserved at least until $k = 5$. This demonstrates that the three clusters analysed in this work are well-defined in the data space and robust with respect to finer clustering (see online supplemental appendix A, section 1.7).

Statistical analysis of differences between clusters

Several statistically significant differences in clinical scores were found. Full results are provided in online supplemental appendix B, and here we only present highlights of those findings. All figures cited in this section are provided in online supplemental appendix B.

- Clusters 2 and 3 had a higher percentage of women than cluster 1, and cluster 3 had a higher mean BMI (online supplemental figure B15).
- There was no difference in median age and range, smoking status, comorbidities and use of OA medication (online supplemental figure B14) across the clusters.
- Cluster 3 had statistically more patients experiencing substantial pain when standing (KOOS_P09, online supplemental figure B9), burning sensation (pain detect 09, online supplemental appendix B, B14) and more pain now and on average over the past 4 weeks (pain detect 01 and 03, online supplemental figure B14) than clusters 1 and 2. Patients in cluster 2 also experienced more pain in the past week than those in cluster 1 (pain detect 03, B14). Maximum Numeric Rating Scale (NRS) pain for hands were higher in cluster 3 (online supplemental figure B15), as well as having worse overall health self-assessment (SF36_11d, online supplemental figure B17).
- Cluster 1 has higher knee JSW (mean) than cluster 2 and less severe carpometacarpal Kellgren-Lawrence scores compared with cluster 3 (online supplemental figure B17).

External validation using FNIH/OAI data

We reduced the set of data features to the common subset of 11 biomarkers across the IMI-APPROACH and FNIH/OAI cohorts and applied the same clustering pipeline to both datasets. Figure 6 shows the comparison of obtained clusters. Despite the removal of five biomarkers, the IMI-APPROACH clusters still corresponded to structural damage, inflammatory and low tissue turnover endotypes. The FNIH/OAI clusters were found to consistently exhibit the same patterns, demonstrating cross-cohort robustness of our approach (see online supplemental appendix C).

Analysis of age and gender-based bias

We found no statistical difference between clusters in terms of age, as well as no statistical difference between male and female subcohorts in terms of age (see online supplemental figure D1). However, the male and female subcohorts had statistically different distributions for the following eight biomarkers: S_ARGS, S_C10C, S_COLL2_1, S_COLL2_1NO2, S_CTXI, S_NMID, U_CTXII and U_CTXI_ALPHA. Moreover, the clusters were significantly different in terms of gender, suggesting that it plays an important role in driving the clustering results (online supplemental figure D4). Similar patterns could be found for the FNIH/OAI cohort (see online supplemental appendix D).

Relation of clusters to progression

Table 2 summarises the progression status of the clusters. While we found progressors in all clusters, they were not distributed uniformly by progression type. There was more pain-related

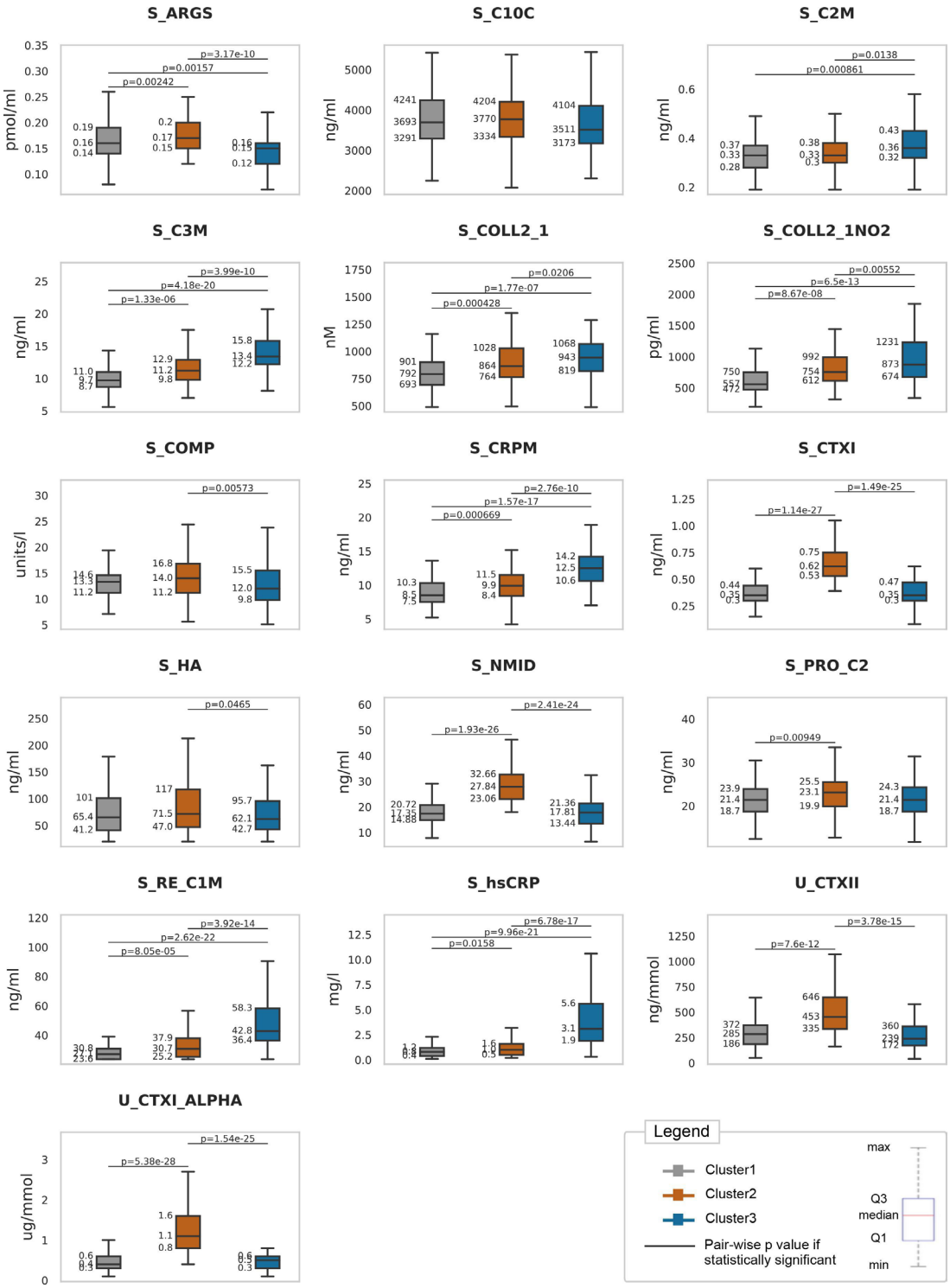

Figure 5 Comparison of biochemical markers' distributions in each cluster, and the statistical relevance of differences between them.

progressors and combined pain and structure progressors in the inflammation cluster (C3). Similarly, there were more structure-related progressors in the structural damage cluster (C2). The highest relative number of non-progressive patients was found in the low tissue turnover cluster (C1) and the lowest in the inflammation cluster (C3).

DISCUSSION

The aim of this work was to test if ML techniques can be used to identify biologically meaningful subgroups of patients with OA in the APPROACH cohort based on selected biochemical markers. By using clustering, that is, an unsupervised ML approach that does not exploit domain knowledge, we were able to identify molecular endotypes from 16 well-defined biochemical markers reflecting different molecular pathways and ongoing pathophysiological processes. We discovered three distinct OA phenotypes associated with the clusters (endotypes): C1—a low tissue turnover phenotype, C2—a structural damage phenotype and C3—systemic inflammation phenotype. The clustering reflects well the current biological and mechanistic understanding of

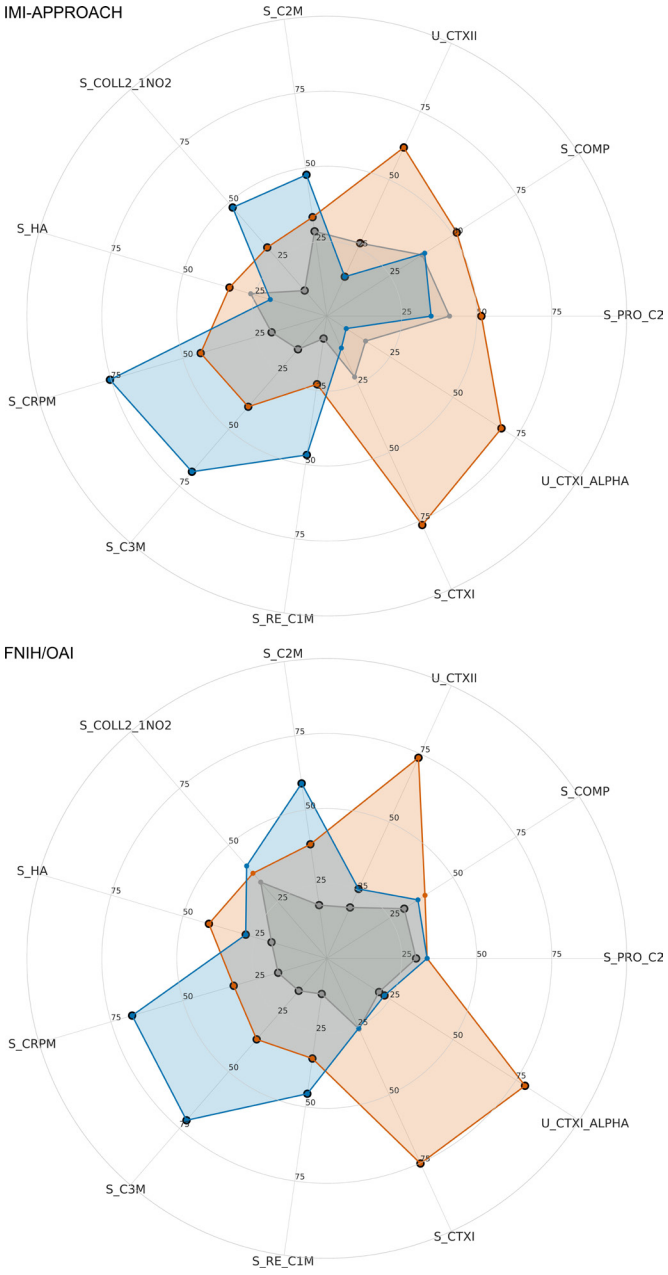

Figure 6 Radar plots comparing clusters found in the IMI-APPROACH and FNIH/OAI cohorts, using common subset of biomarkers. The median biomarker concentration for each cluster is shown. When the difference between the medians is statistically significant, it is marked with a circle (instead of a dot).

the respective biomarkers, in that distinct patterns could be identified for the subtypes. In particular, the combination of different markers describes the underlying biology in the clusters. This result is in line with published results from the FNIH/OAI biomarker initiative,[29 33] and the progression status of the members of each cluster is consistent with the cluster interpretation provided above: C1 has the highest proportion of non-progressors, C2 has the highest proportion of structural progressors and C3 has the highest proportion of pain-related progressors, and those progressing both in pain and in structure. However, although the proportions varied (ranging from 43% to 56%) progressive patients were found in all clusters. This means that the clusters represent different disease subtypes, within which the progression may occur.

Table 2 Distribution of progressive IMI-APPROACH patients across clusters.

Cluster (members)	No progression	Only pain	Only structure	Both
C1 (69)	39 (57%)	20 (29%)	7 (10%)	3 (4%)
C2 (84)	45 (54%)	21 (25%)	16 (19%)	2 (2%)
C3 (68)	30 (44%)	25 (37%)	7 (10%)	6 (9%)

Putting this in context of the work conducted on markers in clinical interventional trials, a few things can be learnt. Oral salmon calcitonin was tested as an antiresorptive treatment for OA. The phase III clinical trials failed to meet their clinical endpoints. Interestingly, calcitonin did significantly modulate CTX-I and CTX-II.[34] There are likely several reasons why this study failed, however, it begs to wonder what would the outcome have been if the study was enriched for C2 patients? Another failed trial was testing the efficacy of the IL-1 monoclonal antibody in OA and found markers from C3 modulated by treatment.[35] Would it still fail if it was enriched for C3 patients?

Despite a large and growing disease burden in OA, many pharmaceutical companies have de-emphasized or even abandoned OA drug development due to perceived hurdles. Crucial in this is the lack of appropriate predictive and outcome measures that can robustly identify patients early in the disease, which may benefit from a specific therapy. The lack of specific and sensitive baseline characteristics and subsequent endpoints to differentiate between responders and non-responders, both at the level of pain and tissue structure modification (ie, DMOAD), has led to trials that included hundreds of patients in each arm with at least 3-year follow-up. Despite these enormous trials, European Medicines Agency and Food and Drug Administration have not approved any DMOAD yet.[36] There is a general lack of understanding of OA pathogenesis which appears rather variable and likely reflects different phenotypes with fundamental differences in disease aetiology, tissue alterations, clinical manifestations (pain/mobility) and disease progression. Although the current mindset for drug treatment in the field is moving to a more personalised medicine and patient stratification approach, there are no accepted methods or guidelines to classify patients with OA, for example, to predict the underlying pathophysiology, to select patients according to their prognosis or to differentiate between patients in terms of diagnosis methodology and treatment plan. However, several initiatives have been initiated to generate more focus on the development of projects for identifying endotypes. For example, a framework for conducting and reporting phenotyping research was provided[37]—this may very well be the first step toward integrating the concept of phenotyping in research.

A better understanding of disease stratification and acceptance of a guideline to classify patients with OA will provide clear phenotype-directed protocols for DMOAD trials that enable us to target subgroups with OA that have uniform disease characteristics, thereby increasing the chances of success. We propose that the biomarker clustering analysis performed herein can be used to stratify patients with OA into groups with distinct molecular endotypes. This approach could potentially drive OA clinical trials stratification and serve as the basis for precision medicine strategies for OA progression in the future. Although there are limited data publicly available, there have been a few attempts to identify multimarker endotypes in OA. Sonh *et al* showed that several cytokines were elevated in synovial fluid and serum of patients with OA compared with normal samples when looking at an average level; however, it was also obvious

that the pattern was very heterogeneous.[38] Werdyani *et al* identified three distinct endotypes using metabolomics.[39] One of those clusters showed some association with muscle weakness. These data suggest that a subset of patients could belong to an inflammatory endotype.

Moreover, we focused on biochemical markers measured at the baseline of the study, and not their longitudinal changes, as this analysis would be more useful to inform future clinical trials. Longitudinal monitoring of biomarkers can give insight in the pharmacodynamic effects or provide early proof of effectiveness of a compound in interventional clinical trials, however often fail to predict progression in the study population in these trials.[34 40–42] Therefore, although longitudinal monitoring of individual biomarkers are only modestly predictive (if at all) of knee OA progression, they might have some utility as patient stratification like described herein for enriching OA trials for progressors.[29]

As more longitudinal data of the IMI-APPROACH cohort becomes available (currently an ongoing process), future investigations could explore the longitudinal data on biomarkers, imaging and other markers in IMI-APPROACH to further refine the description of the phenotypes and possibly explore more detailed stratifications. This analysis could take many different directions, for example, analyse cluster membership differences between visits or on comparison of the entire patient trajectories over 2 years of the study.

The main limitations of this work were the small numbers of patients in the IMI-APPROACH cohort and being able to perform only a partial validation with an external cohort, limited to a common subset of biomarkers. It would be beneficial for the field if future biomarker studies use a superset of the FNIH/OAI and IMI-APPROACH biomarkers, to allow for a complete validation of the discovered clusters. The use of predefined set of biochemical markers limits the discovery potential to certain molecular mechanisms. This could be avoided if clustering was performed on data generated by an untargeted platform (eg, RNA-seq); however, the analysis of such high-dimensional data is often much less robust, especially on small sample sizes. Finally, more research should be conducted on more abundant cohorts to fully evaluate the gender bias in clustering analysis of OA-related biochemical markers. From our analysis in the IMI-APPROACH and FNIH/OAI cohorts, we believe it is advisable for future studies to consider male and female patients separately and possibly draw conclusions that are gender based, if sample sizes are large enough.

Author affiliations
[1]School of Computing, Newcastle University, Newcastle upon Tyne, UK
[2]Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland
[3]Department of Regenerative Medicine, State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania
[4]Rheumatology & Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands
[5]Department of Joint Surgery, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, People's Republic of China
[6]World Health Organization Collaborating Centre for Public Health Aspects of Musculoskeletal Health and Aging, Liege, Belgium
[7]ImmunoScience, Nordic Bioscience, Herlev, Denmark
[8]Faculty of Medicine, Department of Clinical Sciences Lund, Orthopaedics, Lund University, Lund, Sweden
[9]Artialis SA, Liège, Belgium
[10]Center for Interdisciplinary Research on Medicines (CIRM), University of Liège, Liège, Belgium
[11]Rheumatology, Leiden Universitair Medisch Centrum, Leiden, The Netherlands
[12]Department of Clinical Epidemiology, Leiden Universitair Medisch Centrum, Leiden, The Netherlands
[13]Servicio de Reumatologia, INIBIC-Hospital Universitario A Coruña, A Coruña, Spain
[14]Division of Rheumatology and Research, Diakonhjemmet Hospital, Oslo, Norway
[15]Institut national de la santé et de la recherche médicale, Sorbonne Université, Paris, France
[16]BioBone BV, Darmstadt, Germany
[17]GlaxoSmithKline USA, Philadelphia, Pennsylvania, USA

Twitter Francis Berenbaum @larhumato and Jaume Bacardit @jaumebp

Acknowledgements We thank Agnes Lalande (Institut de Recherches Internationales Servier, Suresnes, France) for clinical study coordination, Thomas Lohr (GSK) for sample handling logistics and Janneke Boere and colleagues at Lygature for coordinating the research activity.

Contributors FA, PW, JBa contributed to the conception and design of the study, AS, MU, YH, ACM, MK, FJB, IKH, FB and ACBJ contributed to the acquisition of data. AM, JBl, CL, JL and ACB-J contributed to the analysis and interpretation of data. JBa is the paper's guarantor. All authors contributed to drafting the article or revising it critically for important intellectual content. All authors gave final approval of the version to be submitted.

Funding The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement no 115770, resources of which are composed of financial contributions from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution. See http://www.imi.europa.eu/ and http://wwwapproachprojecteu/.

Competing interests ACB-J is a full-time employee and shareholder of Nordic Bioscience, a privately owned company involved in the development and commercialisation of biomarkers for fibroinflammatory disorders. YH is the founder and president of Artialis, and MU is a full-time employéé of Artialis, a spin-off company of the University of Liège. YH has also received fees from Tilman, Genequine, Seikagaku, Expanscience, Nestlé, Immubio, Biose and Labhra. CL was an employee of Merck at project start. IKH consults for Abbvie and Novartis and has received funding from Pfizer. JL is employed by and shareholder in GlaxoSmithKline. FB reports personal fees from AstraZeneca, Boehringer, Bone Therapeutics, CellProthera, Expanscience, Galapagos, Gilead, Grunenthal, GSK, Eli Lilly, Merck Sereno, MSD, Nordic, Nordic Bioscience, Novartis, Pfizer, Roche, Sandoz, Sanofi, Servier, UCB, Peptinov, 4P Pharma, 4Moving Biotech and grants from TRB Chemedica, outside the submitted work. FJB reports funding from Gedeon Richter, Bristol-Myers Squibb, Sun Pharma Global FZE, Celgene, Janssen Cilag, Janssen Research & Development, Viela Bio, Astrazeneca, UCB BIOSCIENCES, UCB BIOPHARMA SPRL, AbbVie Deutschland, Merck, Amgen, Novartis Farmacéutica, Boehringer Ingelheim España, CSL Behring, Glaxosmithkline Research & Development, Pfizer, Lilly, Corbus Pharmaceuticals, Biohope Scientific Solutions for Human Health, Centrexion Therapeutics, Sanofi, TEDEC-MEIJI FARMA, Kiniksa Pharmaceuticals, Fundación para la Investigación Biomédica Del Hospital Clínico San Carlos, Grünenthal and Galapagos. MK receives consulting fees from Abbvie, Pfizer, Levicept, GlaxoSmithKline, Merck-Serono, Kiniksa, Flexion, Galapagos, Jansen, CHDR, Novartis, UCB. AM receives fees/funding from Merck, Kolon TissueGene, Pfizer, Galapagos-Servier, Image Analysis Group (IAG), Artialis, Aché Laboratórios Farmacêuticos, AbbVie, Guidepoint Global, Alphasights, Science Branding Communications, GSK, Flexion Therapeutics, Pacira Biosciences, Sterifarma, Bioiberica, SANOFI, Genacol, Kolon Life Science, BRASIT/BRASOS, GEOS, MCI Group, Alcimed, Abbot, Laboratoires Expansciences, SPRIM Communications, Frontiers Media and University Health Network Toronto.

Patient consent for publication Not required.

Ethics approval This study involves human participants. The study is being conducted in compliance with the protocol, Good Clinical Practice (GCP), the Declaration of Helsinki, and applicable ethical and legal regulatory requirements (for all countries involved), and is registered under ClinicalTrials.gov: NCT03883568. All participants have received oral and written information and provided written informed consent. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data are available on request to the APPROACH Steering Committee.

ORCID iDs
Federico Angelini http://orcid.org/0000-0001-8333-656X
Paweł Widera http://orcid.org/0000-0003-4955-36530
Ali Mobasheri http://orcid.org/0000-0001-6261-1286
Joseph Blair http://orcid.org/0000-0003-3248-7039
André Struglics http://orcid.org/0000-0003-4289-1393
Melanie Uebelhoer http://orcid.org/0000-0002-7911-3205
Yves Henrotin http://orcid.org/0000-0003-1073-449X
Margreet Kloppenburg http://orcid.org/0000-0002-9294-2307
Francisco J Blanco http://orcid.org/0000-0001-9821-7635
Ida K Haugen http://orcid.org/0000-0001-7810-2216
Francis Berenbaum http://orcid.org/0000-0001-8252-7815
Christoph Ladel http://orcid.org/0000-0001-8657-6219
Jonathan Larkin http://orcid.org/0000-0002-1202-9287
Anne C Bay-Jensen http://orcid.org/0000-0001-7952-9297
Jaume Bacardit http://orcid.org/0000-0002-2692-7205

REFERENCES

1 Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. *Lancet* 2020;396:1711–2.
2 Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet* 2019;393:1745–59.
3 Martel-Pelletier J, Barr AJ, Cicuttini FM, *et al*. Osteoarthritis. *Nat Rev Dis Primers* 2016;2:16072.
4 Cross M, Smith E, Hoy D, *et al*. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* 2014;73:1323–30.
5 Safiri S, Kolahi A-A, Smith E, *et al*. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the global burden of disease study 2017. *Ann Rheum Dis* 2020;79:819–28.
6 Zhang W, Ouyang H, Dass CR, *et al*. Current research on pharmacologic and regenerative therapies for osteoarthritis. *Bone Res* 2016;4:15040.
7 Ghouri A, Conaghan PG. Prospects for therapies in osteoarthritis. *Calcif Tissue Int* 2021;109:339–50.
8 Deveza LA, Loeser RF. Is osteoarthritis one disease or a collection of many? *Rheumatology* 2018;57:iv34–42.
9 Deveza LA, Nelson AE, Loeser RF. Phenotypes of osteoarthritis: current state and future implications. *Clin Exp Rheumatol* 2019;37 Suppl 120:64–72.
10 Karsdal MA, Christiansen C, Ladel C, *et al*. Osteoarthritis-a case for personalized health care? *Osteoarthritis Cartilage* 2014;22:7–16.
11 Driban JB, Sitler MR, Barbe MF, *et al*. Is osteoarthritis a heterogeneous disease that can be stratified into subsets? *Clin Rheumatol* 2010;29:123–31.
12 Luo Y, Samuels J, Krasnokutsky S, *et al*. A low cartilage formation and repair endotype predicts radiographic progression of symptomatic knee osteoarthritis. *J Orthop Traumatol* 2021;22:10.
13 Ching K, Houard X, Berenbaum F. Hypertension meets osteoarthritis — revisiting the vascular aetiology hypothesis. *Nat Rev Rheumatol* 2021:1–17.
14 Dell'Isola A, Allan R, Smith SL, *et al*. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. *BMC Musculoskelet Disord* 2016;17:425.
15 Dell'Isola A, Steultjens M. Classification of patients with knee osteoarthritis in clinical phenotypes: data from the osteoarthritis initiative. *PLoS One* 2018;13:e0191045.
16 Mobasheri A, Saarakkala S, Finnilä M, *et al*. Recent advances in understanding the phenotypes of osteoarthritis. *F1000Res* 2019;8. doi:10.12688/f1000research.20575.1. [Epub ahead of print: 12 Dec 2019].
17 Mobasheri A, van Spil WE, Budd E, *et al*. Molecular taxonomy of osteoarthritis for patient stratification, disease management and drug development: biochemical markers associated with emerging clinical phenotypes and molecular endotypes. *Curr Opin Rheumatol* 2019;31:80–9.
18 Thudium CS, Nielsen SH, Sardar S, *et al*. Bone phenotypes in rheumatology - there is more to bone than just bone. *BMC Musculoskelet Disord* 2020;21:789.
19 Kim Y, Levin G, Nikolov NP, *et al*. Concept endpoints informing design considerations for confirmatory clinical trials in osteoarthritis. *Arthritis Care Res* 2020. doi:10.1002/acr.24549. [Epub ahead of print: 20 Dec 2020].
20 Bonakdari H, Jamshidi A, Pelletier J-P, *et al*. A warning machine learning algorithm for early knee osteoarthritis structural progressor patient screening. *Ther Adv Musculoskelet Dis* 2021;13:1759720X21993254.
21 Tiulpin A, Thevenot J, Rahtu E, *et al*. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep Learning-Based approach. *Sci Rep* 2018;8:1727.
22 Tiulpin A, Klein S, Bierma-Zeinstra SMA, *et al*. Multimodal machine Learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. *Sci Rep* 2019;9:20038.
23 van Helvoort EM, van Spil WE, Jansen MP, *et al*. Cohort profile: the applied public-private research enabling osteoarthritis clinical Headway (IMI-APPROACH) study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical markers. *BMJ Open* 2020;10:e035101.
24 Widera P, Welsing PMJ, Ladel C, *et al*. Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. *Sci Rep* 2020;10:8427.
25 Christian H, Nielsen AB, Thorsen-Meyer H-C. Survival prediction in intensive-care units based on aggregation of long-term disease history and acute physiology a retrospective study of the Danish national patient registry and electronic patient records articles survival prediction in intensive-care. *Lancet Digit Heal* 2019;1.
26 Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12.
27 Lundberg SM, Erion G, Chen H, *et al*. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2:2522–5839.
28 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
29 Kraus VB, Collins JE, Hargrove D, *et al*. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH oa biomarkers Consortium. *Ann Rheum Dis* 2017;76:186–95.
30 Rudzki PJ, Biecek P, Kaza M. Incurred sample reanalysis: time to change the sample size calculation? *AAPS J* 2019;21:28.
31 Fluhler E, Vazvaei F, Singhal P, *et al*. Repeat analysis and incurred sample reanalysis: recommendation for best practices and harmonization from the global bioanalysis Consortium harmonization team. *Aaps J* 2014;16:1167–74.
32 McInnes L, Healy J, Melville J. Umap: uniform manifold approximation and projection for dimension reduction. *arXiv* 2018:1802.03426.
33 Kraus VB, Karsdal MA. Osteoarthritis: current molecular biomarkers and the way forward. *Calcif Tissue Int* 2021;109:1–10.
34 Karsdal MA, Byrjalsen I, Alexandersen P, *et al*. Treatment of symptomatic knee osteoarthritis with oral salmon calcitonin: results from two phase 3 trials. *Osteoarthr Cartil* 2015;23:532–43.
35 Wang SX, Abramson SB, Attur M, *et al*. Safety, tolerability, and pharmacodynamics of an anti-interleukin-1α/β dual variable domain immunoglobulin in patients with osteoarthritis of the knee: a randomized phase 1 study. *Osteoarthritis Cartilage* 2017;25:1952–61.
36 Oo WM, Little C, Duong V, *et al*. The development of disease-modifying therapies for osteoarthritis (DMOADs): the evidence to date. *Drug Des Devel Ther* 2021;15:2921–45.
37 van Spil WE, Bierma-Zeinstra SMA, Deveza LA, *et al*. A consensus-based framework for conducting and reporting osteoarthritis phenotype research. *Arthritis Res Ther* 2020;22:54.
38 Sohn DH, Sokolove J, Sharpe O, *et al*. Plasma proteins present in osteoarthritic synovial fluid can stimulate cytokine production via Toll-like receptor 4. *Arthritis Res Ther* 2012;14:R7.
39 Werdyani S, Liu M, Zhang H, *et al*. Endotypes of primary osteoarthritis identified by plasma metabolomics analysis. *Rheumatology* 2021;60:2735–44.
40 Bay-Jensen AC, Mobasheri A, Thudium CS, *et al*. Blood and urine biomarkers in osteoarthritis - an update on cartilage associated type II collagen and aggrecan markers. *Curr Opin Rheumatol* 2022;34:54–60.
41 Bay-Jensen AC, Manginelli AA, Karsdal M, *et al*. Low levels of type II collagen formation (PRO-C2) are associated with response to sprifermin: a pre-defined, exploratory biomarker analysis from the forward study. *Osteoarthritis Cartilage* 2022;30:92–9.
42 van der Aar E, Deckx H, Dupont S, *et al*. Safety, pharmacokinetics, and pharmacodynamics of the ADAMTS-5 inhibitor GLPG1972/S201086 in healthy volunteers and participants with osteoarthritis of the knee or hip. *Clin Pharmacol Drug Dev* 2022;11:112–22.
43 Loeser RF, Beavers DP, Bay-Jensen AC, *et al*. Effects of dietary weight loss with and without exercise on interstitial matrix turnover and tissue inflammation biomarkers in adults with knee osteoarthritis: the intensive diet and exercise for arthritis trial (idea). *Osteoarthritis Cartilage* 2017;25:1822–8.
44 Bay-Jensen A-C, Bihlet A, Byrjalsen I, *et al*. Serum C-reactive protein metabolite (CRPM) is associated with incidence of contralateral knee osteoarthritis. *Sci Rep* 2021;11:6583.
45 Larsson S, Lohmander LS, Struglics A. An ARGS-aggrecan assay for analysis in blood and synovial fluid. *Osteoarthritis Cartilage* 2014;22:242–9.
46 He Y, Manon-Jensen T, Arendt-Nielsen L, *et al*. Potential diagnostic value of a type X collagen neo-epitope biomarker for knee osteoarthritis. *Osteoarthritis Cartilage* 2019;27:611–20.
47 Bay-Jensen A-C, Liu Q, Byrjalsen I, *et al*. Enzyme-linked immunosorbent assay (ELISAs) for metalloproteinase derived type II collagen neoepitope, CIIM--increased serum CIIM in subjects with severe radiographic osteoarthritis. *Clin Biochem* 2011;44:423–9.
48 Mobasheri A, Lambert C, Henrotin Y. Coll2-1 and Coll2-1NO2 as exemplars of collagen extracellular matrix turnover - biomarkers to facilitate the treatment of osteoarthritis? *Expert Rev Mol Diagn* 2019;19:803–12.
49 Kluzek S, Bay-Jensen A-C, Judge A, *et al*. Serum cartilage oligomeric matrix protein and development of radiographic and painful knee osteoarthritis. A community-based cohort of middle-aged women. *Biomarkers* 2015;20:557–64.
50 Kraus VB, Collins JE, Charles HC, *et al*. Predictive validity of radiographic trabecular bone texture in knee osteoarthritis: the osteoarthritis research Society International/Foundation for the National Institutes of health osteoarthritis biomarkers Consortium. *Arthritis Rheumatol* 2018;70:80–7.

51 Pearson TA, Mensah GA, Alexander RW, *et al*. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 2003;107:499–511.
52 Ravn P, Clemmesen B, Christiansen C. Biochemical markers can predict the response in bone mass during alendronate treatment in early postmenopausal women. Alendronate Osteoporosis Prevention Study Group. *Bone* 1999;24:237–44.
53 Radojčić MR, Thudium CS, Henriksen K, *et al*. Biomarker of extracellular matrix remodelling C1M and proinflammatory cytokine interleukin 6 are related to synovitis and pain in end-stage knee osteoarthritis patients. *Pain* 2017;158:1254–63.
54 Huebner JL, Bay-Jensen AC, Huffman KM, *et al*. Alpha C-telopeptide of type I collagen is associated with subchondral bone turnover and predicts progression of joint space narrowing and osteophytes in osteoarthritis. *Arthritis Rheumatol* 2014;66:2440–9.
55 Taylor J, Dekker S, Jurg D, *et al*. Making the patient voice heard in a research consortium: experiences from an EU project (IMI-APPROACH). *Res Involv Engagem* 2021;7:24.