RHEUMATOLOGY

Original article

Endotypes of primary osteoarthritis identified by plasma metabolomics analysis

Salem Werdyani [1], Ming Liu[1], Hongwei Zhang[2], Guang Sun[2], Andrew Furey[3], Edward W. Randell[4], Proton Rahman[2] and Guangju Zhai [1]

Abstract

Objective. To identify endotypes of osteoarthritis (OA) by a metabolomics analysis.

Methods. Study participants included hip/knee OA patients and controls. Fasting plasma samples were metabolomically profiled. Common factor analysis and K-means clustering were applied to the metabolomics data to identify the endotypes of OA patients. Logistic regression was utilized to identify the most significant metabolites contributing to the endotypes. Clinical and epidemiological factors were examined in relation to the identified OA endotypes.

Results. Six hundred and fifteen primary OA patients and 237 controls were included. Among the 186 metabolites measured, 162 passed the quality control analysis. The 615 OA patients were classified in three clusters (A, 66; B, 200; and C, 349). Patients in cluster A had a significantly higher concentration of butyrylcarnitine (C4) than other clusters and controls (all $P < 0.0002$). Elevated C4 is thought to be related to muscle weakness and wasting. Patients in cluster B had a significantly lower arginine concentration than other clusters and controls (all $P < 7.98 \times 10^{-11}$). Cluster C patients had a significantly lower concentration of lysophosphatidylcholine (with palmitic acid), which is a pro-inflammatory bioactive compound, than other clusters and controls ($P < 3.79 \times 10^{-6}$). Further, cluster A had a higher BMI and prevalence of diabetes than other clusters (all $P \leq 0.0009$), and also a higher prevalence of coronary heart disease than cluster C ($P = 0.04$). Cluster B had a higher prevalence of coronary heart disease than cluster C ($P = 0.003$) whereas cluster C had a higher prevalence of osteoporosis ($P = 0.009$).

Conclusion. Our data suggest three possible clinically actionable endotypes in primary OA: muscle weakness, arginine deficit and low inflammatory OA.

Key words: osteoarthritis, endotypes, butyrylcarnitine, arginine, lysophosphatidylcholine

> **Rheumatology key messages**
>
> - Data demonstrated that metabolomics was a novel tool to classify OA patients into subtypes.
> - Three endotypes existed in OA, characterized by C4, arginine and lysophosphatidylcholine levels in plasma.
> - The findings provided a better understanding of OA pathogenesis toward developing personalized tools for OA management.

CLINICAL SCIENCE

Introduction

Osteoarthritis (OA) is the most common chronic progressive joint condition and one of the 10 most disabling diseases in developed countries [1]. Since the fifties of the last century, the prevalence of OA has doubled, and about 240 million individuals have been diagnosed with OA worldwide. OA is a heterogeneous group of overlapping distinct conditions that have different aetiologies, but similar clinical manifestations [2]. OA heterogeneity represents a major obstacle to the detection of the efficacy of disease-modifying OA drugs. Efforts have been made to classify subtypes of OA patients based on epidemiological factors [3], joint structure changes seen on MRI [4], disease mechanism of onset and pathophysiology [5], anatomical components [6], clinical manifestations [7], disease stage [8], affected joints [9] and

[1]Division of Biomedical Sciences (Genetics), [2]Discipline of Medicine, [3]Discipline of Surgery, [4]Department of Laboratory Medicine, Faculty of Medicine, Memorial University of Newfoundland, St John's, NL, Canada

Correspondence to: Guangju Zhai, Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, Craig L. Dobbin Genetics Research Centre, St John's, NL, Canada, A1B 3V6, Email: guangju.zhai@med.mun.ca

inflammation [10], but significant overlapping features among the OA subtypes defined by these methods limit their clinical application [11]. Hence, there is a need for novel tools that can provide clinically actionable classifications of OA subtypes for personalized medicine in OA management [11, 12].

Metabolomics provides a snapshot of the entire physiology of the host and its response to the environment and genetics, which can be associated with the outcome phenotype and endotypes [13]. Recently, application of metabolomics to OA research has identified several promising and potentially clinically actionable metabolic markers [12, 14]. We therefore hypothesized that endotypes of OA patients exist and can be identified by a metabolomics analysis, and tested our hypothesis in a well-established study—the Newfoundland Osteoarthritis Study (NFOAS) [15].

Methods

Study participants

The subjects were total hip or knee replacement patients due to primary OA who were recruited to the NFOAS [15] between November 2011 and September 2017 in St Clare's Mercy Hospital and Health Science Centre General Hospital in St John's, Newfoundland and Labrador (NL), Canada [11]. OA diagnosis was made based on the American College of Rheumatology OA clinical diagnostic criteria [16]. Pathology reports on cartilage and osteophytic irregularities were investigated post-surgery to confirm OA diagnosis. Self-reported OA-free controls were derived from the same source population that was originally recruited to the Complex Diseases in the Newfoundland Population: Environment and Genetics (CODING) study [17]. The study was approved by the Health Research Ethics Authority of NL (reference number 11.311), and written consent was obtained from all study participants.

Demographic and medical information

A self-administered questionnaire was used to collect the patients' demographic and medical information including age, sex, BMI and comorbidities. A patient's age was calculated at the total joint replacement surgery date. BMI was calculated by dividing a patient's weight in kilograms by the squared height in metres. Comorbidities including diabetes, hypertension, high cholesterol, coronary heart disease, gout, osteoporosis, breast cancer, colon cancer, skin cancer, melanoma, basal cell carcinoma and squamous cell carcinoma were self-reported by the OA patients [11].

Metabolic profiling

Blood samples were collected after at least 8 hours fasting, and plasma was separated from the whole blood by a standard protocol and stored in $-80°C$ freezers until

analysis [18]. Metabolic profiling was performed on plasma using the Biocrates AbsoluteIDQ p180 kit (Biocrates Life Sciences AG, Innsbruck, Austria), which measures the concentrations of 186 metabolites. Supplementary Table S1, available at *Rheumatology* online, provides the full list of the metabolites measured in the study. The profiling was done using an API4000 Qtrap® tandem mass spectrometry instrument (Applied Biosystems/MDS Analytical Technologies, Foster City, CA, USA) equipped with an Agilent 1100 HPLC system (Agilent Technologies, Santa Clara, CA, USA) at the Metabolomics Innovation Centre (https://www.metabolomicscentre.ca). The complete analytical process of targeted metabolite concentrations was performed using the MetIQ software package, which is an integral part of the AbsoluteIDQ® p180 kit, and the concentrations were reported in micromolar [19]. Our in-house reproducibility of the assay was performed in 23 samples as previously described [20]; the mean coefficient of variation (CV) for all metabolites was 0.07 (0.05) μM.

Statistical analysis

Quality control procedures removed metabolites from subsequent analysis if >10% of the samples had values below the limit of detection to minimize the false-positive results as a standard practice in metabolomics studies [21]. For metabolites with values below the limit of detection in <10% of samples, missing values were imputed by the mean of the given metabolites. Principal component analysis demonstrated that we did not have any batch effect in our experiment; therefore, no correction for batch effects was performed.

Bartlett's test of sphericity and the Kaiser–Meyer–Olkin measure of sampling adequacy were used to evaluate the factorability of the data and to determine whether there were meaningful latent factors within the metabolomics data [22]. Then, the number of factors was determined using a scree plot and parallel analysis based on the calculated eigenvalues from the correlation matrices [22]. Based on the suggested number of factors in the metabolomics data, common factor analysis was performed to reduce the dimensionality of the metabolomics data [22, 23]. The identified factors were utilized in the subsequent clustering analysis if their corresponding eigenvalues were >1.0 along with rotated absolute factor loading for each metabolite >0.3 detected in only one factor [24]. Factor scores for each study participant were calculated by adding up all the metabolite concentrations in each identified factor and used in the clustering analysis [24].

Subsequently, the Hopkins statistic was used to assess the clustering tendency of the calculated factor scores from 615 OA patients to investigate whether the metabolomics data were clusterable with any inherent grouping structure [25]. Then, the optimal number of clusters was determined using the silhouette width and elbow plot methods based on the percentage of variance between expected clusters identified by F-test [26]. Thereafter, the factor scores from all 615 OA

patients were utilized in the clustering analysis using the most commonly unsupervised machine learning K-means clustering algorithm to identify the endotypes of OA patients based on the similarity and minimized variance between the clustered subjects [26].

Then, multivariable logistic regression was performed to identify the most significant metabolites that contribute to the classification of the endotypes of OA patients and differentiate the patients in each group from healthy controls. Metabolite ratios between those identified metabolites as proxies for enzymatic reaction were examined to identify the most likely metabolic pathways [27]. Receiver operating characteristic (ROC) analysis was conducted to evaluate the performance of identified metabolites and ratios in the classification of each endotype. The area under the curve (AUC) was calculated and optimal cut-off values were determined using the maximum sensitivity and specificity simultaneously (MaxSpSe) method. Furthermore, we examined 15 clinical and epidemiological variables in relation to the identified endotypes, including age, sex, BMI and comorbidities. Joint specificity (knee/hip) for the identified endotypes was also tested. Significance level was defined as $P \leq 0.0003$ after correction of multiple testing of 162 metabolites with the Bonferroni method. All the analyses were performed in R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) with psych, GPArotation, corpcor, Factoextra, NbClust, cluster, clValid, stats, pROC, caret, OptimalCutpoints, dplyr and ggplot2 packages.

To complement our data analytic method, we also performed a split sample analysis although it was not advisable because of the insufficient sample size. Specifically, we randomly split our cohort equally into training and validation datasets with similar distribution of age, sex and BMI between the two datasets and applied exactly the same methods described above to the training dataset and replicated the results in the validation dataset. The sample splitting was done in R using the CreateDataPartition function in the caret package [28].

Results

A total of 615 primary OA patients and 237 OA-free controls were included in the study. OA patients were significantly older ($P = 0.0001$) and had a higher BMI than controls ($P = 0.0001$), but there was no difference in sex distribution between OA patients and controls ($P = 0.27$; Table 1).

Among the 186 metabolite concentrations measured, 162 metabolites passed the quality control criteria and were included in the analysis. The evaluation of the factorability of the metabolomics data using Bartlett's test of sphericity showed statistical significance ($P = 0.0001$), suggesting a significant difference of the correlation matrix from the identity matrix, thereby indicating that the metabolomics data are factorable. Also, the Kaiser–Meyer–Olkin test indicated that the metabolomics data were adequate for factor analysis with overall measure of sampling adequacy of 0.93. A scree plot (Fig. 1A) and parallel analysis (Fig. 1B) estimated that the metabolomics data can be categorized by 17 factors based on their correlation and relationship. Accordingly, common factor analysis retrieved a rotated factor matrix of 17 distinct factors from 93 metabolites that had a rotated factor loading $>|0.3|$ in only one factor. These 93 metabolites included 12 amino acids, three biogenic amines, 22 acylcarnitines, 48 glycerophospholipids, seven sphingolipids and one sugar hexose ($>90\%$ is glucose) (Supplementary Table S2, available at *Rheumatology* online).

The Hopkins statistic indicated that the metabolomics dataset of OA patients was significantly clusterable with Hopkins statistic value of 0.18. The minimum silhouette width and elbow plot anticipated that the 615 OA patients could be classified into three main clusters based on the metabolomics factor scores (Fig. 1C and D). K-means clustering analysis determined three distinct main clusters (cluster A, 66; cluster B, 200; and cluster C, 349) of OA patients with between sum of squares/total sum of squares $= 77.0\%$ (Fig. 2).

Logistic regression identified butyrylcarnitine (C4), arginine and lysophosphatidylcholine with palmitic acid (lysoPC a C16:0) to be the major contributing factors for the classification and differentiation between clusters A, B and C. The cluster A patients were characterized by a significantly higher concentration of C4 than the other two clusters ($P = 2.04 \times 10^{-4}$) and OA-free controls ($P = 4.08 \times 10^{-8}$) (Table 2). While it did not reach the pre-defined significance, the second metabolite was a phosphatidylcholine with 40 carbons and three double bonds (PC ae C40:3) for which the patients in cluster A had a lower concentration than other OA patients ($P = 0.01$) and controls ($P = 1.30 \times 10^{-11}$) (Table 2). When the ratio between these two metabolites was considered, the significance became stronger. The PC ae C40:3 to C4 ratio was significantly

TABLE 1 The characteristics of the 615 OA patients and 237 OA-free controls

	OA patients (*n* = 615)	OA-free controls (*n* = 237)	*P*-value
Sex, female, %	55.28	59.49	0.265
Age, mean (S.D.), years	66.05 (8.55)	49.23 (12.71)	0.0001
BMI, mean (S.D.), kg/m^2	33.74 (6.86)	29.03 (5.05)	0.0001

P-values were obtained from chi-square test for sex distribution and Student's *t*-test for continuous variables.

Fig. 1 Estimation of the factor and cluster numbers from the concentrations of 162 metabolites of 615 OA patients
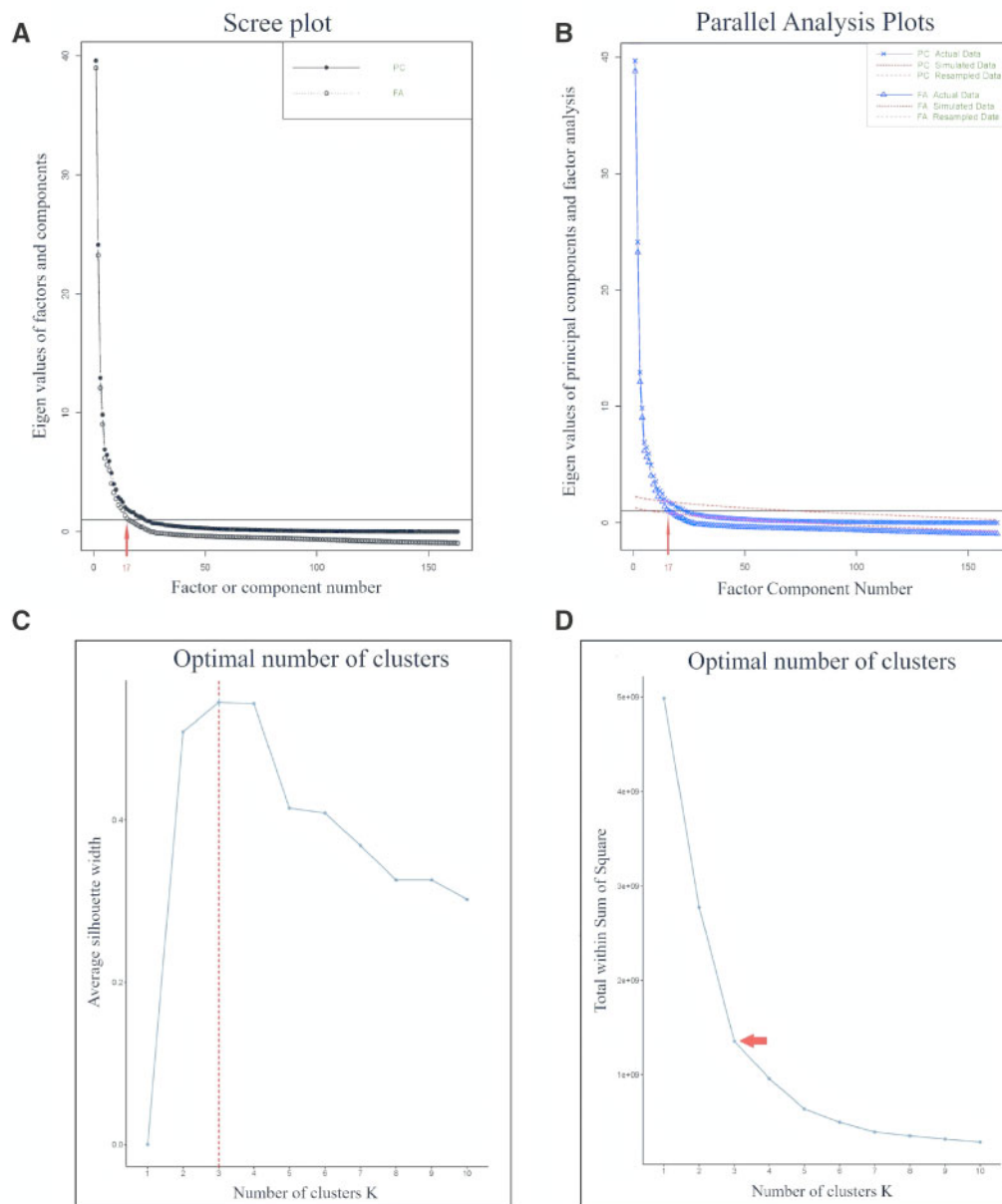
The red arrows in the Scree plot (**A**) and parallel analysis plot (**B**) indicate that the metabolite data could be categorized by 17 factors based on their correlation and relationship. The red vertical dash line in the minimum silhouette width plot (**C**) and the red arrow in the elbow plot (**D**) indicate that the 615 OA patients can be classified into three main clusters based on the calculated metabolite factor scores from the 17 identified factors. FA: factor analysis; PC: principal component analysis.

lower in cluster A patients than in patients of other clusters ($P = 4.89 \times 10^{-7}$) and controls ($P = 1.31 \times 10^{-11}$) (Table 2). ROC analysis showed that the ratio had an AUC of 0.92 (95% CI: 0.89–0.95) to distinguish cluster A patients from OA-free controls with a sensitivity of 0.83 and specificity of 0.83 at the optimal cutoff value of 4.75. The ratio had a moderate discriminatory power to distinguish patients in cluster A from patients in other clusters with an AUC of 0.68 (95% CI: 0.62–0.75) (Fig. 3A).

The cluster B patients had a significantly lower arginine concentration than those of clusters A and C ($P = 3.44 \times 10^{-16}$) and controls ($P = 7.98 \times 10^{-11}$) (Table 2). The ROC curve analyses showed that arginine had an AUC of 0.82 (95% CI: 0.78–0.85), and a sensitivity of 0.74 and specificity of 0.75 to discriminate patients in cluster B from patients in the other two clusters at an optimal cutoff value of 28.30 μM; and an AUC of 0.99 (95% CI: 0.98–1.00) with a sensitivity of 0.95 and a specificity of

Fig. 2 Three distinct clusters of OA patients identified by K-means clustering method coupled with metabolomics data
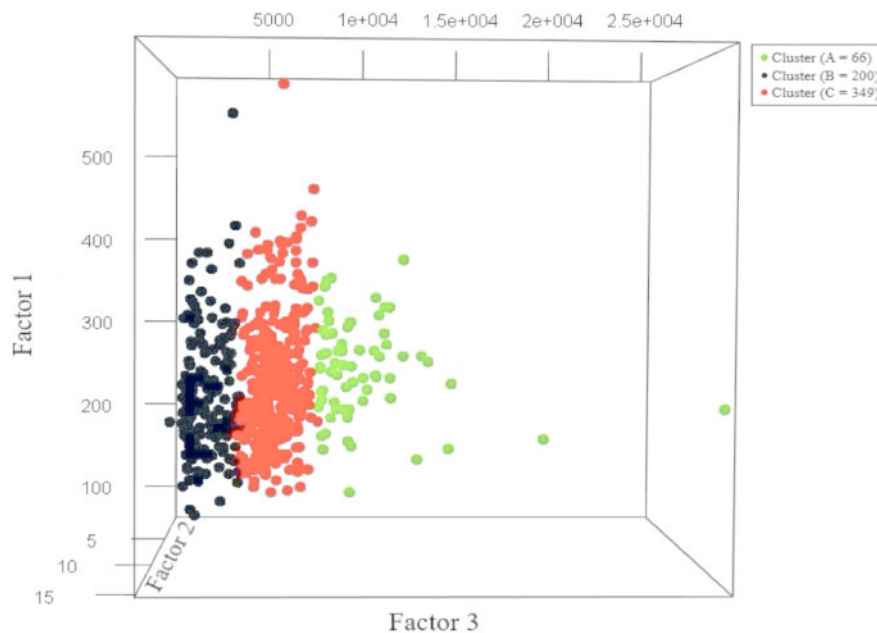

Table 2 Most significant metabolites contributing to the classification of endotypes of primary OA patients

| Metabolite	Mean (s.d.)	Other OA patients		Controls	
		Mean (s.d.)	*P*-value	Mean (s.d.)	*P*-value
Cluster A ($n = 66$)					
C4 concentration, μM	0.34 (0.25)	0.26 (0.16)	0.0002	0.23 (0.14)	4.08×10^{-8}
PC ae C40:3 concentration, μM	0.85 (0.32)	0.97 (0.37)	0.01	1.67 (0.58)	1.30×10^{-11}
PC ae C40:3 to C4 ratio	3.11 (1.57)	4.51 (2.39)	4.89×10^{-7}	8.81 (4.51)	1.31×10^{-11}
Cluster B ($n = 200$)					
Arginine concentration, μM	21.39 (17.83)	41.76 (19.48)	3.44×10^{-16}	115.08 (50.77)	7.98×10^{-11}
Cluster C ($n = 349$)					
LysoPC a C16:0 concentration, μM	79.02 (25.46)	130.47 (65.10)	1.42×10^{-12}	82.61 (37.02)	3.79×10^{-6}
PC ae C38:2 concentration, μM	1.94 (1.44)	1.74 (1.20)	0.006	4.54 (2.30)	1.33×10^{-7}
LysoPC a C16:0 to PC ae C38:2 ratio	53.15 (25.09)	90.93 (50.67)	0.0001	22.34 (12.13)	0.0001

C4: butyrylcarnitine; lysoPC a C16:0: lysophosphatidylcholine with 16 carbons and no double bond; PC ae C38:2: phosphatidylcholine acyl-alkyl with 38 carbons and two double bonds; PC ae C40:3: phosphatidylcholine acyl-alkyl with 40 carbons and three double bonds.

0.96 to discriminate patients in cluster B from controls at the optimal cutoff value of 56.10 μM (Fig. 3B).

Patients in cluster C were distinguished from patients in the other two clusters and OA-free controls by a lower concentration of lysoPC a C16:0 (all $P \leq 3.79 \times 10^{-6}$) (Table 2). Although it did not reach the pre-defined significance level, the second top associated metabolite was a phosphatidylcholine with 38 carbons and two double bonds (PC ae C38:2)

($P \leq 0.006$; Table 2). When examining the ratio of these two metabolites, we found that the significance became stronger and the lysoPC a C16:0 to PC ae C38:2 ratio was significantly lower in patients in cluster C than in patients of the other clusters ($P = 0.0001$) but significantly higher than in OA-free controls ($P = 0.0001$) (Table 2). The ratio was also significantly higher in patients of clusters A and B than in controls ($P = 0.0001$). The ROC curve analyses showed that the

Fig. 3 ROC curve analysis results
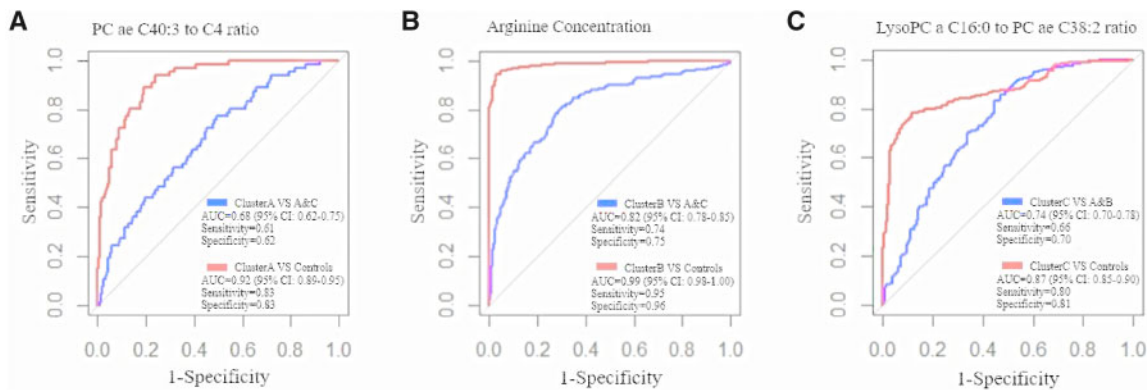
(**A**) The PC ae C40:3 to C4 ratio distinguishing patients in cluster A from other clusters and controls. (**B**) Arginine distinguishing patients in cluster B from other clusters and controls. (**C**) LysoPC a C16:0 to PC ae C38:2 ratio distinguishing patients in cluster C from other clusters and controls. C4: butyrylcarnitine; lysoPC a C16:0: lysophosphatidylcholine with 16 carbons and no double bond; PC ae C38:2: phosphatidylcholine acyl-alkyl with 38 carbons and two double bonds; PC ae C40:3: phosphatidylcholine acyl-alkyl with 40 carbons and three double bonds; ROC: receiver operating characteristics.

ratio had an AUC of 0.74 (95% CI: 0.70–0.78), and sensitivity of 0.66 and specificity of 0.70 to discriminate patients in cluster C from patients in the other clusters at an optimal cutoff value of 63.28, and an AUC of 0.87 (95% CI: 0.85–0.90) with a sensitivity of 0.80 and specificity of 0.81 to discriminate cluster C patients from OA-free controls with the optimal cutoff value of 32.01 (Fig. 3C).

In the split sample analysis, a total of 615 OA patients were randomly divided into a training dataset ($n = 308$) and a validation dataset ($n = 307$). The minimum silhouette width and elbow plot estimated that the training dataset could be classified into three main clusters (Supplementary Fig. S1, available at *Rheumatology* online), which was validated in the validation dataset (Supplementary Fig. S2, available at *Rheumatology* online). K-means clustering analysis identified three distinct clusters (TA, 39; TB, 95; and TC, 174) of OA patients in the training dataset, which was also validated in the validation dataset (VA, 10; VB, 112; and VC, 185). Logistic regression analysis identified exactly the same metabolites—arginine ($P = 4.28 \times 10^{-8}$) and lysoPC a C16:0 ($P = 4.79 \times 10^{-10}$)—that were the key contributors for clusters B and C as found in the entire cohort analysis, which was also confirmed in the validation dataset (with $P = 1.38 \times 10^{-7}$ and $P = 2.98 \times 10^{-9}$ for arginine and lysoPC a C16:0, respectively; Supplementary Table S3, available at *Rheumatology* online). C4 was also identified as the key contributor for cluster A in the training dataset and confirmed in the validation dataset, but the P-value (0.01) did not reach the pre-defined significance level because of the small sample size (Supplementary Table S3, available at *Rheumatology* online).

With regards to the clinical and epidemiological factors, there was no sex difference among the three

identified clusters (all $P > 0.22$), but patients in cluster A had a higher BMI and a higher prevalence of diabetes than those of the other two clusters (all $P \leq 0.0009$), as well as being significantly older than those of cluster B ($P = 0.018$) and had a higher prevalence of CHD than those cluster C ($P = 0.039$) (Table 3). Patients in cluster B had a significantly higher prevalence of CHD than those of cluster C ($P = 0.003$), whereas patients in cluster C had a significantly higher prevalence of osteoporosis than those of cluster B ($P = 0.009$). The prevalence of osteoporosis in cluster C was also higher than that in cluster A (22.06% *vs* 15.15%) but was not statistically significant (Table 3).

In addition, the study cohort included 68% of knee OA patients and 32% of hip OA patients. However, we found that there was no difference in the distribution of knee and hip OA among the three identified OA clusters ($P = 0.43$). The proportion of knee OA in clusters A, B and C was 72.73%, 62.32% and 69.86%, respectively.

Discussion

In this metabolomics analysis with a large sample size, we were able to identify three distinct endotypes of OA patients. The significant metabolite contributors to each of the three endotypes implied that the primary OA patients can be classified as having muscle weakness, arginine deficient, and low inflammatory OA. The findings provide new insights into the pathogenesis of primary OA and could help to develop personalized tools for OA management.

In a preliminary study with only 80 OA patients, we previously found that the OA patients can be classified into two distinct groups [11]. Eleven percent of the 80

TABLE 3 Significant epidemiological factors and comorbidities associated with each of the identified clusters

Factors	Cluster A	Cluster B	Cluster C	P-value		
				A vs B	A vs C	B vs C
Age, mean (s.d.), years	67.78 (7.37)	65.03 (8.32)	66.32 (8.82)	0.018	0.207	0.094
BMI, mean (s.d.), kg/m^2	36.57 (7.27)	33.29 (6.70)	33.46 (6.74)	0.0009	0.0008	0.776
Diabetes (%)	54.55	13.50	12.89	1.039×10^{-11}	4.906×10^{-15}	0.840
CHD (%)	7.57	8.00	2.58	0.912	0.039	0.003
Osteoporosis (%)	15.15	13.00	22.06	0.658	0.206	0.009

P-values were obtained from chi-square test or Student's t-test where appropriate. CHD: coronary heart disease.

patients were classified into one group characterized by a high concentration of acetylcarnitines in the synovial fluid. The findings of the current study were consistent with it, and we found that 11% of the 615 primary OA patients were classified into cluster A based on plasma metabolic profiles, and a specific acetylcarnitine, C4, was the key contributor to the clustering. The patients in cluster A had on average a 33% increase in C4 compared with the OA patients in the other two clusters and a 52% increase compared with OA-free controls.

Acetylcarnitines are used to transport fatty acids from cytosol into the mitochondrial matrix for energy production [11]. C4 is a short-chain acetylcarnitine and responsible for the transfer of short-chain fatty acids. It has been reported that accumulation of C4 reflects the abnormal concentration of tissue butyryl-CoA due to a defect or inhibition of short-chain acyl-CoA dehydrogenase, which is the key enzyme involved in the short-chain fatty acid β-oxidation pathway in mitochondria, leading to an energy pathway defect and generalized muscle weakness [29]. An elevated concentration of C4 in blood is one of the diagnostic parameters for short-chain acyl-CoA dehydrogenase (SCAD) deficiency. SCAD deficiency is a clinically heterogeneous disorder with variable clinical phenotypes ranging from fatal metabolic decompensation in early life to subtle adult onset with asymptomatic phenotypes in some patients [30]. Adult patients are more likely to have problems related to muscle weakness and wasting [30, 31]. Muscle weakness has been associated with OA [32]. Thus, our findings suggest that patients in cluster A might have weakened muscle strength that makes them susceptible to OA. To the best of our knowledge, elevated plasma concentration of C4 had not been reported in OA patients previously, but it had been associated with disorders including diabetes, obesity and cardiovascular diseases [33, 34]. Interestingly, the analysis of 15 clinical and epidemiological factors in relation to the identified clusters found that the majority of cluster A patients were diabetic and had a higher BMI than the other study participants. This cluster also had a significantly higher prevalence of CHD than cluster C.

Previously, we found that both knee OA patients and diabetic patients had a lower concentration of two phosphatidylcholines (PC ae C34:3 and PC ae C36:3) than controls [35]. In the current study, we found that patients in cluster A were also associated with a phosphatidylcholine, PC ae C40:3. Although the statistical test only reached the pre-defined significance when comparing cluster A with controls, the strength of association became much stronger when examining the ratio of these two top metabolites, supporting a hypothesis that cluster A had an impaired fatty acid oxidation for energy production leading to insufficient energy levels for muscle functions. Thus, OA patients in cluster A would benefit from interventions or therapies that improve muscle strength.

Cluster B with 200 OA patients was characterized by a significant reduction of plasma free arginine concentration. Previous studies have reported that blood arginine levels were reduced by 24–31% in OA patients [36, 37]. An animal model of OA showed a significant reduction of arginine concentration after anterior cruciate ligament transection (ACLT) in the ACLT rabbit model with a negative association between the post-ACLT arginine concentration and severity of OA [38], suggesting the potential mechanism of the reduced arginine concentration in OA patients is increased demand of arginine for cartilage repair in OA [14] and inability of the body to meet the demand.

Further, arginine has antihypertensive and antioxidant properties, which influences blood viscosity and the coagulation system, and affects the metabolism of glucose, lipids and proteins [39]. Evidence shows that arginine intake in cardiovascular patients reverses endothelial dysfunction associated with major cardiovascular risk factors, such as hypercholesterolaemia, smoking, hypertension, diabetes, obesity, insulin resistance and ageing [40]. Also, multiple studies have reported an important role of arginine in improving blood flow in the arteries of the heart, which may improve symptoms of clogged arteries, chest pain or angina, and CHD [41]. Interestingly, our analysis is in agreement with the findings of these studies and showed that cluster B patients had a significantly higher prevalence of CHD especially relative to cluster C. Hence, supplementation of arginine might be beneficial to OA patients, particularly cluster B patients. Further, arginine is a natural inhibitor of

cathepsins such as cathepsin B and K, proteases that break down cartilage [35], and thus the depletion of arginine in cluster B OA patients reduces its inhibitory effect and leads to overactivity of cathepsin that in turn leads to cartilage breakdown. Functional studies are needed to confirm this.

Significant differences in the concentration of lysoPC a C16:0 contributed to the categorization of cluster C of 349 OA patients as distinct from other OA patients and controls. LysoPCs are bioactive lipids that contribute to a variety of cellular functions [42, 43], such as proliferation, apoptosis, smooth muscle contraction, wound healing and tumour cell invasiveness [44], and they have been reported to stimulate pro-inflammatory cytokines, such as IL-1β, TNF and IL-6, leading to the initiation and progression of OA [45]. The other possible pathway to produce lysoPCs from phosphatidylcholines uses the reactive oxygen species (ROS) in neutrophils [18]. Under the oxidative stress that is implicated in the pathogenesis of OA, the majority of OA joint cells can produce large amounts of ROS and NO in response to biomechanical or biochemical stimuli. Then, the mixture of proteolytic enzymes released from neutrophils combined with different ROS may stimulate the cartilage damage at the OA joints [18, 46].

Although the second metabolite, PC ae C38:2, only reached the pre-defined significance when comparing patients in cluster C with controls, the ratio of lysoPC a C16:0 to PC ae C38:2 was statistically significant. The ratio was significantly higher in all three identified OA clusters than in OA-free controls, which is in agreement with our previous studies in which we documented a significant increase of the lysoPCs to PCs ratio as being associated with knee OA risk [18], knee cartilage volume changes in 2-year follow-up [47], and an increased risk for undergoing total knee joint replacement in 10-year follow-up [18]. The ratio has also been found to be able to predict an OA patient's response to symptomatic drugs [48]. The elevated ratio indicated that the conversion pathway of PC to lysoPC was overactivated in OA and led to the stimulation of inflammation. Thus, the lysoPCs to PCs ratio has been suggested as a possible biomarker for monitoring anti-inflammatory treatment in rheumatoid arthritis as well [49]. Between identified clusters, the lysoPC a C16:0 to PC ae C38:2 ratio was significantly lower in cluster C patients compared with OA patients in the other clusters, suggesting that cluster C might have a lower level of inflammation. Further analysis found a significantly lower percentage of cluster C patients to have diabetes and CHD, but a higher percentage of these patients had osteoporosis compared with the other two clusters. These findings support a proposed hypothesis that bone loss might be an initiation factor for OA development at least for patients in cluster C [50]. Thus, identifying cluster C patients for clinical trials of anti-osteoporotic drugs in OA would be helpful for determining its efficacy.

There are a number of limitations in the study. The metabolic profiling was done with a commercially available metabolomics assay kit that has limited coverage of metabolites. Thus, we might have missed some metabolites that may contribute to the endotypes of OA. Also, study participants included both knee and hip OA patients. While knee and hip OA share a number of risk factors, the aetiology might be different between them. However, we did not find a different distribution of knee and hip OA in the identified three clusters. Further studies of a knee and hip OA cohorts with sufficient sample sizes are needed to confirm the findings. Similarly, prevalence of OA, particularly knee OA, was different between men and women; however, we did not find a sex difference in the three identified endotypes of OA. Different endotypes might have different severity of the disease or different observed characteristics such as muscle weakness in cluster A, but we did not have data on the severity or muscle strength, and further studies are needed to confirm these. Lastly, all the study participants were from Newfoundland and Labrador, which is a genetically and ethnically homogeneous population that may limit the generalizability of our results to other populations.

In conclusion, our data demonstrated that at least three distinct endotypes existed in primary OA, suggesting muscle weakness, arginine deficiency and low inflammatory OA subtypes that can be distinguished by specific blood metabolic markers. While confirmation is needed, these findings provide new insights into the understanding of OA pathogenesis and hold promise in developing personalized tools for OA management toward a reduction of economic burden and a better quality of life for OA patients.

Data availability statement

Data are available upon reasonable request by any qualified researchers who engage in rigorous, independent scientific research, and will be provided following review and approval of a research proposal and Statistical Analysis Plan (SAP) and execution of a Data Sharing Agreement (DSA). All data relevant to the study are included in the article.

References

1 Hiligsmann M, Cooper C, Guillemin F *et al.* A reference case for economic evaluations in osteoarthritis: an expert consensus article from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Semin Arthritis Rheum 2014;44: 271–82.

2 Cicuttini FM, Wluka AE. Osteoarthritis: is OA a mechanical or systemic disease? Nat Rev Rheumatol 2014;10:515–6.

3 Felson DT. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthritis Cartilage 2010;18: 601–4.

4 Van Spil WE, Kubassova O, Boesen M, Bay-Jensen A, Mobasheri A. Osteoarthritis phenotypes and novel therapeutic targets. Biochem Pharmacol 2019;165:41–8.

5 Herrero-Beaumont G, Roman-Blas JA, Castaneda S, Jimenez SA. Primary osteoarthritis no longer primary: three subsets with distinct etiological, clinical, and therapeutic characteristics. Semin Arthritis Rheum 2009; 39:71–80.

6 McGonagle D, Tan AL, Carey J, Benjamin M. The anatomical basis for a novel classification of osteoarthritis and allied disorders. J Anat 2010;216: 279–91.

7 Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011; 377:2115–26.

8 Scanzello CR, Umoh E, Pessler F *et al.* Local cytokine profiles in knee osteoarthritis: elevated synovial fluid interleukin-15 differentiates early from end-stage disease. Osteoarthritis Cartilage 2009;17:1040–8.

9 MacGregor AJ, Li Q, Spector TD, Williams FM. The genetic influence on radiographic osteoarthritis is site specific at the hand, hip and knee. Rheumatology (Oxford) 2008;48:277–80.

10 Haynes MK, Hume EL, Smith JB. Phenotypic characterization of inflammatory cells from osteoarthritic synovium and synovial fluids. Clin Immunol 2002;105: 315–25.

11 Zhang W, Likhodii S, Zhang Y *et al.* Classification of osteoarthritis phenotypes by metabolomics analysis. BMJ Open 2014;4:e006286.

12 Zhai G, Randell EW, Rahman P. Metabolomics of osteoarthritis: emerging novel markers and their potential clinical utility. Rheumatology 2018;57:2087–95.

13 Idle JR, Gonzalez FJ. Metabolomics. Cell Metab 2007;6: 348–51.

14 Zhai G. Alteration of metabolic pathways in osteoarthritis. Metabolites 2019;9:11.

15 Zhai G, Aref-Eshghi E, Rahman P *et al.* Attempt to replicate the published osteoarthritis-associated genetic variants in the Newfoundland & Labrador population. J Orthopedics Rheumatol 2014;1:5.

16 Altman R, Alarcon G, Appelrouth D *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.

17 Fontaine-Bisson B, Thorburn J, Gregory A, Zhang H, Sun G. Melanin-concentrating hormone receptor 1 polymorphisms are associated with components of energy balance in the Complex Diseases in the Newfoundland Population: Environment and Genetics (CODING) study. Am J Clin Nutr 2014;99:384–91.

18 Zhang W, Sun G, Aitken D *et al.* Lysophosphatidylcholines to phosphatidylcholines ratio predicts advanced knee osteoarthritis. Rheumatology 2016;55:1566–74.

19 Zhang W, Likhodii S, Aref-Eshghi E *et al.* Relationship between blood plasma and synovial fluid metabolite concentrations in patients with osteoarthritis. J Rheumatol 2015;42:859–65.

20 Zhai G, Wang-Sattler R, Hart DJ *et al.* Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. Ann Rheum Dis 2010; 69:1227–31.

21 Mittelstrass K, Ried JS, Yu Z *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet 2011;7:e1002215.

22 Williams B, Onsman A, Brown T. Exploratory factor analysis: a five-step guide for novices. Aust J Paramed 2010;8:1–13.

23 Thompson B. Factor analysis. The Blackwell Encyclopedia of Sociology 2007, doi: 10.1002/9781405165518.wbeosf003.

24 Dinning PG, Jones M, Hunt L *et al.* Factor analysis identifies subgroups of constipation. World J Gastroenterol 2011;17:1468–74.

25 Lawson RG, Jurs PC. New index for clustering tendency and its application to chemical problems. J Chem Inf Comput Sci 1990;30:36–41.

26 Andrew Ng. Clustering with the k-means algorithm. In: Blockeel H, ed. Mach Learning. USA: Kluwer/Springer, 2012. https://www.youtube.com/watch?v=hDmNF9JG3lo.

27 Petersen A, Krumsiek J, Wägele B *et al.* On the hypothesis-free testing of metabolite ratios in genome-wide and metabolome-wide association studies. BMC Bioinformatics 2012;13:120.

28 Williams CK, Engelhardt A, Cooper T *et al.* Package 'caret', 2015. http://topepo.github.io/caret/data-splitting.html.

29 Mels C, Jansen van Rensburg P, van der Westhuizen Francois H *et al.* Increased excretion of C4-carnitine species after a therapeutic acetylsalicylic acid dose: evidence for an inhibitory effect on short-chain fatty acid metabolism. ISRN pharmacology 2011;2011: 851870.

30 Nagan N, Kruckeberg KE, Tauscher AL *et al.* The frequency of short-chain acyl-CoA dehydrogenase gene variants in the US population and correlation with the C4-acylcarnitine concentration in newborn blood spots. Mol Genet Metab 2003;78:239–46.

31 Pedersen CB, Kølvraa S, Kølvraa A *et al.* The ACADS gene variation spectrum in 114 patients with short-chain acyl-CoA dehydrogenase (SCAD) deficiency is dominated by missense variations leading to protein misfolding at the cellular level. Hum Genet 2008;124:43–56.

32 O'Reilly S, Jones A, Doherty M. Muscle weakness in osteoarthritis. Curr Opin Rheumatol 1997;9:259–62.

33 Villarreal-Pérez JZ, Villarreal-Martínez JZ, Lavalle-González FJ *et al.* Plasma and urine metabolic profiles are reflective of altered beta-oxidation in non-diabetic obese subjects and patients with type 2 diabetes mellitus. Diabetology & metabolic syndrome 2014;6:129.

34 Makrecka-Kuka M, Sevostjanovs E, Vilks K *et al.* Plasma acylcarnitine concentrations reflect the acylcarnitine profile in cardiac tissues. Sci Rep 2017;7:17528.

35 Zhang W, Sun G, Likhodii S *et al.* Metabolomic analysis of human synovial fluid and plasma reveals that phosphatidylcholine metabolism is associated with both osteoarthritis and diabetes mellitus. Metabolomics 2016; 12:24.

36 Zhang W, Sun G, Likhodii S *et al.* Metabolomic analysis of human plasma reveals that arginine is depleted in knee osteoarthritis patients. Osteoarthritis and cartilage 2016;24:827–34.

37 Pascale V, Pascale W, Lavanga V *et al.* L-arginine, asymmetric dimethylarginine, and symmetric dimethylarginine in plasma and synovial fluid of patients with knee osteoarthritis. Med Sci Monit 2013;19: 1057–62.

38 Ohnishi A, Osaki T, Matahira Y *et al.* Correlation of plasma amino acid concentrations and chondroprotective effects of glucosamine and fish collagen peptide on the development of osteoarthritis. J Vet Med Sci 2013;75:497–502.

39 Tousoulis D, Antoniades C, Tentolouris C *et al.* L-arginine in cardiovascular disease: dream or reality? Vascular Medicine 2002;7:203–11.

40 Gornik HL, Creager MA. Arginine and endothelial and vascular health. J Nutr 2004;134(10 Suppl):2880–7S; discussion 2895S.

41 Sudar-Milovanovic E, Obradovic M, Jovanovic A *et al.* Benefits of L-arginine on cardiovascular system. Mini Rev Med Chem 2015;16:94–103.

42 Goetzl EJ. Pleiotypic mechanisms of cellular responses to biologically active lysophospholipids. Prostaglandins Other Lipid Mediat 2001;64:11–20.

43 Kabarowski JH, Xu Y, Witte ON. Lysophosphatidylcholine as a ligand for immunoregulation. Biochem Pharmacol 2002;64:161–7.

44 Hsieh C, Yen M, Liu H, Lau Y. Lysophosphatidylcholine induces apoptotic and non-apoptotic death in vascular smooth muscle cells: in comparison with oxidized LDL. Atherosclerosis 2000;151:481–91.

45 Leistad L, Feuerherm AJ, Faxvaag A, Johansen B. Multiple phospholipase A2 enzymes participate in the inflammatory process in osteoarthritic cartilage. Scand J Rheumatol 2011;40:308–16.

46 Hilbert N, Schiller J, Arnhold J, Arnold K. Cartilage degradation by stimulated human neutrophils: elastase is mainly responsible for cartilage damage. Bioorg Chem 2002;30:119–32.

47 Zhai G, Pelletier J, Liu M *et al.* Activation of the phosphatidylcholine to lysophosphatidylcholine pathway is associated with osteoarthritis knee cartilage volume loss over time. Sci Rep 2019;9:9648.

48 Zhai G, Pelletier J, Liu M *et al.* Serum lysophosphatidylcholines to phosphatidylcholines ratio is associated with symptomatic responders to symptomatic drugs in knee osteoarthritis patients. Arthritis Res Ther 2019;21:224.

49 Fuchs B, Schiller J, Wagner U, Häntzschel H, Arnold K. The phosphatidylcholine/lysophosphatidylcholine ratio in human plasma is an indicator of the severity of rheumatoid arthritis: investigations by ^{31}P NMR and MALDI-TOF MS. Clin Biochem 2005;38:925–33.

50 Bultink IE, Lems WF. Osteoarthritis and osteoporosis: what is the overlap? Curr Rheumatol Rep 2013;15:328.