Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

A machine learning approach to knee osteoarthritis phenotyping: data from the FNIH Biomarkers Consortium

A.E. Nelson † ‡ *, F. Fang §, L. Arbeeva †, R.J. Cleveland † ‡, T.A. Schwartz † ||, L.F. Callahan † ‡, J.S. Marron §, R.F. Loeser † ‡

† Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
‡ Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
§ Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
|| Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

ARTICLE INFO

SUMMARY

Objective: Knee osteoarthritis (KOA) is a heterogeneous condition representing a variety of potentially distinct phenotypes. The purpose of this study was to apply innovative machine learning approaches to KOA phenotyping in order to define progression phenotypes that are potentially more responsive to interventions.

Design: We used publicly available data from the Foundation for the National Institutes of Health (FNIH) osteoarthritis (OA) Biomarkers Consortium, where radiographic (medial joint space narrowing of ≥0.7 mm), and pain progression (increase of ≥9 Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC] points) were defined at 48 months, as four mutually exclusive outcome groups (none, both, pain only, radiographic only), along with an extensive set of covariates. We applied distance weighted discrimination (DWD), direction-projection-permutation (DiProPerm) testing, and clustering methods to focus on the contrast (z-scores) between those progressing by both criteria ("progressors") and those progressing by neither ("non-progressors").

Results: Using all observations (597 individuals, 59% women, mean age 62 years and BMI 31 kg/m^2) and all 73 baseline variables available in the dataset, there was a clear separation among progressors and non-progressors ($z = 10.1$). Higher z-scores were seen for the magnetic resonance imaging (MRI)-based variables than for demographic/clinical variables or biochemical markers. Baseline variables with the greatest contribution to non-progression at 48 months included WOMAC pain, lateral meniscal extrusion, and serum N-terminal pro-peptide of collagen IIA (PIIANP), while those contributing to progression included bone marrow lesions, osteophytes, medial meniscal extrusion, and urine C-terminal crosslinked telopeptide type II collagen (CTX-II).

Conclusions: Using methods that provide a way to assess numerous variables of different types and scalings simultaneously in relation to an outcome of interest enabled a data-driven approach that identified key variables associated with a progression phenotype.

© 2019 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Knee osteoarthritis (KOA) is a heterogeneous condition characterized by changes in a variety of joint tissues and driven by a number of different potential mechanisms[1]. A variety of diverse risk factors, such as aging, body weight, joint injury, genetics, and biomechanical factors, contribute to disease, sometimes alone but more often in combination. Such common osteoarthritis (OA) risk factors may lead to a different mechanistic pathway to OA disease, such that the key mediators promoting OA development or progression in older adults may be quite different than those that contribute to development of post-traumatic OA. There are still no effective disease modifying drugs for KOA, likely due in part to the fact that clinical trials have treated all KOA, regardless of etiology or risk factors, as the same disease.

* Address correspondence and reprint requests to: A. E. Nelson, University of North Carolina at Chapel Hill, 3300 Doc J. Thurston Building, Campus Box #7280, Chapel Hill, NC 27599-7280, USA. Tel.: 1-919-966-0553; Fax: 1-919-966-1739.

E-mail addresses: aenelson@med.unc.edu (A.E. Nelson), fangf@live.unc.edu (F. Fang), liubov_arbeeva@med.unc.edu (L. Arbeeva), Rebecca_cleveland@med.unc.edu (R.J. Cleveland), tschwart@email.unc.edu (T.A. Schwartz), leigh_callahan@med.unc.edu (L.F. Callahan), marron@unc.edu (J.S. Marron), Richard_loeser@med.unc.edu (R.F. Loeser).

The variations in observable characteristics of individuals, resulting from genetic and environmental factors, constitute a phenotype[1]. Understanding underlying phenotypes of KOA that represent different pathways to disease could lead to new treatments for this common and debilitating condition. To date, most phenotypes have been postulated based on our current understanding of KOA, such as those related to inflammation, metabolic disturbances, and biomechanical stresses. Two recent systematic reviews have approached the question of OA phenotypes in somewhat different ways. Dell'Isola *et al.* identified six phenotypes from 24 published studies, including pain sensitization, inflammation, metabolic syndrome, bone/cartilage metabolism, malalignment, and minimal joint disease[2]. They then applied these phenotypes to the Foundation for the National Institutes of Health (FNIH) Biomarkers Consortium dataset, classifying the majority into at least one phenotype, however, nearly 1 in 5 did not meet criteria for any of the postulated phenotypes[3]. Deveza *et al.* focused instead on important phenotypic characteristics in their review of 34 studies, finding a number of features to be associated with clinical vs structural phenotypes[4].

Data-driven approaches are also of great interest as these do not require an *a priori* hypothesis and are therefore able to identify unanticipated patterns in the data that may reveal new subgroups. Some of these methods, such as clustering and latent class analysis, are widely used but only recently applied to KOA phenotyping. Others, such as those employed here, are machine learning based and not previously applied to these types of questions, and so offer the potential to provide new insights into this complex problem.

The purpose of this study was to apply innovative machine learning approaches, specialized for a high dimension low sample size setting (e.g., many measurements on a relatively limited sample size), to phenotyping in KOA in order to better define progression phenotypes that may be more homogeneous and responsive to potential disease modifying interventions.

Methods

This analysis used publicly available data from the FNIH OA Biomarkers Consortium (available at https://nda.nih.gov/oai). Details of the study design have been published previously[5]. The published dataset includes 600 individuals with 76 demographic, imaging (quantitative and semi-quantitative magnetic resonance imaging [MRI]), and biochemical variables. Details regarding the acquisition of the imaging and other biomarkers are also published[6–9]. This dataset includes four mutually exclusive outcome groups (one knee per person) defined as: 1) neither X-ray nor pain progression ($n = 200$); 2) pain progression only ($n = 103$); 3) X-ray progression only ($n = 103$); 4) both X-ray and pain progression ($n = 194$). X-ray progression, as defined by the FNIH OA Biomarkers Consortium, was based on medial joint space loss (loss in joint space width of at least 0.7 mm), while pain progression was based on an increase in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score (persistent increase of nine or more points), at 48 months. For each eligible subject, the FNIH OA Biomarkers Consortium randomly selected one knee as the index knee, and these were frequency-matched for Kellgren–Lawrence (KL) grade and BMI category[10]. As a first step in the current analysis, problematic variables ($n = 3$: unique participant identifier [non-informative], posterior cruciate ligament repair and tear [both frequently missing]) and observations ($n = 3$ due to missing or incorrect data) were removed, leaving 597 observations and 73 variables for the remainder of our analysis.

A supervised approach (i.e., contrasting known or specified classes) was employed to identify features associated with existing phenotypes based on progression status. For this purpose, we focused on contrasting "progressors," ($n = 192$) defined as knees that progressed by both X-ray and pain, vs "non-progressors," ($n = 200$) those that did not progress by either. Our hypothesis was that a set of predictive measures would be able to identify significant differences between those who did and did not progress in this cohort. We considered all predictive measures together as well as in groups (e.g., demographic, imaging, and biochemical markers). Additional hypotheses considering other comparisons (e.g., X-ray only or pain only progressors) were also considered in an exploratory way (see Supplemental Table (A)). The data were transformed to reduce skewness and standardized to address differences in scale, prior to application of the machine learning methodology described below (code for these methods is available at http://marronwebfiles.sites.oasis.unc.edu/Matlab7Software/).

Distance weighted discrimination (DWD) is a linear discriminant analysis method allowing maximal separation of data points by class[11], and utilized in our prior publication on hip shape in OA[12]. DWD is particularly suited to cases where the dimension of the data vector exceeds the number of samples (i.e., a large number of measurements relative to the sample size). Once defined, the difference between two distributions obtained using DWD can be tested for statistical significance using the Direction-Projection-Permutation (DiProPerm) test[13]. DiProPerm ensures statistical specificity of the hypothesis test for two previously defined populations (e.g., progressors vs non-progressors) by first finding a separating direction (e.g. DWD), then projecting the data and using a one dimensional summary of the separation (e.g. the difference of the means as described in Ref. 13). Statistical significance is obtained by a permutation approach (using 100 permutations), where the class labels are randomly shuffled and the DWD direction, projections, and the test statistic, are recomputed, giving a null distribution whose quantiles are used to compute *P*-values. Since this method treats the overall vector of features as a single data object, there is less of a need for adjustment for multiple comparisons.

As a next step, we examined loadings of the variables on the DWD direction. These loadings demonstrate the relative contribution of each variable to the class difference (e.g. progressors compared with non-progressors). Additionally, we explored k-means clustering to partition the observations into subgroups, with the number of subgroups based on published statistical indices[14–17]. A z-score of at least 3 (corresponding to a *P*-value of ~0.001) was considered statistically significant.

An alternate visualization technique, t-distributed stochastic neighbor embedding (t-SNE), was also applied[18]. This unsupervised machine learning algorithm is designed to embed high-dimensional data into a two to three-dimensional space to allow improved visual clustering, through first constructing a probability distribution in high dimensional space and then mapping this to a parallel distribution in low-dimensional space (while minimizing the relative entropy). We considered a range of perplexity tuning parameters (which control the visual impression of clustering), from 1 to 100, and 5,000 iterations (in the t-SNE optimization). Because t-SNE visually distinguishes clusters but hides differing relative distances between those clusters, DiProPerm (described above) was used to quantify the difference between each pair of t-SNE clusters. Principal components analysis (PCA) was applied, using colors based on clusters obtained from t-SNE, in order to visualize the contribution of the individual variables to these different modes of variation. A bar plot of the PCA loadings (entries of the eigenvector) was used to demonstrate which of the variables were the most important factors in a given component, where the height of the bar is representative of the loading, or importance, of each variable.

Results

The overall FNIH cohort study characteristics have been published[5,10]. The final dataset for the present analysis included 597 participants, 59% women, with a mean age of 62 years and mean BMI of 31 kg/m^2.

When considering all observations and all variables simultaneously, the progressors and non-progressors clearly separated with a z-score of 10.10 (Table I, Fig. 1). We explored k-means clustering to partition the observations into two subgroups (the number most supported by the noted statistical indices). No statistical improvement was seen with partitioning the observations into 2 clusters using k-means, so we did not further investigate the potential drivers of these clusters. For example, using all variables, there were significant z-scores for both clusters (6.19 and 4.77) but neither was more significant than the overall z-score for all observations (10.10). We were also able to study the four sets of variables separately, finding higher z-scores for the MRI variables (11.62 for quantitative and 10.28 for semi-quantitative), with lower scores for demographic and biochemical markers (1.47 and 2.43, respectively). Again, lower z-scores were seen for the 2 clusters compared with using all observations. Exploratory results of contrasts other than progressors vs non-progressors are shown in Supplemental Table (A); the strongest results were seen for all observations rather than the 2-cluster scenario for all hypotheses with significant z-scores. The most significant results were seen for the progressor vs non-progressor contrast discussed here and for the contrast between those who did and did not progress by radiographic criteria (regardless of pain, z-score = 11.86 using all observations).

We were specifically interested in the relative contributions of each variable to the overall difference between progressors and non-progressors, that is, which baseline variables were most important in separating these two classes. Figure 2 shows the DWD loadings, represented in a bar plot for the 40 variables with the greatest contribution to the observed class difference. The greatest contribution to non-progression (below the null) was seen for greater WOMAC pain, lateral meniscal extrusion, and serum N-terminal pro-peptide of collagen IIA (PIIANP). The largest contributors to progression (above the null) were from the number of subregions with bone marrow lesions, the number of locations with any osteophyte, medial meniscal extrusion, and urine C-terminal crosslinked telopeptide type II collagen (CTX-II, all greater). Loadings of the 40 variables with the greatest contribution when the observations were divided into 2 clusters are shown in the Supplement (Supplemental Figures (A) and (B)). For Cluster 1 (Supplemental Figure (A)), the variables with the greatest contribution were similar to those observed using the overall data (Fig. 2), while for Cluster 2 (Supplemental Figure (B)), there was a greater contribution to progression (above the null) from osteophyte number and pain medication use, but a smaller contribution from urine biomarkers and bone marrow lesions. Cluster two also demonstrated a greater contribution to non-progression (below the null) from the serum biomarkers cartilage oligomeric matrix protein (COMP) and crosslinked N-telopeptide of type I collagen (NTX-1).

Using the t-SNE approach, four clusters were identified within the dataset. Figure 3 shows the results using a perplexity parameter of 20. On the left (**A**) is the t-SNE visualization with the points colored based solely on the visual clusters. The right side of figure (**B**), enables interpretation of the t-SNE clusters by associating the colors with outcome and sex (chosen given its large contribution to PC1), with the symbols identified in the legend. This demonstrates that the clustering is primarily based on outcome and sex: the red cluster is all men who did not progress, the green cluster is all women who did not progress. The dark and light blue clusters represent progressors but are not as clearly divided by sex. The dark blue cluster primarily represents male progressors and the light blue cluster female progressors, but there are some exceptions [Fig. 3(B)].

Next, because t-SNE tends to hide relative distances between clusters, the DiProPerm test was applied to quantify the difference between the visual clusters obtained using t-SNE (Fig. 4). The results of DiProPerm testing between clusters demonstrates that the difference between progressors and non-progressors within the same sex (e.g., male on the left side of the figure in red and purple, z-score = 4.03, and female on the right side of the figure in green and cyan, z-score = 6.54) is much less than the difference between men (red and purple clusters) and women (cyan and green clusters), regardless of progression status (z-scores = 38.22–47.18, Fig. 4).

Additionally, we used PCA to visualize the driving factors of differences among the four clusters identified by t-SNE (Supplemental Figure (C)). Nearly 50% of the total variation is seen in the first four principal components. To understand which variables are the dominant factors driving the variation in each principal component, we can assess the loadings as done above (Fig. 2) in the DWD direction. Thus, in Fig. 5, the loadings for principal component one demonstrate the largest contribution from female sex (as expected from the large sex difference noted above), but also from several of the quantitative MRI variables related to femoral and tibial cartilage volume, meniscal volume, and subchondral bone area of the medial and lateral femur, patella, and tibia. For principal component 2 (Supplemental Figure (D)), the variables which contribute most to the differences between clusters are those related to bone, including the number and maximal bone marrow lesion score and the number and maximum osteophyte score, as well as several of the cartilage morphology assessments from semi-quantitative scoring, such as cartilage thickness and surface scores. For principal component 3 (Supplemental Figure (E)), the largest contributions are from use of pain medication and several biochemical markers: serum C-terminal crosslinked telopeptide of type I collagen (CTX-I) and NTX-1, urine CTX-II and NTX-I, and urine CTX-I alpha and beta. Principal component four was essentially driven by side (right vs left, data not shown).

Discussion

Through application of novel machine learning methodologies to the FNIH cohort study data, we were able to simultaneously utilize all of the data, including MRI assessments, demographic and clinical variables, and biochemical markers using a single statistical

Table I
Z-scores derived from the DiProPerm test for the difference between those progressing by both radiographic and pain criteria vs non-progressors on either

		All variables	Demographic variables	Quantitative MRI variables	Semi-Quantitative MRI variables	Biochemical variables
All observations		**10.10**	1.47	**11.62**	**10.28**	2.43
2 clusters	1	**6.19**	1.49	**6.51**	**5.72**	2.01
	2	**4.77**	1.39	**4.73**	**5.78**	0.97

z-scores >3 are statistically significant at $P < 0.001$, in **bold**.
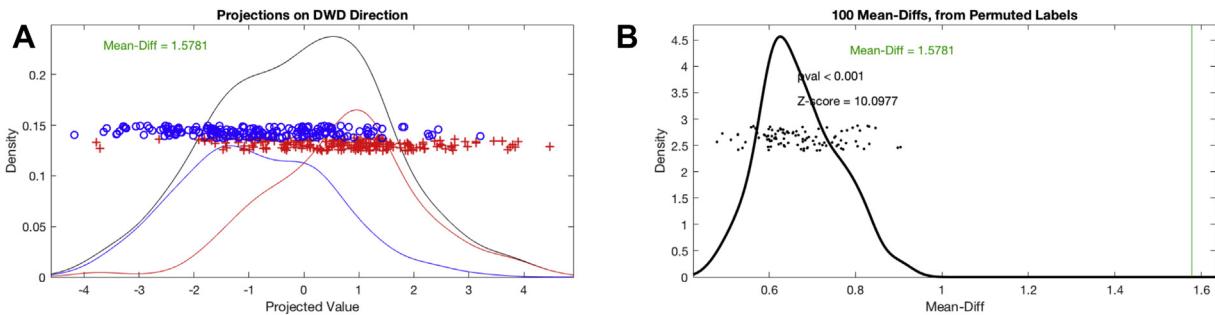
Fig. 1. The difference between progressors and non-progressors in the Distance Weighted Discrimination (DWD) direction utilizing all data and all observations. A) Actual data projected onto the DWD direction, with non-progressors in red and progressors in blue. B) The proportion of permuted values (black dots) greater than the test statistic (green vertical line) is indicated by the *P*-value (~0) and the corresponding z-score (10.10), indicating a significant difference between the two classes.
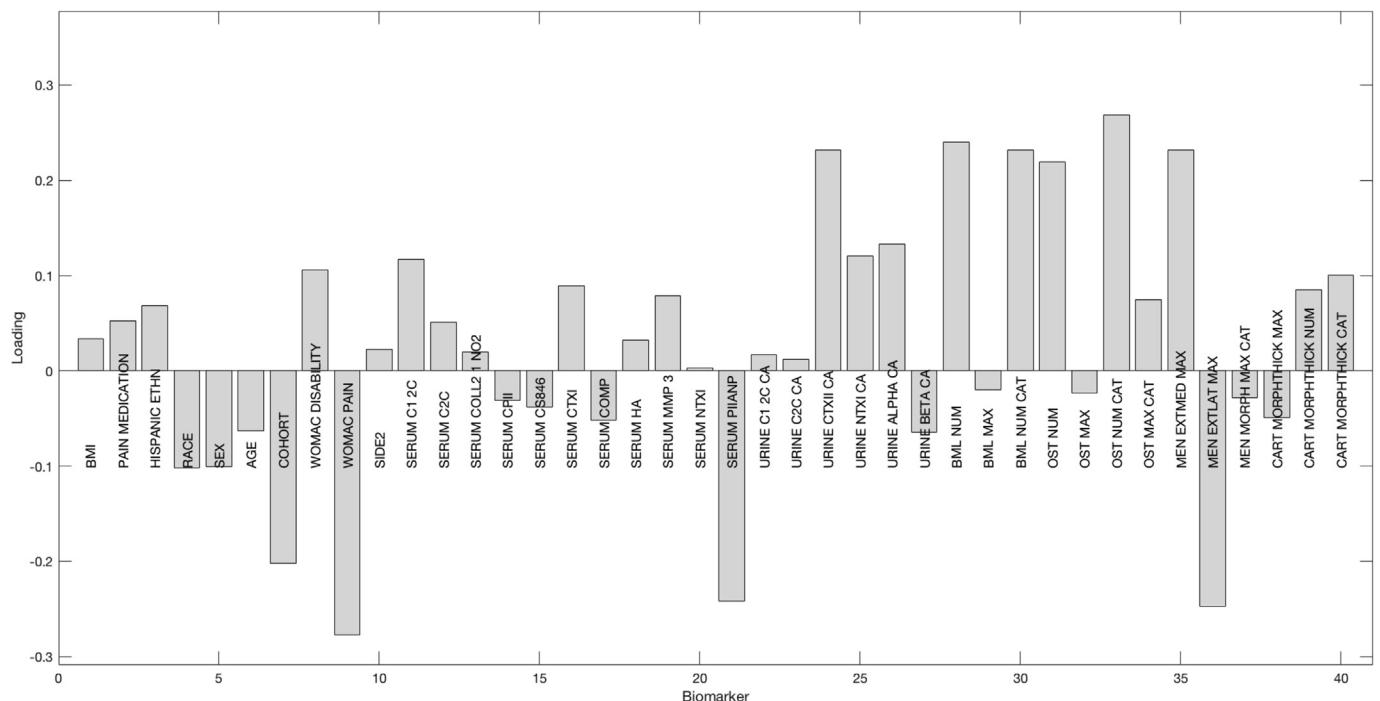
Fig. 2. Loadings on the DWD direction showing the 40 measures with the greatest contribution to the difference between progressors and non-progressors (to progression above the null and to non-progression below the null). BMI: Body mass index; CA: Creatinine adjusted; C12C: Col2-3/4 C-terminal cleavage product of types I and II collagen; C2C: Col2-3/4 C-terminal cleavage product of human type II collagen; Coll21NO2: Nitrated epitope of the alpha-helical region of type II collagen; CPII: C-propeptide of type II collagen; CS846: Chondroitin sulfate 846 epitope; CTXI, alpha, beta: C-terminal crosslinked telopeptide of type I collagen, alpha and beta; cartilage oligomeric matrix protein (COMP): Cartilage oligomeric matrix protein; HA: Hyaluronan; MMP3: Matrix metalloproteinase 3; NTXI: Crosslinked N-telopeptide of type I collagen; PIIANP: N-terminal pro-peptide of collagen IIA; CTXII: C-terminal crosslinked telopeptide type II collagen; Num: Number/count of regions/locations; Max: Maximum score; Cat: Categories; BML: Bone marrow lesion; OST: Osteophyte; Men ExtMed: Meniscal extrusion medial; Men ExtLat: Meniscal extrusion lateral; Men Morph: Meniscal morphology; Cart MorphThick: Cartilage thickness score.

hypothesis test, hence eliminating the need to adjust for multiple comparisons to control for Type I error inflation. This type of data-driven analysis can provide unanticipated insights into patterns in data that are not easily observable through more traditional methods and can therefore be hypothesis-generating. However, we recognize that this analysis is a preliminary step, requiring further internal validation using other methodology in this cohort as well as external validation in other datasets to test the robustness of the findings.

First, we were able to demonstrate clear separation of progressors from non-progressors using all of the available data simultaneously with a highly significant z-score. In this part of the analysis, each participant is represented by a point in high dimensional space which includes information about all of their other demographic, clinical, MRI, and biomarker characteristics,

such that all of this information is used when separating participants into outcome classes. This data-driven approach can therefore identify unsuspected patterns in the data more effectively and efficiently than individual hypothesis-driven testing of the association between each potential variable of interest and the outcome. In doing this, we find that those variables with the greatest contribution to identifying non-progressors included: WOMAC pain, lateral meniscal extrusion, and serum PIIANP. While WOMAC pain is counter-intuitive, this could indicate that those knees that have greater pain at baseline are less likely to experience progression by the pain criterion (that is, their baseline pain was already high and did not worsen).

Ipsilateral meniscal damage is known to be a risk factor for OA incidence and progression[19–22]. Radiographic progression in the FNIH cohort was defined as medial joint space loss of at least
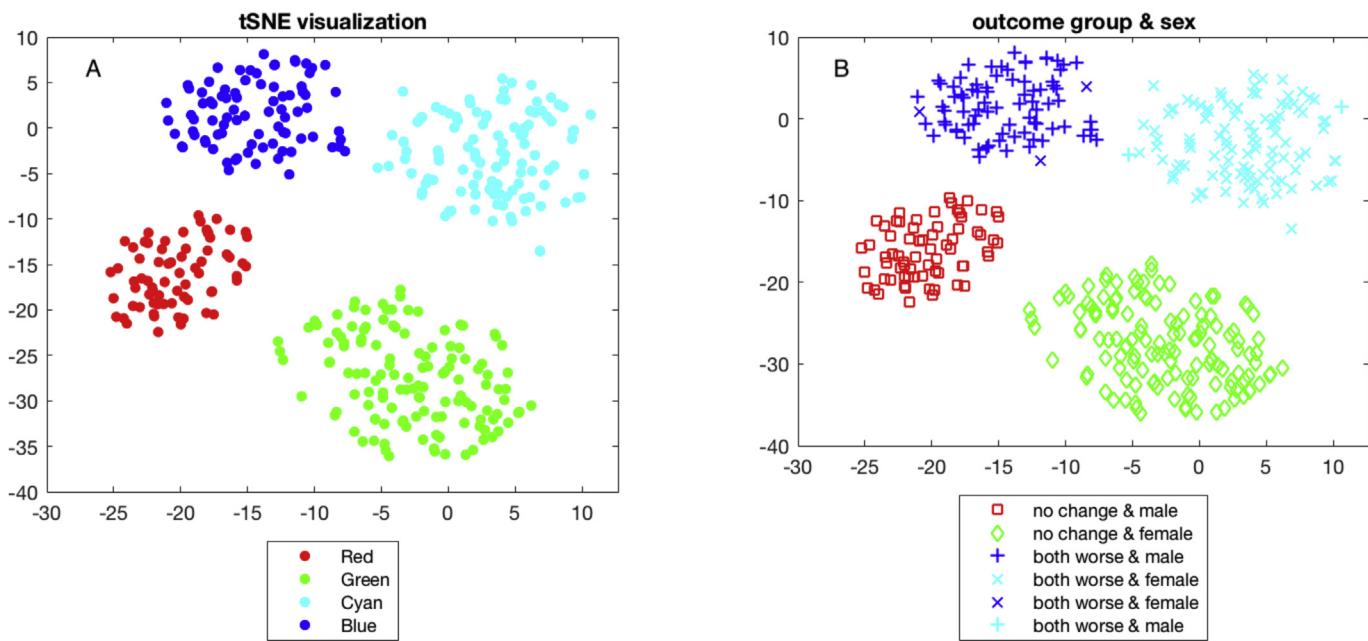
Fig. 3. Groupings using t-distributed stochastic neighbor embedding (t-SNE), four machine learning-generated clusters (A, with arbitrary two-dimensional *x* and *y* axes and color labels) and labeled based on progressor status (no change = non-progressor; both worse = progressor) and sex (B).
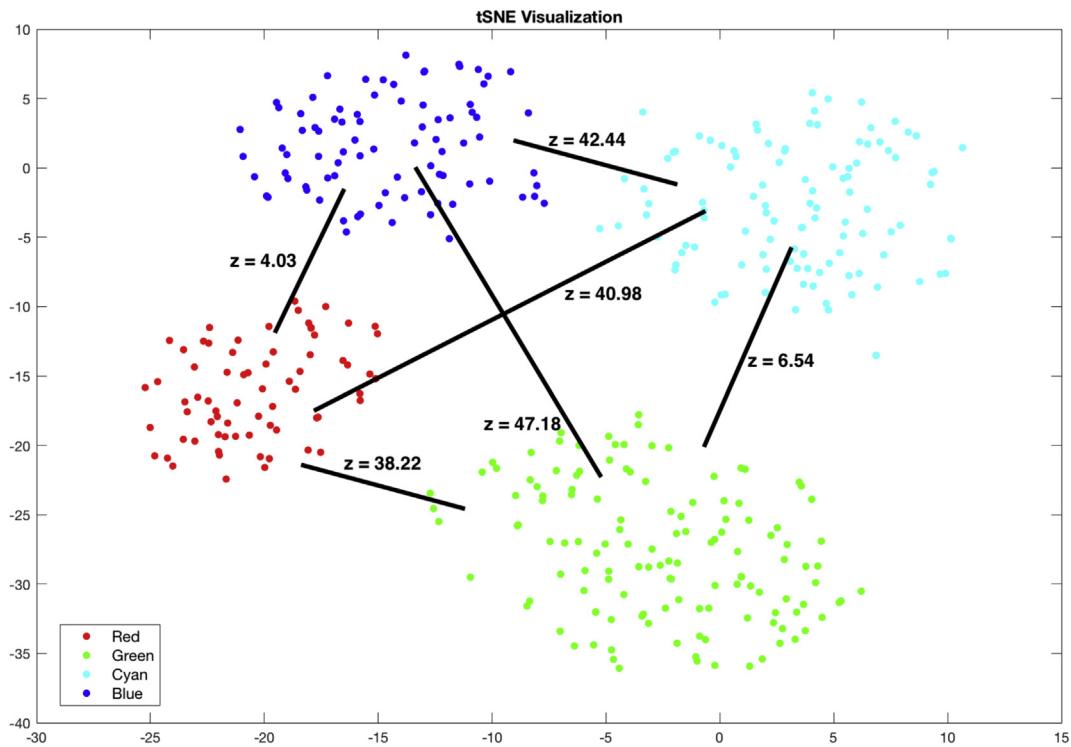

Fig. 4. Differences (z-scores from Direction-Projection-Permutation [DiProPerm]) between the 4 clusters identified using t-SNE. The colors again represent the clusters identified by the algorithm (as in Fig. 3A). The lines indicate the difference being tested with DiProPerm and the resultant z-score.

0.7 mm, so it is not unexpected that medial meniscal extrusion was a strong contributor to progression, while lateral meniscal extrusion contributed to non-progression in the medial compartment. The contribution of serum PIIANP (a marker of collagen synthesis) to the non-progressor classification is also supported in the literature, as elevated levels of this marker were associated with a lesser burden of knee osteophytes and hip joint space narrowing (JSN)

(although interestingly not knee JSN) in the Genetics of Generalized OA study[23].

In contrast, the variables with the largest contribution to progression included: medial meniscal extrusion (discussed above), the number of subregions with bone marrow lesions, the number of locations with any osteophyte, and urine CTX-II. Here again, the findings of this data-driven approach are supported by the
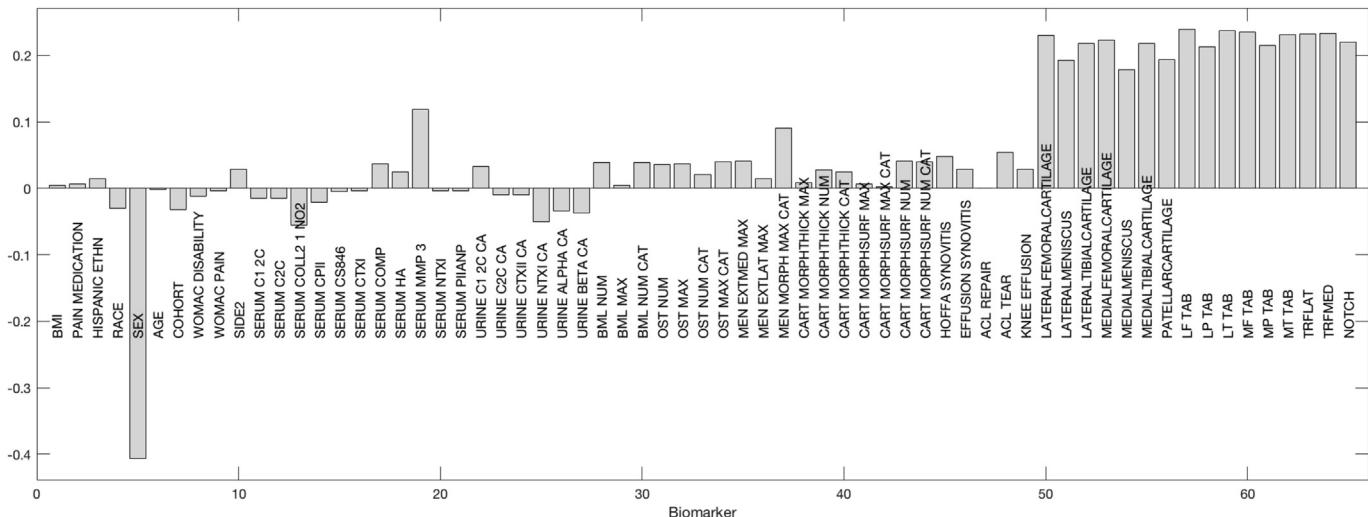
Fig. 5. Principal Components Analysis (PCA) was used to demonstrate which variables were driving the differences among clusters. Here, the loadings (strength of loading represented by height of bar) onto the first principal component based on the t-SNE clusters are represented. Cart MorphSurf: Cartilage surface score; ACL: Anterior cruciate ligament; TAB: Total Area of Subchondral Bone; LF/P/T: Lateral femur/patella/tibia; MF/P/T: Medial femur/patella/tibia; TRFLAT/Med: Total area of subchondral bone femoral trochlea lateral/medial; Notch: Total area of subchondral bone femoral notch.

literature. Bone marrow lesions are known to be associated with knee OA progression[24,25], and are often associated with ipsilateral meniscal damage[26]. Baseline osteophytes are predictive of radiographic progression[27], although they may be indicative of mechanical derangements[28]. Urine CTX-II is arguably the most consistent biomarker of knee OA progression[29] and was previously found to be predictive of case status in this cohort[6]. Interestingly, synovitis by MRI was not identified by this analysis as a key predictor of progression, despite prior research suggesting its importance in this and other cohorts[7,10,30].

When the observations were divided into clusters, cluster one mirrored the overall results while some additional features were noted in cluster 2, including a greater contribution to progression from osteophyte number and pain medication use (potentially indicating greater baseline disease and therefore risk of worsening), but a smaller contribution from urine biomarkers and bone marrow lesions. Cluster two also demonstrated a greater contribution to non-progression (above the null) from the serum biomarkers COMP (which is more often associated with incident rather than progressive KOA[31,32], and NTX-1 associated with reduced cartilage loss[33]).

In this first application of t-SNE methodology in OA, we note dramatic differences between men and women, potentially arguing for sex-stratified studies in KOA. Interestingly, groups of features appeared in each principal component in reasonable fashion. The first principal component from the t-SNE analysis was driven primarily by the sex difference, but also by quantitative MRI features such as cartilage and meniscal volume and subchondral bone area. The mode of variation reflected in the principal component two included bone marrow lesions, osteophytes, cartilage thickness and surface scores from the semi-quantitative MRI assessments, while principal component three included several of the baseline serum and urine biomarkers, most notably serum CTX-I and NTX-1, urine CTX-II and NTX-I, and urine CTX-I alpha and beta. These biomarkers largely mirror those found to be significant in the prior publication focused on baseline and 24-month time-integrated concentrations of biomarkers in this cohort, representing six of the eight biomarkers (all but serum Hyaluronan [HA] and urine Col2-3/4 C-terminal cleavage product of human type II collagen [C2C]) associated with worsening (defined there as pain, joint space

narrowing, or both)[6]. In that analysis, baseline urine CTX-II and urine CTX-I alpha also significantly predicted progression[6].

This analysis has many strengths, including the application of novel machine learning methodology well-suited to the situation of a large number of measures on a relatively small cohort (high dimension low sample size; HDLSS). It allows a broad overview of potential associations across many variables simultaneously. The FNIH cohort is a well-characterized, publicly available cohort with state-of-the-art imaging and biochemical biomarkers. However, as it was designed to look for the best available imaging and biochemical biomarkers, it did not include potentially novel markers, limiting the capacity for new discoveries in analyses using this dataset. The focus on proven markers and thus preselection of likely key indicators in this cohort also limits our ability to consider other predictive factors, and additional outcomes were not available. Other limitations of this analysis include a relative lack of heterogeneity in the cohort overall, unknown generalizability to other populations and time frames, and the focus on disease progression only (rather than prevalent or incident disease). Ideally, external validation would be employed to confirm our findings; however, there is a limited ability to generalize findings from this cohort given the extensive set of assessments obtained that is not readily replicable in other groups or in routine clinical practice. We focused on a primary comparison between both progressors and non-progressors, but future work could study additional comparisons in greater detail.

Conclusion

Utilizing novel machine learning methodology, we have efficiently identified multiple variables that are most associated with progressor status (e.g., bone marrow lesions, osteophytes, medial meniscal extrusion, and urine CTX-II) in a large KOA dataset, and noted a marked difference by sex. These innovative machine learning methods provide a way to assess numerous variables of different types and scalings simultaneously in relation to KOA progression, and could be readily applied to other outcomes of interest. While it is beyond the scope of the present analysis, such methodology could identify both known and novel KOA phenotypes, potentially improving patient selection for specific

interventions, a goal of Precision Medicine, and providing insight into pathophysiology in this heterogeneous condition.

Author contributions

Conception and design: AEN, FF, LFC, JSM, RFL.

Analysis and interpretation of data: AEN, FF, LA, RJC, TAS, JSM.

Drafting the article: AEN, FF.

Critically revising the article for important intellectual content: AEN, FF, LA, RJC, TAS, LFC, JSM, RFL.

Final approval of the submitted version: AEN, FF, LA, RJC, TAS, LFC, JSM, RFL.

Competing interests

The authors have no relevant competing interests to disclose.

Data Statement

This analysis used publicly available data from the FNIH OA Biomarkers Consortium, available at https://nda.nih.gov/oai.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2018.12.027.

References

1. Deveza LA, Loeser RF. Is osteoarthritis one disease or a collection of many? Rheumatology 2017;57, https://doi.org/10.1093/rheumatology/kex417.
2. Dell'Isola A, Allan R, Smith SL, Marreiros SS, Steultjens M. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. BMC Muscoskelet Disord 2016;17:425, https://doi.org/10.1186/s12891-016-1286-2.
3. Dell'Isola A, Steultjens M. Classification of patients with knee osteoarthritis in clinical phenotypes: data from the osteoarthritis initiative. PLoS One 2018;13, e0191045, https://doi.org/10.1371/journal.pone.0191045.
4. Deveza LA, Melo L, Yamato TP, Mills K, Ravi V, Hunter DJ. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. Osteoarthritis Cartilage 2017;25:1926–41, https://doi.org/10.1016/j.joca.2017.08.009.
5. Hunter DJ, Nevitt M, Losina E, Kraus V. Biomarkers for osteoarthritis: current position and steps towards further validation. Best Pract Res Clin Rheumatol 2014;28:61–71, https://doi.org/10.1016/j.berh.2014.01.007.
6. Kraus VB, Collins JE, Hargrove D, Losina E, Nevitt M, Katz JN, et al. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH OA biomarkers consortium. Ann Rheum Dis 2017;76:186–95, https://doi.org/10.1136/annrheumdis-2016-209252.
7. Roemer FW, Guermazi A, Collins JE, Losina E, Nevitt MC, Lynch JA, et al. Semi-quantitative MRI biomarkers of knee osteoarthritis progression in the FNIH biomarkers consortium cohort - Methodologic aspects and definition of change. BMC Muscoskelet Disord 2016;17:466, https://doi.org/10.1186/s12891-016-1310-6.
8. Eckstein F, Collins JE, Nevitt MC, Lynch JA, Kraus VB, Katz JN, et al. Brief report: cartilage thickness change as an imaging biomarker of knee osteoarthritis progression: data from the Foundation for the National Institutes of Health osteoarthritis biomarkers consortium. Arthritis Rheum 2015;67:3184–9, https://doi.org/10.1002/art.39324.
9. Hunter D, Nevitt M, Lynch J, Kraus VB, Katz JN, Collins JE, et al. Longitudinal validation of periarticular bone area and 3D shape as biomarkers for knee OA progression? Data from the FNIH OA biomarkers consortium. Ann Rheum Dis 2016;75:1607–14, https://doi.org/10.1136/annrheumdis-2015-207602.
10. Collins JE, Losina E, Nevitt MC, Roemer FW, Guermazi A, Lynch JA, et al. Semiquantitative imaging biomarkers of knee osteoarthritis progression: data from the Foundation for the National Institutes of Health osteoarthritis biomarkers consortium. Arthritis Rheum 2016;68:2422–31, https://doi.org/10.1002/art.39731.
11. Marron JS, Todd MJ, Ahn J. Distance-weighted discrimination. J Am Stat Assoc 2007;102:1267–71, https://doi.org/10.1198/016214507000001120.
12. An H, Marron JS, Schwartz TA, Renner JB, Liu F, Lynch JA, et al. Novel statistical methodology reveals that hip shape is associated with incident radiographic hip osteoarthritis among African American women. Osteoarthritis Cartilage 2016;24:640–6, https://doi.org/10.1016/j.joca.2015.11.013.
13. Wei S, Lee C, Wichers L, Marron JS. Direction-projection-permutation for high-dimensional hypothesis tests. J Comput Graph Stat 2016;25:549–69, https://doi.org/10.1080/10618600.2015.1027773.
14. Davies DL, Bouldin DW. Cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;1:224–7, https://doi.org/10.1109/Tpami.1979.4766909.
15. Rousseeuw PJ. Silhouettes - a graphical aid to the interpretation and validation of cluster-analysis. J Comput Appl Math 1987;20:53–65, https://doi.org/10.1016/0377-0427(87)90125-7.
16. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Methodol 2001;63:411–23, https://doi.org/10.1111/1467-9868.00293.
17. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat 1974;3:1–27, https://doi.org/10.1080/03610927408827101.
18. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9:2579–605.
19. Sharma L, Eckstein F, Song J, Guermazi A, Prasad P, Kapoor D, et al. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. Arthritis Rheum 2008;58:1716–26, https://doi.org/10.1002/art.23462.
20. Englund M, Guermazi A, Roemer FW, Aliabadi P, Yang M, Lewis CE, et al. Meniscal tear in knees without surgery and the development of radiographic osteoarthritis among middle-aged and elderly persons: the multicenter osteoarthritis study. Arthritis Rheum 2009;60:831–9, https://doi.org/10.1002/art.24383.
21. Roemer FW, Kwoh CK, Hannon MJ, Hunter DJ, Eckstein F, Fujii T, et al. What comes first? Multitissue involvement leading to radiographic osteoarthritis: magnetic resonance

imaging-based trajectory analysis over four years in the osteoarthritis initiative. Arthritis Rheum 2015;67:2085–96, https://doi.org/10.1002/art.39176.

22. Guermazi A, Eckstein F, Hayashi D, Roemer FW, Wirth W, Yang T, *et al.* Baseline radiographic osteoarthritis and semi-quantitatively assessed meniscal damage and extrusion and cartilage damage on MRI is related to quantitatively defined cartilage thickness loss in knee osteoarthritis: the multicenter osteoarthritis study. Osteoarthritis Cartilage 2015;23:2191–8, https://doi.org/10.1016/j.joca.2015.06.017.

23. Daghestani HN, Jordan JM, Renner JB, Doherty M, Wilson AG, Kraus VB. Serum N-propeptide of collagen IIA (PIIANP) as a marker of radiographic osteoarthritis burden. PLoS One 2017;12, e0190251, https://doi.org/10.1371/journal.pone.0190251.

24. Tanamas SK, Wluka AE, Pelletier JP, Pelletier JM, Abram F, Berry PA, *et al.* Bone marrow lesions in people with knee osteoarthritis predict progression of disease and joint replacement: a longitudinal study. Rheumatology 2010;49: 2413–9, https://doi.org/10.1093/rheumatology/keq286.

25. Roemer FW, Guermazi A, Javaid MK, Lynch JA, Niu J, Zhang Y, *et al.* Change in MRI-detected subchondral bone marrow lesions is associated with cartilage loss: the MOST Study. A longitudinal multicentre study of knee osteoarthritis. Ann Rheum Dis 2009;68:1461–5, https://doi.org/10.1136/ard.2008.096834.

26. Lim YZ, Wang Y, Wluka AE, Davies-Tuck ML, Teichtahl A, Urquhart DM, *et al.* Are biomechanical factors, meniscal pathology, and physical activity risk factors for bone marrow lesions at the knee? A systematic review. Semin Arthritis Rheum 2013;43:187–94, https://doi.org/10.1016/j.semarthrit.2013.03.002.

27. Saunders J, Ding C, Cicuttini F, Jones G. Radiographic osteoarthritis and pain are independent predictors of knee cartilage loss: a prospective study. Intern Med J 2012;42:274–80, https://doi.org/10.1111/j.1445-5994.2011.02438.x.

28. Felson DT, Gale DR, Elon Gale M, Niu J, Hunter DJ, Goggins J, *et al.* Osteophytes and progression of knee osteoarthritis. Rheumatology 2005;44:100–4, https://doi.org/10.1093/rheumatology/keh411.

29. Hosnijeh FS, Runhaar J, van Meurs JB, Bierma-Zeinstra SM. Biomarkers for osteoarthritis: can they be used for risk assessment? A systematic review. Maturitas 2015;82:36–49, https://doi.org/10.1016/j.maturitas.2015.04.004.

30. Felson DT, Niu J, Neogi T, Goggins J, Nevitt MC, Roemer F, *et al.* Synovitis and the risk of knee osteoarthritis: the MOST Study. Osteoarthritis Cartilage 2016;24:458–64, https://doi.org/10.1016/j.joca.2015.09.013.

31. Kluzek S, Bay-Jensen AC, Judge A, Karsdal MA, Shorthose M, Spector T, *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, https://doi.org/10.3109/1354750X.2015.1105498.

32. Zhang J. Meta-analysis of serum C-reactive protein and cartilage oligomeric matrix protein levels as biomarkers for clinical knee osteoarthritis. BMC Muscoskelet Disord 2018;19:22, https://doi.org/10.1186/s12891-018-1932-y.

33. Berry PA, Maciewicz RA, Cicuttini FM, Jones MD, Hellawell CJ, Wluka AE. Markers of bone formation and resorption identify subgroups of patients with clinical knee osteoarthritis who have reduced rates of cartilage loss. J Rheumatol 2010;37: 1252–9, https://doi.org/10.3899/jrheum.091055.