Osteoarthritis and Cartilage

A clinical model including protein biomarkers predicts radiographic knee osteoarthritis: a prospective study using data from the Osteoarthritis Initiative

--Manuscript Draft--

Manuscript Number:	OAC10948R2
Article Type:	Manuscript
Keywords:	osteoarthritis; knee; biomarkers; proteomics; diagnosis; arthritis
Corresponding Author:	Francisco J. Blanco, MD INIBIC-Hospital Universitario A Coruña A Coruña, SPAIN
First Author:	Lucia Lourido
Order of Authors:	Lucia Lourido
	Cristina Ruiz-Romero
	Vanesa Balboa-Barreiro
	Ignacio Rego-Perez
	Maria Camacho-Encina
	Rocio Paz
	Valentina Calamia
	Natividad Oreiro
	Peter Nilsson
	Francisco J. Blanco, MD
Abstract:	Objective : We aimed to provide a model to predict the prospective development of radiographic KOA (rKOA). Method : Baseline sera from 333 non-radiographic KOA subjects belonging to OA Initiative (OAI) who developed or not, rKOA during a follow-up period of 96 months were used in this study. The exploratory cohort included 200 subjects, whereas the replication cohort included 133. The levels of inter-alpha trypsin inhibitor heavy chain 1 (ITIH1), complement C3 (C3) and calcyclin (S100A6), identified in previous large proteomic analysis, were analyzed by using sandwich immunoassays on suspension bead arrays. The association of protein levels and clinical covariates with rKOA incidence was assessed by combining logistic regression analysis, Receiver Operating Characteristic analysis, Integrated Discrimination Improvement (IDI) analysis and Kaplan-Meier curves. Results: Levels of ITIH1, C3 and S100A6 were significantly associated with the prospective development of rKOA, showing an area under the curve (AUC) of 0.713 (0.624-0.802), 0.708 (0.618-0.799) and 0.654 (0.559-0.749), respectively to predict rKOA in the replication cohort. The inclusion of ITIH1 in the clinical model (age, gender, BMI, previous knee injury and WOMAC pain) improved the predictive capacity of the clinical covariates (AUC=0.754 [0.670-0.838]) producing the model with the highest AUC (0.786 [0.705-0.867]) and the highest IDI index (9%). High levels of ITIH1 were also associated with an earlier onset of the disease. Conclusion: A clinical model including protein biomarkers that predicts incident rKOA has been developed. Among the tested biomarkers, ITIH1 showed potential to improve the capacity to predict rKOA incidence in clinical practice.

A clinical model including protein biomarkers predicts radiographic knee osteoarthritis: a prospective study using data from the Osteoarthritis Initiative

Lucía Lourido[1], Vanesa Balboa-Barreiro[2], Cristina Ruiz-Romero[1,3], Ignacio Rego-Pérez[1], María Camacho-Encina[1], Rocío Paz-González[1], Valentina Calamia[1], Natividad Oreiro[1], Peter Nilsson[4], Francisco J. Blanco[1,2]

Affiliations:

[1] Grupo de Investigación de Reumatología (GIR), Instituto de Investigación Biomédica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), Sergas. Universidade da Coruña (UDC). As Xubias de arriba 84, 15006, A Coruña, Spain.

[2] Universidade da Coruña (UDC), Grupo de Investigación de Reumatología y Salud (GIR-S). Departamento de Fisioterapia, Medicina y Ciencias Biomédicas, Facultad de Fisioterapia, Campus de Oza, 15008, A Coruña, Spain.

[3] Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN). Av. Monforte de Lemos, 3-5. Pabellón 11, 28029, Madrid, Spain.

[4] Division of Affinity Proteomics, Department of Protein Science, KTH Royal Institute of Technology, SciLifeLab, SE-100 44 Stockholm, Sweden.

*** Correspondence and request for materials should be addressed to:**
Dr. Francisco J. Blanco
As Xubias, s/n 15006. A Coruña, Spain

Telephone: +34981176399

Fax: +34981176398
email: fblagar@sergas.es

Abstract

Objective: We aimed to provide a model to predict the prospective development of radiographic KOA (rKOA).

Method: Baseline sera from 333 non-radiographic KOA subjects belonging to OA Initiative (OAI) who developed or not, rKOA during a follow-up period of 96 months were used in this study. The exploratory cohort included 200 subjects, whereas the replication cohort included 133. The levels of inter-alpha trypsin inhibitor heavy chain 1 **(**ITIH1), complement C3 (C3) and calcyclin (S100A6), identified in previous large proteomic analysis, were analyzed by using sandwich immunoassays on suspension bead arrays. The association of protein levels and clinical covariates with rKOA incidence was assessed by combining logistic regression analysis, Receiver Operating Characteristic analysis, Integrated Discrimination Improvement (IDI) analysis and Kaplan-Meier curves.

Results: Levels of ITIH1, C3 and S100A6 were significantly associated with the prospective development of rKOA, showing an area under the curve (AUC) of 0.713 (0.624-0.802), 0.708 (0.618-0.799) and 0.654 (0.559-0.749), respectively to predict rKOA in the replication cohort. The inclusion of ITIH1 in the clinical model (age, gender, BMI, previous knee injury and WOMAC pain) improved the predictive capacity of the clinical covariates (AUC=0.754 [0.670-0.838]) producing the model with the highest AUC (0.786 [0.705-0.867]) and the highest IDI index (9%). High levels of ITIH1 were also associated with an earlier onset of the disease.

Conclusion: A clinical model including protein biomarkers that predicts incident rKOA has been developed. Among the tested biomarkers, ITIH1 showed potential to improve the capacity to predict rKOA incidence in clinical practice.

Keywords: knee osteoarthritis, incidence, predictive model, ITIH1

Running title: Predicting incident knee osteoarthritis

1. Introduction

Osteoarthritis (OA) is one of the highest contributors to global disability and a major public health burden, mostly as a consequence of the knee OA (KOA) that is affecting almost 10% of worldwide population(1). KOA is a disease with a long asymptomatic period(2). Diagnosis is routinely based on clinical symptoms in combination with radiography, which is insensitive to measure molecular changes and thus the diagnosis occurs when the disease is already in advanced stages(3). KOA has been described to be associated with gender, age, obesity, and joint trauma(4-7). Nevertheless, it is a multifactorial disease in which a combination of both environmental and genetic factors interact(8), which turns incident KOA into a highly heterogeneous disease and difficult to predict by currently available imaging and clinical measurements.

Despite the urgency driven by its prevalence, impact of disability, and socioeconomic costs(9), current therapies available to treat OA are designed to provide symptomatic relief rather than prevent the disease or slow down joint destruction(10). Without a change in the OA management, the prevalence of KOA is expected to increase by 12.5 million over the next 40 years, secondary to the aging and the increasing rates of obesity in the general population(11).

One of the strategies to prevent KOA would be to identify biochemical markers and clinical models associated to the incidence and the progression of this disease(12). This would allow identifying molecular events on the early stages and stratifying individuals at high risk of developing the disease. Furthermore, this may open the possibility of future development of prevention and treatment strategies to avoid or delay the advance of the disease. Therefore, the identification of biochemical markers to predict the KOA incidence at an early stage has been the focus of much research over the past few years(13-15).

In a previous large-scale proteomic approach, we verified that the levels in serum of inter-alpha trypsin inhibitor heavy chain 1 (ITIH1), complement C3 (C3), and calcyclin (S100A6), among a total of 78 proteins analyzed, were significantly elevated in patients with established KOA compared to healthy controls(16). In the present study, we aimed to qualify the potential clinical endpoint of these three proteins by investigating if their levels in serum at baseline could predict the prospective occurrence of radiographic KOA (rKOA) and be useful to develop a model to predict incident rKOA in clinics.

2. Method

2.1 Study design and population

This work was carried out using serum samples and data from the Osteoarthritis Initiative (OAI) (https://nda.nih.gov/oai). This is a well-described multicenter prospective observational cohort study of knee OA (KOA). All methods were conducted according to the Declaration of Helsinki. An informed written consent was obtained from all participants before inclusion.

The current study included the analysis of serum at baseline from two prospective subcohorts of Caucasian subjects of the OAI defined as exploratory subcohort (N=200) and replication subcohort (N=133). The OAI participants were randomly selected from those available who fulfilled three inclusion criteria: non-radiographic KOA at baseline, defined as having a Kellgren and Lawrence (KL) grade equal or lower than one (K/L≤1) in at least one knee (target knee) at baseline; data of follow-up at 96-months and, all clinical variables at baseline. Clinical variables that may have predictive value for KOA development were selected based on the specific eligibility risk factor criteria of OAI and prior published evidence (17). A sample size of 200 participants allowed to estimate

an area under the curve (AUC) with a 95% confidence interval and a precision of ±0.04, assuming an allocation ratio of 1.3 between incident and non-incident subjects at 96 months follow-up in the entire OAI cohort, and an AUC of 0.7 for all the proteins under the study.

After 96-months of follow-up, two specific outcome groups, with one target knee per subject, were defined: radiographic KOA (rKOA) cases, that acquire relevant radiographic KOA (K/L≥2) during the follow-up; and no radiographic KOA (no rKOA) subjects, lacking the feature of relevant radiographic KOA at the end of the study. Total knee replacements or osteotomies were not considered as rKOA. For each patient, we defined the duration in time between the first visit and the date that incident rKOA was firstly recorded (12, 24, 36, 48, 72 or 96 months).

2.2 Protein analysis

Sandwich immunoassays on suspension bead arrays were developed to detect the proteins ITIH1, S100A6, and C3 separately in sera. Briefly, 2 μg of capture antibodies were coupled to $5x10^5$ activated color-coded magnetic beads (MagPlex- C, Luminex, Corp.) together with one bare bead (empty bead) and normal rabbit IgG. Efficient coupling was confirmed by using either R-phycoerythrine (PE)-labeled anti-rabbit or R-PE-labeled anti-mouse antibodies (Jackson ImmunoResearch). The detection antibodies were biotinylated in accordance with previously described protocols(18). Serum samples were centrifuged (3000 rpm for 3 min), diluted in assay buffer and heat-treated at 56 °C for 30 min. Then, a pipetting robot was used to add 45 μl of diluted samples to 5 μl beads in randomized layouts across four separate 96-well plates and combined into a 384-well microtiter plate. Following an overnight incubation, beads were washed and incubated with the corresponding biotinylated detection antibody for

1h. After 3x washing of beads, 0.5 µg/ml R-PE-labeled streptavidin (Life Technologies) in PBST was incubated for 20 min and beads were finally washed and measured in PBST using the FlexMap3D instrument (Luminex, Corp.) Signals corresponding to the levels of the proteins were reported as median of at least 30 beads per bead identity as median fluorescence intensities (MFI).

All assays were run blinded to the clinical information. The proteins measured in this study, the capture and detection antibodies, and their manufacturers, as well as the sample dilution for each immunoassay are listed in Supplementary Table 1.

2.3 Statistical analysis

Each protein measurement was normalized based on the fluorescence levels corresponding to the replicates of the control sample pools included within each plate. Samples with missing data were excluded from the data analysis. Receiver operating characteristic (ROC) curves were plotted to assess the area under the curve (AUC) with 95% confidence interval (CI).

Clinical data at baseline were obtained from the OAI database (https://data-archive.nimh.nih.gov/oai). We performed a univariate logistic regression analysis followed by a stepwise logistic regression analysis to define a clinical model to predict rKOA incidence.

The clinical model, the biomarker model, and the model combining clinical covariates and the biomarkers were performed on the exploratory subcohort. Model performance was examined on the replication subcohort. The performance between the exploratory and the replication subcohorts was compared using the following metrics of ROC curve analysis: AUC positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity. For calculating AUC, the probability threshold was set based

on highest sum of sensitivity with specificity. All discrimination measures were presented with 95% bootstrap confidence intervals, with bootstrapping techniques based on 2000 bootstrapped samples. Moreover, the integrated discrimination improvement (IDI), was included as a measure to assess performance of a prediction model. The comparison of the models was carried out through the comparison of the AUC, according to the methods described by DeLong (1988)(19) or Venkatraman and Begg (1996)(20). The roc.test function available in the pROC package of the R statistical software was used. Moreover, the integrated discrimination improvement (IDI), was included as a measure to assess performance of a prediction model. Turnbull's extension of the Kaplan-Meier curve to interval-censored data(21) was used to estimate the cumulative probability of incidence KOA over time (survival function) according to ITIH1 levels. Therefore, ITIH1 was included as a categorical variable and grouped in tertiles, considering low levels of ITIH1 at baseline as a reference group. The rest of the categories (intermediate and high) were subsequently compared with the reference group (low). An extended Cox proportional hazard model was also used for multivariable analysis(22) and the respective 95% confidence intervals (CI) for the hazard ratios (HR) were obtained.

All statistical analyses were done using IBM SPSS version 24 and R software version 3.4.254.

3. Results

We studied 200 participants (121 females and 79 males) in the exploratory subcohort and 133 participants (78 females and 55 males) in the replication subcohort. In the exploratory cohort, 86 participants were classified as rKOA patients after 96 months follow-up and 114 as no rKOA subjects, whereas in the replication cohort, 57 were

classified as rKOA subjects and 76 as no rKOA subjects. The descriptive clinical characteristics of study participants and protein levels for the three biomarkers at baseline are presented in Table 1 along with the number of cases for the binary outcome. The univariate analysis indicated that ITIH1, C3, and S100A6 were significantly associated with incident rKOA (Table 2). The ORs corresponding to the three biomarkers were the following: 1.6 (1.4-1.8; p=1.428E-9) for ITIH1, 1.2 (1.1-1.2; p=8E-6) for C3, and 1.2 (1.0-1.4; p=0.008) for S100A6. These results indicate that an increase of 100 units of fluorescence for ITIH1 raises the odds of incident rKOA up to 57.3%, whereas for C3 and S100A6 the increase on the odds is lower, 16.5% and 19.5% respectively. On the basis of receiver-operating characteristic (ROC) curve analysis we found that the three proteins were significant predictors of incident rKOA (Table 2). The highest predictive value for incident rKOA was observed for ITIH1 [AUC= 0.783; p=7.242E-12; 95% CI (0.718-0.849)]. C3 displays an AUC of 0.727 [p=3.787E-8; 95% CI (0.656-0.799)] whereas S100A6 shows an AUC of 0.605 [p=0.011; 95% CI (0.526-0.684)].

We next performed a univariate logistic regression analysis to study the individual effects of clinical covariates on the binary clinical outcome (rKOA *vs* no rKOA). Based on univariate models, eight variables met the $p<0.05$ threshold and were included in step-wise multivariable regression analyses: age, gender, BMI, frequent knee bending activity, alignment type (varus *vs* valgus), alignment degrees, history of knee injury and WOMAC pain score at baseline (Supplementary Table 2). Step-wise multivariable regression analysis resulted in the identification of a clinical model for incident rKOA which included complete clinical information of 200 participants from the exploratory cohort (Supplementary Table 3). The clinical model including age, gender, BMI, history of knee injury and WOMAC pain at baseline yielded an AUC of 0.816

[p=2.19E-14; 95% CI (0.756-0.875)] to discriminate rKOA and no rKOA cases. The same model tested in the replication subcohort yielded an AUC of 0.754 [p= 5,760E-7; 95% CI (0.670-0.838)].

The capacity of ITIH1, S100A6 and C3 to predict incident rKOA was firstly tested on the exploratory subcohort. The clinical relevance of ITIH1, S100A6 and C3 was assessed by comparing the AUCs between the different predictive models: the clinical model, the models including each biomarker alone, the model combining the three biomarkers, and the models including the clinical covariates (gender, age, BMI, history of knee injury and WOMAC pain) with the biomarkers, both separately or in combination (Figure 1A, 1B, 1C, 1D). Therefore, we observed that the clinical model showed higher ability than C3 and S100A6 to predict the rKOA development. However, no significant differences were observed between the predictive capacity of ITIH1 alone and the clinical model (p=0.453). The model combining the three biomarkers did not provide a significantly higher predictive capacity than the clinical model (p=0.599). No significant differences were either found between ITIH1 and the combination of the three biomarkers (data not shown). The addition of each protein separately as well as the addition of the three biomarkers improved the AUC of the clinical model. However, only the addition of the biomarker ITIH1 to the clinical variables (full model) improved the predictive value achieved by the clinical model alone (AUC=0.871; p=0.0056 ;95% CI [0.822-0.920]) in the exploratory subcohort. The characteristics of the ROC analysis for the different predictive models are specified in Supplementary Table 4. The parameters of the full model are provided in Supplementary Table 5.

The predictive models performed on the exploratory subcohort were tested on the replication subcohort (Supplementary Table 6). The AUCs observed in the replication

analysis for ITIH1, C3 and S100A6 were 0.713 (p=2.7E-5; 95% CI [0.624-0.802]), 0.708 (p=4E-5; 95% CI [0.618-0.799]) and 0.654 (p= 0.0026; 95% CI [0.559-0.749]), respectively to predict rKOA development. The model combining clinical covariates and ITIH1 (full model) showed the highest AUC (0.786; p=1.8148E-8; 95% CI [0.705-0.867]), although it was lower than the one obtained in the exploratory subcohort (AUC=0.871; p= 0.0056; 95%CI [0.822-0.920]) (Figure 2, Table 3).

IDI analysis was also performed to further explore the value of each biomarker within the clinical model in the replication subcohort. Therefore, we observed that the inclusion of ITIH1 in the clinical model produced a significant improvement of 9% (p= 0.002; 95% CI[3.4%-14.5%]) in its the predictive capacity, being less when introducing C3 and S100A6, and thus supporting the relevance of ITIH1 in the predictive model (Supplementary Table 6).

Exploring the differences on the predictive capacities of ITIH1, the clinical covariates and the full model between the exploratory and the replication subcohorts, we did not observe significant differences in terms of AUC (Figure 2, Table 3), indicating that the full model showed similar predictive capacity in both sample sets. For the full model the specificity and predictive values remained very similar in both exploratory and replication subcohorts.

We also explored the influence of ITIH1 levels at baseline on the time to rKOA development considering the total population study (N=333). With this aim, we categorized the variable ITIH1 in tertiles according to the different fluorescence levels that were observed: low (MFI<810.159), intermediate (810.159<MFI>1048.739) and high (MFI>1048.739) levels of ITIH1. In the Cox regression analysis, we observed that individuals with intermediate and high levels of ITIH have significantly higher risk to

develop rKOA (HR=1.75; 95% CI[1.04-2.97] and HR=3.058; 95% CI[1.86-5.03], respectively) compared to those that have low levels (Supplementary Figure 1A). The cumulative probability of incidence rKOA over time was estimated by each interval censured. Our results show that patients with high and intermediate levels of ITIH1 at baseline have a higher expected hazard to develop rKOA within 12 months compared to those with low levels, showing probabilities of 30.9%, 18.6% and 4.9% for high, intermediate and low levels of ITIH1, respectively. This trend was observed for all intervals of time analyzed (Supplementary Table 7). These results are depicted in a Kaplan-Meier curve (Supplementary Figure 1B).

4. Discussion

Here, we validate the association of serum ITIH1, C3, and S100A6 with KOA described in a previous work(16) and report for the first time the potential of these biomarkers to predict the prospective development of rKOA.

One of the features of a prognostic marker is the ability to predict the future occurrence of a certain disease(17). In the course of this prospective study, we showed that baseline levels of ITIH1, C3, and S100A6 were significantly associated with the development of rKOA as it shown in our univariate regression analysis (OR=1.6; OR=1.2; OR=1.2, respectively).

The ultimate aim of this work was to develop a tool to predict rKOA incidence combining clinical variables and biochemical markers. This tool would avoid the radiation of the patient and the time and costs consuming resources related. Accordingly, our model to predict rKOA does not include any radiological variable but other widely known clinical variables associated with the risk of KOA development namely age, gender, BMI, previous injury, and WOMAC knee pain at baseline(23).

This clinical model showed a great capacity to predict rKOA incidence (AUC= 0.816) in the exploratory subcohort but we aimed to improve it by the addition of any of the three biomarkers. In the process of searching for the best model performance to predict rKOA, we tested the prediction capacity of the three biomarkers. Therefore, the S100A6 only model showed a modest ability to predict radiographic KOA development (AUC<0.7) whereas ITIH1 and C3 showed an acceptable capacity (AUC>0.7) to identify subjects in risk of suffering rKOA. The combination of clinical covariates with the ITIH1 measurement at baseline resulted in the best performing model (full model) to predict the development of rKOA in an exploratory cohort of non-radiographic KOA subjects (N=200), yielding an AUC of 0.871. The predictive value of the full-model was replicated on another cohort of non-radiographic KOA subjects (N=133). Promisingly, the addition of ITIH1 increased the AUC of the clinical model from 0.754 (0.670-0.838) to 0.786 (0.704-0.867) and the IDI (9%) in the full model. As the AUCs confidence intervals of the clinical model and the full model have some overlapping, we cannot confirm that ITIH1 certainly improves the discriminatory accuracy of the clinical model in terms of AUC. However, our results provide information related to the potential ability of ITIH1 to improve the prediction capacity of the clinical model.

Clinical practice needs also prognostic tools for identifying individuals at risk of developing rKOA in the short term, which might allow new interventions targeting this population(24). In this regard, we provide data supporting the association of the baseline ITIH1 levels with the time of rKOA appearance. We show that subjects with the highest levels of ITIH1 at the beginning of the study hold the highest risk (HR=3.08) for an earlier onset of rKOA. This finding points out the potential usefulness of ITIH1 to identify subjects with high risk to develop the disease in the short term.

316 The limiting factors, such as the heterogeneity of the KOA condition, result in an
317 ineffective management of the disease(25). The early identification of subjects more
318 likely to develop the disease is crucial for handling tailored preventive and therapeutic
319 approaches(24). Despite much work on searching for prognostic biomarkers to predict
320 incident KOA and try to prevent its development, so far only two cartilage-derived
321 proteins, Urinary C-terminal telopeptide of collagen type II (uCTX-II), and serum
322 COMP, have been validated as protein biomarkers to be associated with the course of
323 this disease(26). Nevertheless, to date these biomarkers have not demonstrated to
324 achieve a sufficient value to predict KOA development in clinical practice(27).
325 Therefore, robust prognostic biomarkers to identify subjects in risk to develop KOA
326 remain to be identified. Regarding this, our study shows that ITIH1 is a prognostic
327 biomarker showing a potential capacity (AUC= 0.713 alone and AUC=0.786 in the full
328 model) to predict incident rKOA.

329 ITIH1 is a glycoprotein of the family of inter-alpha trypsin inhibitor (IαI) serum
330 proteins. This protein covalently binds the hyaluronic acid (HA) molecules through
331 their heavy chains (HC)(28, 29). The presence of HA-HC of serum IαI proteins in the
332 extracellular matrix (ECM) has been described to play an important role not only in the
333 stabilization of ECM(30, 31), but also in the onset of inflammation(32, 33). In the
334 context of OA, ITIH1 has been related to the early degradation process of OA articular
335 cartilage(34), being significantly increased in OA synovial fluid compared to RA(35),
336 and also elevated in serum from OA patients compared to RA and healthy
337 individuals(16). All these evidences indicated that ITIH1 is associated to OA pathology.
338 The present work, supported by the samples and data of OAI, validates therefore the
339 link between the protein ITIH1 and OA pathology. The clinical relevance of our work
340 is that we report for the first time the potential capacity of ITIH1 to identify individuals

at high risk to develop rKOA either alone or in combination with clinical covariates through a two-phase approach that consisted of screening and replication on two independent sample subcohorts.

Our results suggest that the measurement of circulating ITIH1 in combination with some clinical variables might have the potential to be incorporated in clinical practice to detect KOA in a pre-radiographic stage before radiographic and functional alterations in the joint integrity have occurred. Although further independent validation of the model in additional independent cohorts is required, this tool would open the window to target the potential therapeutic strategies on the high-risk individuals in order to delay the disease development.

Acknowledgements

The Osteoarthritis Initiative (OAI) is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Pfizer, Inc.; Novartis Pharmaceuticals Corporation; Merck Research Laboratories; and GlaxoSmithKline. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

This work has been supported by grants from *Fondo Investigación Sanitaria* (PI16/02124, PI17/00404, PI19/01206, DTS17/00200, CIBER-CB06/01/0040 and RETIC-RIER-RD16/0012/0002), integrated in the National Plan for Scientific Program, Development and Technological Innovation 2013–2016 and funded by the

ISCIII-General Subdirection of Assessment and Promotion of Research-European Regional Development Fund (FEDER) "A way of making Europe". This study has been also supported by grants IN607A2017/11, IN607D2020/10 and AE CICA-INIBIC (ED431E 2018/03) from Xunta de Galicia. The Proteomics Unit of GIR belongs to ProteoRed, PRB3- ISCIII (PT17/0019/0014). L.L. is supported by *Xunta de Galicia* (IN606B-2016/2005) and Contrato Sara Borrell (CD19/00229) Fondo de Investigación Sanitaria, ISCIII. I.R.P. is supported by Contrato Miguel Servet-II Fondo de Investigación Sanitaria (CPII17/00026). This work was also supported by the KTH Center for Applied Precision Medicine funded by the Ehrling Persson foundation, as well as the Human Protein Atlas project funded by Knut and Alice Wallenberg Foundation.

Author contributions

L.L, C.R-R. and F.J.B conceived and designed the study. L.L developed and validated the immunoassays with support from M.C-E, R.P, V.C and P.N. I.R-R and V.B-B performed the statistical analysis. L.L, C.R-R, N.O and F.J.B interpreted the data. L.L, C.R-R, V.B-B, P.N and F.J.B drafted the initial manuscript. All authors contributed to revision and editing of the manuscript and approved the version to be submitted.

Role of the funding sources

Funding sources had no role in the contents of this work.

Competing interest statement

We certify that there is no conflict of interest to disclose regarding the materials and data discussed in this manuscript. The contents of this manuscript have not been copyrighted or published previously.

References

1. Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, *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.
2. Rousseau JCh, Garnero P. Biological markers in osteoarthritis. Bone. 2012;51:265-77.
3. Kraus VB, Burnett B, Coindreau J, Cottrell S, Eyre D, Gendreau M, *et al*. Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis. Osteoarthritis Cartilage. 2011;19:515-42.
4. Jiang L, Tian W, Wang Y, Rong J, Bao C, Liu Y, *et al*. Body mass index and susceptibility to knee osteoarthritis: a systematic review and meta-analysis. Joint Bone Spine. 2012;79:291-7.
5. Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. Ann Rheum Dis. 2014;73:1659-64.
6. Chaganti RK, Lane NE. Risk factors for incident osteoarthritis of the hip and knee. Curr Rev Musculoskelet Med. 2011;4:99-104.
7. Blagojevic M, Jinks C, Jeffery A, Jordan KP. Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage. 2010;18:24-33.
8. Yucesoy B, Charles LE, Baker B, Burchfiel CM. Occupational and genetic risk factors for osteoarthritis: a review. Work. 2015;50:261-73.
9. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol. 2014;10:437-41.
10. Chu CR, Millis MB, Olson SA. Osteoarthritis: From Palliation to Prevention: AOA Critical Issues. J Bone Joint Surg Am. 2014;96:e130.
11. Driban JB, Sitler MR, Barbe MF, Balasubramanian E. Is osteoarthritis a heterogeneous disease that can be stratified into subsets? Clin Rheumatol. 2010;29:123-31.
12. Kraus VB, Blanco FJ, Englund M, Karsdal MA, Lohmander LS. Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. Osteoarthritis Cartilage. 2015;23:1233-41.
13. Saberi Hosnijeh F, Bierma-Zeinstra SM, Bay-Jensen AC. Osteoarthritis year in review 2018: biomarkers (biochemical markers). Osteoarthritis Cartilage. 2019;27:412-23.
14. Qvist P, Christiansen C, Karsdal MA, Madsen SH, Sondergaard BC, Bay-Jensen AC. Application of biochemical markers in development of drugs for treatment of osteoarthritis. Biomarkers. 2010;15:1-19.
15. Lotz M, Martel-Pelletier J, Christiansen C, Brandi ML, Bruyère O, Chapurlat R, et al. Value of biomarkers in osteoarthritis: current status and perspectives. Ann Rheum Dis. 2013;72:1756-63.
16. Lourido L, Ayoglu B, Fernández-Tajes J, Oreiro N, Henjes F, Hellström C, *et al*. Discovery of circulating proteins associated to knee radiographic osteoarthritis. Sci Rep. 2017;7:137.
17. Camacho-Encina M, Balboa-Barreiro V, Rego-Perez I, Picchi F, VanDuin J, Qiu J, *et al.* Discovery of an autoantibody signature for the early diagnosis of knee osteoarthritis: data from the Osteoarthritis Initiative. Ann Rheum Dis. 2019;78:1699-705.

18. Dezfouli M, Vickovic S, Iglesias MJ, Nilsson P, Schwenk JM, Ahmadian A. Magnetic bead assisted labeling of antibodies at nanogram scale. Proteomics. 2014;14:14-8.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.
20. Venkatraman ES , Begg CB. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. Biometrika.1996;84: 835–48.
21. Gomez Guadalupe CML, Oller R, Langhor K. Tutorial on methods for interval-censored data and their implementation in R. 2009;9: 259-97.
22. Wei P. Extending the iterative convex minorant algorithm to the Cox model for interval censored data.: Journal of Computational and Graphical Statistics 1999;8:109-20.
23. Heidari B. Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. Caspian J Intern Med. 2011;2:205-12.
24. Chu CR, Williams AA, Coyle CH, Bowers ME. Early diagnosis to enable early treatment of pre-osteoarthritis. Arthritis Res Ther. 2012;14:212.
25. Hussain SM, Neilly DW, Baliga S, Patil S, Meek R. Knee osteoarthritis: a review of management options. Scott Med J. 2016;61:7-16.
26. 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.
27. Valdes AM, Meulenbelt I, Chassaing E, Arden NK, Bierma-Zeinstra S, Hart D, *et al*. Large scale meta-analysis of urinary C-terminal telopeptide, serum cartilage oligomeric protein and matrix metalloprotease degraded type II collagen and their role in prevalence, incidence and progression of osteoarthritis. Osteoarthritis Cartilage. 2014;22:683-9.
28. Chen L, Mao SJ, McLean LR, Powers RW, Larsen WJ. Proteins of the inter-alpha-trypsin inhibitor family stabilize the cumulus extracellular matrix through their direct binding with hyaluronic acid. J Biol Chem. 1994;269:28282-7.
29. Lauer ME, Hascall VC, Green DE, DeAngelis PL, Calabro A. Irreversible heavy chain transfer to chondroitin. J Biol Chem. 2014;289:29171-9.
30. Mukhopadhyay D, Asari A, Rugg MS, Day AJ, Fülöp C. Specificity of the tumor necrosis factor-induced protein 6-mediated heavy chain transfer from inter-alpha-trypsin inhibitor to hyaluronan: implications for the assembly of the cumulus extracellular matrix. J Biol Chem. 2004;279:11119-28.
31. Murasawa Y, Nakamura H, Watanabe K, Kanoh H, Koyama E, Fujii S, *et al*. The Versican G1 Fragment and Serum-Derived Hyaluronan-Associated Proteins Interact and Form a Complex in Granulation Tissue of Pressure Ulcers. Am J Pathol. 2018;188:432-49.
32. Ni K, Gill A, Tseng V, Mikosz AM, Koike K, Beatman EL, *et al*. Rapid clearance of heavy chain-modified hyaluronan during resolving acute lung injury. Respir Res. 2018;19:107.
33. Yamaguchi Y, Noda H, Okaniwa N, Adachi K, Shinmura T, Nakagawa S, *et al*. Serum-Derived Hyaluronan-Associated Protein Is a Novel Biomarker for Inflammatory Bowel Diseases. Digestion. 2017;95:146-55.
34. Lourido L, Calamia V, Mateos J, Fernández-Puente P, Fernández-Tajes J, Blanco FJ, *et al*. Quantitative proteomic profiling of human articular cartilage degradation in osteoarthritis. J Proteome Res. 2014;13:6096-106.

35. Mateos J, Lourido L, Fernández-Puente P, Calamia V, Fernández-López C, Oreiro N, *et al*. Differential protein profiling of synovial fluid from rheumatoid arthritis and osteoarthritis patients using LC-MALDI TOF/TOF. J Proteomics. 2012;75:2869-78.

Tables

Table 1. Characteristics and covariates of the study participants.

	Exploratory subcohort (N=200)		Replication subcohort (N=133)	
Clinical covariates	**rKOA at 96 months (n=86)**	**No KOA at 96 months (n=114)**	**rKOA at 96 months (n=57)**	**No rKOA at 96 months (n=76)**
Age, years	60.7± 8.2	56.7 ± 8.4	60.3± 8.7	57.2± 8.7
Gender, female	59 (68.6)	62 (54.4)	38(52.6)	40(66.7)
BMI, kg/m2	29.2± 4.5	25.6± 3.4	28.6± 4.7	25.9± 4.1
Family history of knee replacement	12 (14.5)	17 (15.2)	7 (12.7)	7 (9.3)
Frequent knee bending activity, n (%) yes	67 (78.8)	58 (51.3)	39 (69.6)	52 (70.3)
Alignment, varus or valgus	15 (17.4) or 46 (53.5)	31 (37.8) or 24 (29.3)	19 (34.5) or 27 (49.1)	12 (25.5) or 19 (40.4)
Alignment valgus negative, degrees	–1.31 ± 3.7	0.15 ± 3.6	–0.78 ± 3.9	–0.55 ± 3.9
PASE	165.8 ± 83.8	183.12± 87.2	180.31± 101.11	156.48± 68.19
History of previous knee injury	23 (26.7)	17 (14.9)	15 (26.3)	9 (11.8)
History of knee surgery	9 (10.5)	9 (7.9)	9 (15.8)	7 (9.3)
Baseline WOMAC pain	1.65 ± 2.4	0.8 ± 2.04	1.82 ± 2.5	0.5± 1.2
Baseline JSW	4.7± 0.9	4.7± 0.8	4.7± 0.7	4.8± 0.8
ITIH1 (A.U)	1104.8±243.5	838.4±245.3	1091.5±297.8	916.1±546.4
C3 (A.U)	2059.5±465.5	1665.6±583	2028.8±580.9	1654.5±637.4
S100A6 (A.U)	808.6±230.7	720.3±219.1	852.8±226.6	714.8±216.4

Values are mean±SD or number of patients with percentage in parentheses.

BMI, Body mass index; PASE: Physical Activity Scale for the Elderly; WOMAC: Western Ontario and McMaster Universities Arthritis Index pain; JSW, joint space width; A.U, arbitrary units corresponding to the levels of the proteins

Table 2. Biomarker assessment. Univariate analysis and prediction capacity of the three biochemical markers analyzed in this study.

Meters	ITIH1	C3	S100A6
OR[a] (95% CI, pvalue)	1.6 (1.4-1.8; 1.428E-9)	1.2 (1.1-1.2; 8E-6)	1.2 (1.0-1.4; 0.008)
c-statistic (AUC) (95% CI)	0.783 (0.718-0.849)	0.727 (0.656-0.799)	0.605 (0.526-0.684)
Sensitivity % (95% CI)	68.6 (59.3-77.0)	46.5 (36.0-56.9)	81.4 (73.2-89.5)
Specificity % (95% CI)	80.7 (73.6-87.0)	90.3 (84.2-95.6)	36.8 (28.0-45.6)
PPV % (95% CI)	72.9 (65.4-80.0)	78.6 (67.9-88.6)	49.3 (45.0-53.8)
NPV % (95% CI)	77.2 (71.8-83.0)	69.1 (64.9-73.6)	72.5 (61.8-82.6)

OR, odd ratio CI, confidence interval; AUC, area under the curve; Sensitivity, Specificity; PPV, positive predictive value; NPV, negative predictive value

[a] OR per 100 units increase

Table 3. Predictive capacity of the ITIH1 model, the clinical model and the full model within the exploratory and replication subcohorts.

Models	Receiver-operating characteristic (ROC) curve	Exploratory subcohort	Replication subcohort	pvalues between the AUCs
ITIH1 model	**AUC (95% CI)**	0.783 (0.718-0.849)	0.713 (0.624-0.802)	0.214
	Sensitivity % (95% CI)	68.3 (59.3-77.0)	73.6 (61.4-84.2)	
	Specificity % (95% CI)	80.7 (73.6-87.0)	64.4 (53.9-75.0)	
	PPV % (95% CI)	72.9 (65.4-80.0)	60.8 (53.4-69.7)	
	NPV % (95% CI)	77.2 (71.8- 83.0)	76.6 (68.4-85.1)	
Clinical model	**AUC (95% CI)**	0.816 (0.756-0.875)	0.754 (0.670-0.838)	0.239
	Sensitivity % (95% CI)	84.8 (76.7-91.8)	71.9 (61.4-82.4)	
	Specificity % (95% CI)	67.5 (58.7-75.4)	71.5 (60.5-80.2)	
	PPV % (95% CI)	66.3 (60.3-72.8)	65.1 (56.6-74.1)	
	NPV % (95% CI)	85.7 (79.5-91.8)	77.1 (69.6-84.8)	
Full model	**AUC (95% CI)**	0.871 (0.822-0.920)	0.786 (0.704-0.867)	0.081
	Sensitivity % (95% CI)	81.4 (73.2-89.5)	54.4 (42.1-68.4)	
	Specificity % (95% CI)	80.7 (73.6-87.7)	94.7 (89.5-98.7)	
	PPV % (95% CI)	76.3 (69.2-83.5)	88.9 (78.1-97.2)	
	NPV % (95% CI)	85.2 (79.3-90.8)	73.4 (68.2-79.8)	

CI, confidence interval; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value

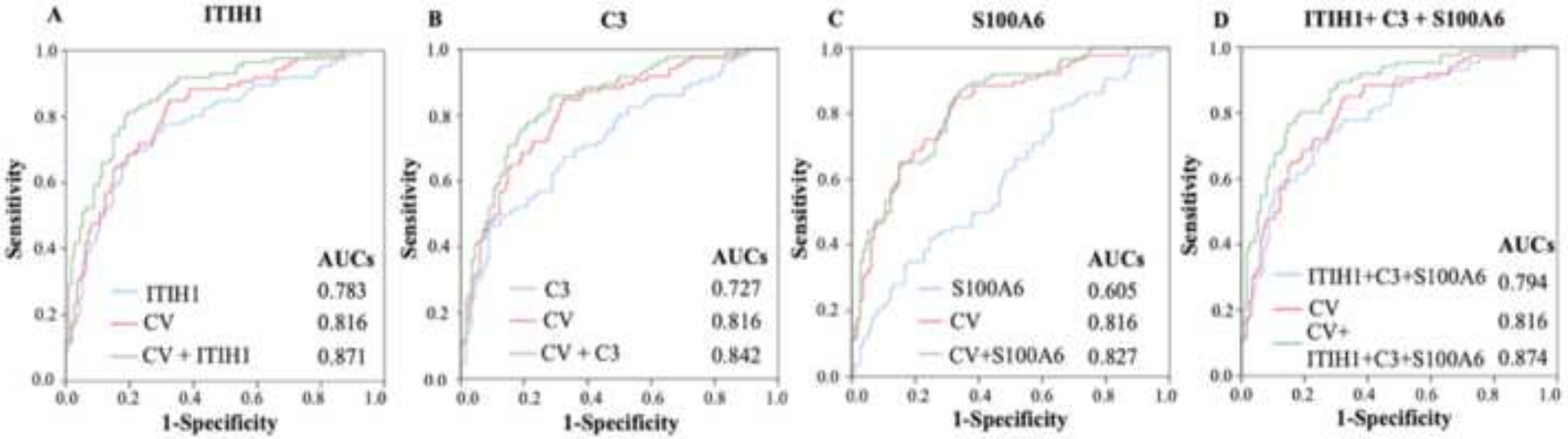
ITIH1 model includes only the biomarker ITIH1; Clinical model includes age, gender, BMI, history of knee injury and WOMAC pain; and full model includes the combination of the clinical model plus the biomarker ITIH1.

Figure legends

Figure 1. Receiver operating characteristic curve and area under the curves (AUCs) for the different predictive models within the exploratory subcohort. A) ITIH1 only, clinical covariates (CV) only and clinical covariates in combination with ITIH1 (B) C3 only, clinical covariates (CV) only and clinical covariates in combination with C3; C) S100A6 only, clinical covariates only and clinical covariates (CV) in combination with S100A6; D) Combination of ITIH1, C3 and S100A6, clinical covariates (CV) only and clinical covariates in combination with ITIH1, C3 and S100A6. Clinical covariates include age, gender, BMI, previous injury and WOMAC pain at baseline. The comparisons between the predictive capacities of the different models are represented in the tables below each graph. *$p<0.05$. NA: not applicable

Figure 2. Receiver operating characteristic curve and area under the curves (AUCs) for the different predictive models in the exploratory and replication subcohorts. ITIH1 only (blue line), only clinical covariates (red line) and clinical covariates (CV) in combination with ITIH1 (full model) (green line).

Figure 1



P	CV	CV+ ITIH1
ITIH1	0.453	0.001*
CV	NA	0.005*

P	CV	CV+ C3
C3	0.032*	0.002*
CV	NA	0.065

P	CV	CV+ S100A6
S100A6	2.51E-5*	2.30E-7*
CV	NA	0.275

P	CV	CV+ ITIH1+C3 +S100A6
ITIH1+C3 +S100A6	0.599	1.24E-3*
clinicals	NA	3.59E-3*

Figure 2