The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Swan, Anna L. and Hillier, Kirsty L. and Smith, Julia R. and Allaway, David and Liddell, Susan and Bacardit, Jaume and Mobasheri, Ali (2013) Analysis of mass spectrometry data from the secretome of an explant model of articular cartilage exposed to pro-inflammatory and anti-inflammatory stimuli using machine learning. BMC Musculoskeletal Disorders, 14 . 349/1-349/12. ISSN 1471-2474

see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact

BMC
Musculoskeletal Disorders

RESEARCH ARTICLE   **Open Access**

Analysis of mass spectrometry data from the secretome of an explant model of articular cartilage exposed to pro-inflammatory and anti-inflammatory stimuli using machine learning

Anna L Swan[1], Kirsty L Hillier[2], Julia R Smith[3], David Allaway[4], Susan Liddell[1,5,7], Jaume Bacardit[6,7,8†]
and Ali Mobasheri[2,7,9,10,11,12,13,14,15*†]

Abstract

Background: Osteoarthritis (OA) is an inflammatory disease of synovial joints involving the loss and degeneration of articular cartilage. The gold standard for evaluating cartilage loss in OA is the measurement of joint space width on standard radiographs. However, in most cases the diagnosis is made well after the onset of the disease, when the symptoms are well established. Identification of early biomarkers of OA can facilitate earlier diagnosis, improve disease monitoring and predict responses to therapeutic interventions.

Methods: This study describes the bioinformatic analysis of data generated from high throughput proteomics for identification of potential biomarkers of OA. The mass spectrometry data was generated using a canine explant model of articular cartilage treated with the pro-inflammatory cytokine interleukin 1 β (IL-1β). The bioinformatics analysis involved the application of machine learning and network analysis to the proteomic mass spectrometry data. A rule based machine learning technique, BioHEL, was used to create a model that classified the samples into their relevant treatment groups by identifying those proteins that separated samples into their respective groups. The proteins identified were considered to be potential biomarkers. Protein networks were also generated; from these networks, proteins pivotal to the classification were identified.

Results: BioHEL correctly classified eighteen out of twenty-three samples, giving a classification accuracy of 78.3% for the dataset. The dataset included the four classes of control, IL-1β, carprofen, and IL-1β and carprofen together. This exceeded the other machine learners that were used for a comparison, on the same dataset, with the exception of another rule-based method, JRip, which performed equally well. The proteins that were most frequently used in rules generated by BioHEL were found to include a number of relevant proteins including matrix metalloproteinase 3, interleukin 8 and matrix gla protein.

Conclusions: Using this protocol, combining an *in vitro* model of OA with bioinformatics analysis, a number of relevant extracellular matrix proteins were identified, thereby supporting the application of these bioinformatics tools for analysis of proteomic data from *in vitro* models of cartilage degradation.

Keywords: Osteoarthritis, Cartilage, Biomarker, Interleukin 1 β, Carprofen, Bioinformatics, Machine learning

* Correspondence: ali.mobasheri.manuscripts@gmail.com
†Equal contributors
[7]The D-BOARD European Consortium for Biomarker Discovery, University of
Nottingham, University Park, NG7 2RD Nottingham, UK
Full list of author information is available at the end of the article

Background

Articular cartilage is a mechanically resilient connective tissue with unique load-bearing and shock-absorbing properties, which are largely dependent on the structural and functional integrity of its highly charged and hydrated extracellular matrix (ECM) [1]. Cartilage contains three principal components: chondrocytes, aggregating proteoglycans and collagens, all of which are embedded within the ECM and contribute to the homeostasis of the tissue [2]. Cartilage relies on oxygen and nutrient delivery from the synovial fluid [3] but is avascular and recalcitrant to repair [4]. Osteoarthritis (OA) is a degenerative disease of synovial joints, involving the loss of articular cartilage, synovial inflammation and changes to the subchondral bone, resulting in impaired articulation, reduced mobility, joint stiffness and pain [5,6]. OA is estimated to affect up to 85% of the human population over 60 years old [7] and is also common in companion animals [8]. There are a number of factors affecting OA, including age, obesity, previous joint trauma or instability, metabolic or endocrine disease and oestrogen status [9,10]. Currently, diagnosis is made through clinical examination and the imaging "gold standard", radiography. However, radiographic diagnosis of OA is usually made when the clinical signs of pain and loss of mobility have already appeared. Consequently, the disease can remain undiagnosed until the later stages, where interventions may not alter the course of progression.

Biomarkers have the capacity to identify early changes in joint tissues and diagnose OA during the pre-radiographic stages of the disease and to determine the course of its progression, as well as aid in drug discovery and clinical trials [11-15]. The term biomarker can be used to describe molecules or molecular fragments that indicate the presence of a biological or disease process. Early detection may also help prioritize treatments to slow progression, such as weight loss and a reduction in high impact load bearing on those joints [16]. Therefore, individual or combination biomarkers must be able to clearly differentiate between healthy and diseased states. Ideally biomarkers should be disease-specific and not be influenced by other disorders. Biomarkers should also be easily measurable in a clinical setting [17]. In rheumatology, biomarkers can be "tissue fingerprints" or combinations of "neo-epitopes", reflecting catabolic effects downstream of inflammatory signals.

Recent advances in post-genomic technologies, including genomics, transcriptomics, proteomics and metabolomics, have allowed the development of novel methods for identification of biomarkers of disease. Proteomics is a particularly promising technology as it allows the identification of individual proteins and their peptides, neo-epitopes and degradation "fingerprints". This information can then be used to develop sensitive, rapid antibody-based assays. In addition proteomic analyses provide an overview of changes in the proteome in biological systems across a range of conditions [18].

Through the combined use of proteomics, transcriptomics and other biochemical and immunological techniques, a number of proteins and protein families have previously been associated with OA. These include ECM proteins such as aggrecan, the major structural proteoglycan found in the cartilage ECM, cartilage oligomeric matrix protein (COMP), a non-collagenous protein involved in the organization and assembly of articular cartilage, and matrix metalloproteinases (MMPs), a family of proteins expressed by chondrocytes, which are involved in the degradation of ECM macromolecules and lead to the fibrillation of articular cartilage [11,19-24]. In the ECM, matrix metalloproteinase-3 (MMP-3) in particular appears to be vital for matrix turnover and homeostasis. This protein is up-regulated in early OA, but has been found to be down-regulated in later stages of the disease [25].

Many omics technologies, such as microarrays, next generation sequencing and mass spectrometry (MS), generate large amounts of data. Therefore, bioinformatic tools play an important role in the analysis of such data and a wide range of methods have been developed for this purpose [26,27]. Supervised machine learning techniques are used, based on a training set of labelled samples, to build models that are able to automatically label previously unclassified samples [28,29]. Samples can be assigned a label (e.g. a treatment group) based on whether or not they contain a certain attribute (e.g. a protein, or a group of proteins) and at what level the attribute is found within the samples [30,31]. There are many types of machine learning techniques, such as decision trees, rule-based learners and support vector machines [28,32]. Rule-based machine learning methods automatically produce human-readable production rules that assign samples to their respective treatment groups. In proteomics-based approaches, the rules created contain proteins that best divide the samples into disease or treatment groups. Proteins most consistently differing between groups are suitable for further investigation as potential biomarkers.

The aim of this study was to identify suitable bioinformatic methods for the analysis of proteomics data generated to investigate cytokine-induced catabolic changes associated with the early stages of OA [33]. This involved using an explant model of cartilage to investigate the secretome of canine articular cartilage. The cartilage explant model was selected because it allows a rapid and 'clean' analysis of secreted proteins in the context of joint disease. Many of the proteins present in the secretome of explant cultures are involved in the control of physiological and pathophysiological processes in the

joint [34] and may enter the blood stream where they may be accessible as systemic biomarkers.

Methods

Animal tissues and statement of ethical approval

Forelimbs and hind limbs were taken from male German Shepherd army dogs, over 5 years of age, that were euthanized for clinical reasons unrelated to research. Therefore, this project does not fall under the Animals (Scientific Procedures) Act 1986[a] or the Veterinary Surgeons Act 1966[b]. Approval for the use of clinical materials was obtained from the Ethics Committee of the School of Veterinary Science and Medicine with input from members of the University of Nottingham's Animal Welfare and Ethical Review Body (AWERB). The British Army owned the animals that were used in this study. Informed consent was obtained for the use of joint tissues.

Cartilage explant culture

Limbs were washed in disinfectant and soaked in sodium hypochlorite prior to spraying with ethanol. The stifle and elbow joints were dissected under sterile conditions and full thickness articular cartilage was placed in serum free collection media. The media consisted of Hyclone® liquid medium: DMEM supplemented with penicillin and streptomycin.

After washing the harvested cartilage, a 3 mm biopsy punch was used to cut discs, which were placed in a randomized manner into wells of a 24 well plate, containing serum free DMEM (as above). The media was removed and the explants were incubated in media alone (control), or supplemented with recombinant canine IL-1β (10 ng/ml), the non-steroidal anti-inflammatory drug carprofen (Rimadyl®, 100 μg/ml), or carprofen and IL-1β combined (100 μg/ml and 10 ng/ml, respectively). For each treatment, three samples were used per dog, giving six samples per treatment. After 5 days in culture, supernatants and explants were removed and processed for mass spectrometric analysis.

Sample preparation and mass spectrometry

Samples from 2 dogs were chosen for MS/MS analysis based on the general profile of proteins as visualized on SDS-PAGE (data not included in the manuscript, see Additional file 1: Table S1; Additional file 2: Table S2; Additional file 3: Table S3; Additional file 4: Table S4; Additional file 5: Figure S1; Additional file 6: Figure S2). Each set of dog samples consisted of three treatments (IL-1β, carprofen, IL-1β + carprofen), with three replicates for each treatment for both dogs. A set of control samples was also analyzed, providing a total of 24 samples (12 samples per dog).

The secretome samples were digested with trypsin before mass spectrometry. Soluble proteins were reduced by the addition of DTT to a final concentration of 10 mM to each sample. The thiol groups were blocked by the addition of iodoacetamide to a final concentration of 55 mM. The proteins were then precipitated with ice-cold acetone before being suspended in trypsin solution (10 ng/μl in 50 mM ammonium bicarbonate) (Trypsin Gold, Mass Spectrometry Grade, Promega). Trypsin digestion was terminated by addition of formic acid to give a final concentration of 0.1%. Before MS analysis, an aliquot of the digestion was desalted and any insoluble particulates removed using a C18 Zip-Tip (Millipore).
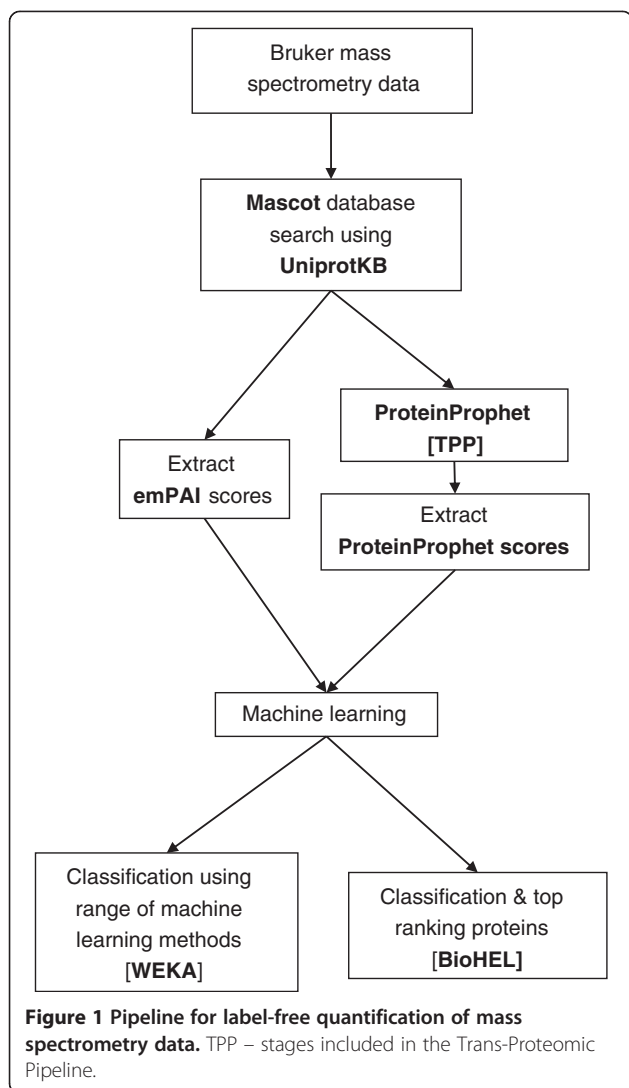
Peptides were separated on a 15 cm C18 PepMap™ column (LC Packings) using a Bruker Easy-nLC platform with a flow rate of 300 nl/min. The sample was added to solvent A (95% v/v H$_2$O, 5% v/v ACN, 0.1% v/v formic acid) and was injected into the HPLC column via the autosampler. Following binding and washing of the sample on the column in solvent A, peptides were separated and eluted in a gradient of solvent B (95% v/v ACN, 5% v/v H$_2$O, 0.1% v/v formic acid).

Eluted peptides were delivered on-line and detected in a Bruker AmaZon ETD ion trap instrument. The five most abundant peptides in each MS scan were selected for fragmentation. The raw data were processed to provide peptide and fragment mass lists which were submitted to the MS/MS ions tool of the Mascot search engine, software which uses protein sequence databases to predict the identity of proteins present in samples, based on the peptides identified. The fragment mass values for each peptide were compared to the mammalian entries from the UniProtKB database. The modifications incorporated into the search were: fixed carbamidomethyl cysteine and variable oxidation of methionine.

One sample, treated with both IL-1β + carprofen, was removed from the dataset at this stage as it was considered to be anomalous due to the very small number of proteins that were identified from it by Mascot. This resulted in 23 samples for further analysis: six samples per treatment, except for IL-1β + carprofen, for which there were five samples.

Further MS data analysis pipeline

The pipeline for the analysis of mass spectrometry data is described in Figure 1. Included in the results generated by Mascot is the exponentially modified protein abundance index (emPAI) score for each protein identified. The emPAI score gives an estimate for the absolute amount of a protein present in a sample [35]. It is based on the protein abundance index (PAI), which is defined as 'the number of peptides identified divided by the number of theoretically observable tryptic peptides' [36]. PAI was then adapted to emPAI to ensure it is proportional to the total protein content in a sample [35].

Bruker mass
spectrometry data

↓

Mascot database
search using
UniprotKB

ProteinProphet
[TPP]

Extract
emPAI scores

Extract
ProteinProphet scores

↓

Machine learning

↓

Classification using
range of machine
learning methods
[WEKA]

Classification & top
ranking proteins
[BioHEL]

**Figure 1 Pipeline for label-free quantification of mass
spectrometry data.** TPP – stages included in the Trans-Proteomic
Pipeline.

Mascot outputs were also submitted to ProteinProphet [37,38], part of the Trans-Proteomic Pipeline [39], used for the statistical validation of protein identifications. Using ProteinProphet a probability score is assigned to each of the protein identifications that was made by Mascot. Strictly speaking, ProteinProphet is not a true quantification method, but the probability scores that it produces are roughly equivalent to a quantitative approach. Therefore it is suitable for further analysis using machine learning techniques.

Both the emPAI and ProteinProphet scores were generated. Machine learning was applied to these datasets using a number of methods from the WEKA machine learning package [32] and BioHEL, a rule based learner [31].

Comparison of machine learning techniques

To determine the most suitable machine learning method for the analysis of canine articular cartilage mass spectrometry dataset, seven different machine-learning

techniques, including BioHEL, were applied to compare their abilities. The other methods used were Naive Bayes, Support Vector Machines, C4.5, IBk, JRip and Random Forest, all implemented in WEKA. The source code and user manual for BioHEL are available at http://ico2s.org/software/biohel.html.

Due to some anomalous identifications in the Mascot results for one of the carprofen + IL-1β samples, where only a very small number of proteins were identified compared to the other samples, it was removed from the dataset. This resulted in a dataset of 23 samples, spanning four treatment classes. As a result of this small number of samples, leave-one-out cross validation was used to divide them into training and test sets [40]; using this method allows for the most information to be extracted from the data available. Twenty-three training sets and the same number of test sets were created. The test sets each contained only one sample, with the remainder of the dataset in the related training set. This allows the ability of classification models to be evaluated.

Significance testing

The significance of the BioHEL classification accuracies achieved was tested by calculating *p*-values using one-tailed permutation testing [41]. A new version of the dataset was created where the samples were randomly assigned to treatments, but maintaining the same number of samples per treatment as in the original data. Afterwards, BioHEL was run, using leave-one-out cross-validation, to compare the accuracies achieved; 50 such permutations were generated for the emPAI, ProteinProphet and combined datasets. The accuracies achieved by these runs were compared to the accuracies achieved on the real, non-randomized, datasets and a *p*-value of the likelihood that the accuracy on the original data belongs to the randomized distribution was computed.

Identification of top ranking proteins

Due to the performance of BioHEL in the comparison with other machine learning methods, analyses using BioHEL continued through the identification of proteins that were pivotal to the classification, using a methodology previously used for the analysis of transcriptomics data [42,43].

BioHEL classifies samples by automatically producing rules sets that consist of a number of rules that use the proteins found in the samples to determine which treatment group they belong to. Each rule within a set uses proteins, when used with mass spectrometry data, to assign samples to treatment classes. A rule within a set uses one or more proteins and assigns samples to the relevant class, shown at the end of rule, if it matches

exactly the protein content specified by the rule. An example of a rule set for this data follows:

1. If the abundance of TPIS is greater than 0.01 then the sample belongs to the IL-1β group
2. If the abundance of IL-8 is greater than 0.02 then the sample belongs to the carprofen+IL-1β group
3. If the abundance of MMP-3 is greater than 0 and the abundance of UBIB is less than 0.2 then the sample belongs to the IL-1β group
4. If the abundance of MGP is greater than 0 and the abundance of A1AT is less than 0.9 then the sample belongs to the carprofen group
5. If the abundance of ALBU is greater than 0.01 then the sample belongs to the carprofen+IL-1β group
6. Any sample not assigned to a group belongs to the control group

The combinations of rules in the rule sets are used to assign samples to their respective treatment groups. Each rule contains one or more proteins and a score (either emPAI or ProteinProphet), which each protein should either be above or below, depending on the sign used. At the end of each line is the treatment class to which each rule relates. For example, the 1st rule of the rule set shown classifies all samples as belonging to the IL-1β class if the value of the protein attribute TPIS is greater than 0.01. There are no rules for the control: all samples that are not assigned to the other three classes by the rules generated will be, by default, considered as a control sample.

Due to the stochastic nature of BioHEL, running it multiple times on the same dataset produces different rule sets. Therefore BioHEL was run 10,000 times to analyze the results and determine recurrent patterns. Proteins were ranked by the number of times they appeared in rules across the 10,000 runs, to highlight those proteins used most frequently. Those ranking at the top are proteins that can be used to most successfully identify between samples of different treatments. As these proteins are the most different between treatment classes, they may be suitable for consideration as biomarkers or further analysis of them may provide information about possible novel methods for diagnosis or treatment.

Network generation

To investigate interactions between proteins with our prediction model we used network analysis, by identifying proteins that were working together in rules generated by BioHEL. Within rules generated by BioHEL protein pairs can be identified, from which networks were generated. These networks can be used to identify relationships between proteins; they also provide a visual way of viewing those proteins that are frequently in rules

through identification of the most connected proteins. In the example of a rule set, shown in "**Identification of top ranking proteins**" **subsection**, there are some rules that use more than one protein; these were used to form protein pairs. For example, in the third rule both apolipoprotein E (APOE) and hyaluronan and proteoglycan link protein 1 (HPLN1) are used and so are considered a protein pair. The 100 protein pairs that were most frequently used within rules, for each individual treatment class, across the 10,000 runs of BioHEL were extracted and a network was generated from them in Cytoscape [44]. The networks consist of nodes that relate to the proteins, found in the BioHEL rules, and edges connect proteins if they were frequently included in rules together. The edges were then coloured based on the treatment class that each pair of proteins relates to.

Results and discussion

Proteomic techniques are increasingly being used for the identification of novel joint disease biomarkers [11,19,20,45]. This study tests the hypothesis that the secretome of canine articular cartilage may provide a simple but well-defined model for studying potential biomarkers of early cartilage damage. To study the secretome of canine articular cartilage in an explant model we used a combination of conventional and high throughput proteomic techniques, followed by the application of bioinformatics techniques.

Although the cartilage explant system has not been used extensively in proteomic studies, a similar equine explant model of articular cartilage has been used to examine changes in the secretome in response to pro-inflammatory and anti-inflammatory stimuli [33]. This present study indicates that canine cartilage explants can also serve as a model for targeted and high throughput proteomic studies. This is supported by the identification of a large number of proteins whose functions are relevant to articular cartilage and biological processes that are relevant to joint disease and OA. Using the explant model this study has demonstrated it is feasible to incorporate pathophysiologically relevant stimuli such as pro-inflammatory cytokines (e.g. IL-1β) to simulate catabolic changes as well as NSAIDs (e.g. carprofen) to simulate pharmacotherapy in a well-controlled model *in vitro*.

The SDS-PAGE protein profiles of the IL-1β stimulated samples illustrate that some proteins are present at a higher level of abundance in the presence of IL-1β. This was demonstrated by the presence of extra bands in the IL-1β treated samples that were not detected in the controls (see Additional file 1: Table S1; Additional file 2: Table S2; Additional file 3: Table S3; Additional file 4: Table S4; Additional file 5: Figure S1; Additional file 6: Figure S2). There was also general consistency in protein

Table 1 Comparison of performance accuracies, as percentage of samples correctly classified, for classification of canine articular cartilage data for seven different machine-learning methods, using leave-one-out cross-validation

Dataseter	Naive Bayes	SVM	k-nearest neighbour	JRip (rule based)	Random forest	C4.5	BioHEL
ProteinProphet	39.1	52.2	34.8	43.5	34.8	52.2	**73.9**
emPAI	52.2	56.5	52.2	**78.3**	39.1	73.9	56.5
ProteinProphet and emPAI combined	52.2	52.2	43.5	**78.3**	26.1	47.8	**78.3**

For the 'ProteinProphet and emPAI combined' the two scores were combined into one dataset. The highest accuracies achieved in each dataset are shown in bold.

profiles across all groups of treated samples for the two animals (see Additional file 1: Table S1; Additional file 2: Table S2; Additional file 3: Table S3; Additional file 4: Table S4; Additional file 5: Figure S1; Additional file 6: Figure S2).

A range of machine learning methods were compared and BioHEL proved to be successful in classifying both the ProteinProphet and emPAI datasets. The accuracies of the range of machine learning techniques tested on the canine articular cartilage data are shown in Table 1. For the BioHEL classifications on each dataset, confusion matrices (that identify, treatment by treatment, how the samples were predicted) were generated to understand which samples were predicted correctly. It can be seen from the matrices for the emPAI, ProteinProphet and combined datasets (Figure 2) that the most frequent incorrect prediction made was predicting control samples as carprofen treated samples. This is due to the similarity between these groups, as carprofen was added in the absence of IL-1β and thus there was no pro-inflammatory present for this NSAID to act on. No IL-1β sample was predicted as a control sample. From

Table 1, it can be seen that BioHEL achieves the highest accuracies for both the ProteinProphet and the dataset that combines both emPAI and ProteinProphet scores; because of this, analysis was continued using BioHEL. The classification was increased by the combination of these two scoring systems. The significance of the BioHEL classification accuracies was supported by the *p*-values, calculated using permutation testing, shown in Table 2, as they were all close to zero. The outcome of this test confirms that the models generated by BioHEL (rule sets) are sound and hence we can safely analyze them to extract rankings of important variables and generate interaction networks.

From the rules generated by BioHEL, the top ranking mammalian proteins for the three treatments are shown in Tables 3 and 4. There is no ranking for the control class because it was used as the default and so did not include any proteins in rules. The default is included at the end of a rule set, so that any sample that has not been assigned to a class by the rules in the set is automatically placed into the default class. Table 5 shows the top ranking mammalian proteins for the emPAI and

emPAI dataset		Predicted treatment			
		Control	IL-1β	Carprofen	Carprofen +IL-1β
		--			
Real treatment	Control	1	0	5	0
	IL-1β	0	4	1	1
	Carprofen	4	0	2	0
	Carprofen +IL-1β	0	0	0	5
ProteinProphet dataset		Control	IL-1β	Carprofen	Carprofen +IL-1β
		--			
Real treatment	Control	2	0	4	0
	IL-1β	0	4	0	2
	Carprofen	0	0	6	0
	Carprofen +IL-1β	0	0	0	5
Combined dataset		Control	IL-1β	Carprofen	Carprofen +IL-1β
		--			
Real treatment	Control	3	0	3	0
	IL-1β	0	4	0	2
	Carprofen	0	0	6	0
	Carprofen +IL-1β	0	0	0	5

Figure 2 Confusion matrices, for the emPAI, ProteinProphet and combined datasets, to show the number of samples in each class and which class they were predicted to be in, using BioHEL.

Table 2 *P*-values generated by significance testing of BioHEL for the emPAI, ProteinProphet and combined datasets

Dataset	*P*-value
emPAI	1.64E-100
ProteinProphet	1.23E-220
ProteinProphet and emPAI combined	0

P-values were generated using a one-tailed permutation test. Classifications were run on datasets with randomized sample labels, using leave-one-out cross-validation; this was performed 50 times.

ProteinProphet combined datasets. It shows that both the emPAI and ProteinProphet scores were useful in the classification as some proteins, including triosephosphate isomerase, MMP-3, IL-8 and HPLN1, are top ranking proteins using both emPAI and ProteinProphet values.

The interaction network generated from the Protein-Prophet probabilities is shown in Figure 3. There are particular proteins (those most connected to other proteins) that can be identified from the network. These proteins include matrix metalloproteinase-3 (MMP-3), interleukin 8 (IL-8), HPLN1, matrix gla protein (MGP) and APOE, and are detailed in Table 6. The interaction network generated from the emPAI scores is shown in Figure 4. In this network there are fewer highly connected proteins, than in the ProteinProphet network, although MMP-3 and IL-8 are again connected to many other proteins. The fewer highly connected proteins in the emPAI network could be due to some proteins having similar emPAI scores but differing ProteinProphet probabilities. Therefore, where in the ProteinProphet network only one protein was suitable, in the emPAI network multiple proteins gave the same results and were interchangeable.

COMP is a noncollagenous ECM protein that is abundantly expressed in articular cartilage and which has been considered by other groups as a possible marker of articular cartilage degradation. This protein was not included in any top ranking protein lists, or in either network generated, because COMP was found at similar levels across all samples, regardless of the type of treatment. Therefore the bioinformatics methods discussed here are useful in determining proteins that may be suitable for use as putative biomarkers, rather than simply proteins that are abundant. We also expected to detect MMPs, a family of proteins expressed by chondrocytes with roles in cartilage development, remodelling and disease [54]. Matrix metalloproteinase-3 (MMP-3), a surrogate biomarker of psoriatic and rheumatoid arthritis [55,56], was pivotal in the classification of IL-1β samples. MMP-3 is a proteolytic enzyme known to degrade components of the ECM, including collagens and cartilage proteoglycans and, as a result, was the top ranking protein for the IL-1β class. No other MMPs were highlighted by the bioinformatics techniques applied. interleukin 8 (IL-8) was dominant in the classification of IL-1β and carprofen treated samples. IL-8 is the major chemotactic factor released in response to pro-inflammatory cytokines in synovial tissues from rheumatoid arthritis and osteoarthritis affected joints [57-59]. Matrix gla protein, involved in inhibition of calcification in cartilage [51], was also frequently found in the Bio-HEL rules from the analysis of the ProteinProphet dataset. This protein was found in many samples across all treatment groups, except for the carprofen + IL-1β group. MMP-3, IL-8 and MGP were also the most connected proteins in the ProteinProphet network. The inclusion of proteins such as these in the top ranking lists and as the most connected proteins, demonstrates the abilities

Table 3 The ten mammalian proteins found most frequently in rules for each of the three classes, not including the default control class, from the ProteinProphet dataset

IL-1β	Carprofen	IL-1β and Carprofen
Matrix metalloproteinase 3 (47.5)	Matrix gla protein (52.5)	interleukin 8 (93.1)
Triosephosphate isomerase (19.6)	Apolipoprotein E (44.8)	Matrix metalloproteinase 3 (6.87)
Enolase (4.56)	Hyaluronan and proteoglycan link protein 1 (22.7)	Pyruvate kinase isozyme M1 (5.59)
interleukin 8 (2.61)	Target of Nesh-SH3 (5.24)	Protein S100-A1 (5.39)
Leukocyte antigen CD37 (2.58)	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN (5.1)	Bardet-Biedl syndrome 10 protein (5.35)
Fibromodulin (2.58)	Extracellular matrix protein 1 (4.84)	50S ribosomal protein L29 (5.26)
Hyaluronan and proteoglycan link protein 1 (1.79)	Alpha-1-antitrypsin (2.89)	Max-like protein X (5.08)
Cartilage intermediate layer protein 1 (1.32)	Keratin (2.82)	Trypsin (4.82)
Metalloproteinase inhibitor 1 (1.22)	Decorin (2.59)	Triosephosphate isomerase (4.74)
Thrombospondin 1(1.13)	Myotubularin-related protein 1 (2.54)	Alpha-2-HS-glycoprotein (4.62)

The percentage of rules each protein was used in is shown in brackets.

Table 4 The ten mammalian proteins found most frequently in rules for each of the three classes, not including the default control class, from the emPAI dataset

IL-1β	Carprofen	IL-1β and Carprofen
Triosephosphate isomerase (73.4)	Alpha-2-HS-glycoprotein (99.3)	interleukin 8 (81.4)
Albumin (40.4)	Tumor necrosis factor receptor superfamily member 11B (15.6)	Lumican (14.9)
Serum amyloid A protein (23)	Pseudouridylate synthase 7 homolog (15.1)	Matrix metalloproteinase 3 (3.1)
Matrix metalloproteinase 3 (21.3)	Lysozyme C (2)	Desmin (2.6)
Vimentin-1 (9.8)	Apolipoprotein E (1.8)	Clusterin (2)
Vimentin (6)	Matrix gla protein (1.8)	Syndecan-4 (1.4)
Enolase B (5.9)	Fibromodulin (1.4)	Ribonuclease 4 (1)
Thrombospondin 1(5.3)	Clusterin (1.3)	Thrombospondin-3 (0.8)
Enolase A (3.4)	Metalloproteinase inhibitor 1 (0.6)	Enolase (0.6)
Keratin (3.8)	Retinoblastoma-like protein 2 (0.4)	Cartilage intermediate layer protein 1 (0.4)

The percentage of rules each protein was used in is shown in brackets.

of these techniques aimed at identifying proteins involved in cartilage degradation. There were other proteins, such as APOE and HPLN1 that were found frequently in the rules. However, the supplementary tables show they are not present in all the samples of any group.

The proteins identified by this protocol were compared to those found using the same proteomics protocol, but without the bioinformatics analysis, using equine explant tissue [33]. There were proteins highlighted in the equine study that were not in this canine study, including COMP, fibronectin and chondroadherin, because, whilst they were abundant in the samples, they were not significantly different across the different treatment groups. Therefore, the bioinformatics methods used provide a way to focus on the most relevant proteins.

The data indicate that in the absence of IL-1β carprofen had little effect on the cartilage explant secretome. Therefore, proteins that aided in the classification may have been included in the classification model, but are not necessarily intrinsically involved in the processes being investigated. This resulted in some non-mammalian proteins identified as top ranking proteins. It is possible traces of contaminating proteins entered the analysis and the proteins have been correctly identified. Alternatively, proteins were incorrectly identified by Mascot; because the selected proteins were not in the database used, in which case the highest-ranking closest protein was used.

The major challenge faced by many proteomic studies is the under representation of the lower abundance proteins that are potentially of interest. This under representation is due to the massive range of protein abundance in complex biological samples such as serum,

Table 5 The ten mammalian proteins found most frequently in rules for each of the three classes, not including the default control class, from the ProteinProphet and emPAI combined dataset

IL-1β		Carprofen		IL-1β and Carprofen	
Triosephosphate isomerase (19.8)	PP	Apolipoprotein E (13.7)	PP	interleukin 8 (33.8)	emPAI
Albumin (10)	emPAI	interleukin 8 (10.8)	emPAI	interleukin 8 (31)	PP
Triosephosphate isomerase (9.3)	emPAI	interleukin 8 (9.9)	PP	Clusterin (5.8)	emPAI
Matrix metalloproteinase 3 (7.6)	PP	Matrix gla protein (9.3)	PP	Matrix metalloproteinase 3 (5.7)	PP
Thrombospondin 1 (6.2)	PP	Hyaluronan and proteoglycan link protein 1 (8.4)	PP	Thrombospondin-3 (2.4)	emPAI
Serum amyloid A protein (3.5)	emPAI	Fibromodulin (7.4)	emPAI	Matrix metalloproteinase 3 (2.2)	emPAI
Matrix metalloproteinase 3 (3.4)	emPAI	Transmembrane protein PVRIG (4.6)	PP	Ribonuclease 4 (1.9)	PP
Enolase B (3)	emPAI	Matrix gla protein (3.2)	emPAI	Ribonuclease 4 (1.7)	emPAI
Keratin (2.5)	emPAI	Hyaluronan and proteoglycan link protein 1 (2.3)	emPAI	Cartilage intermediate layer protein 1 (1)	emPAI
Enolase A (2.4)	emPAI	Clusterin (2)	emPAI	Lumican (1)	emPAI

The percentage of rules each protein was used in is shown in brackets. Alongside the proteins the table reports whether the rules were using the protein with its ProteinProphet probability (PP) or emPAI score.
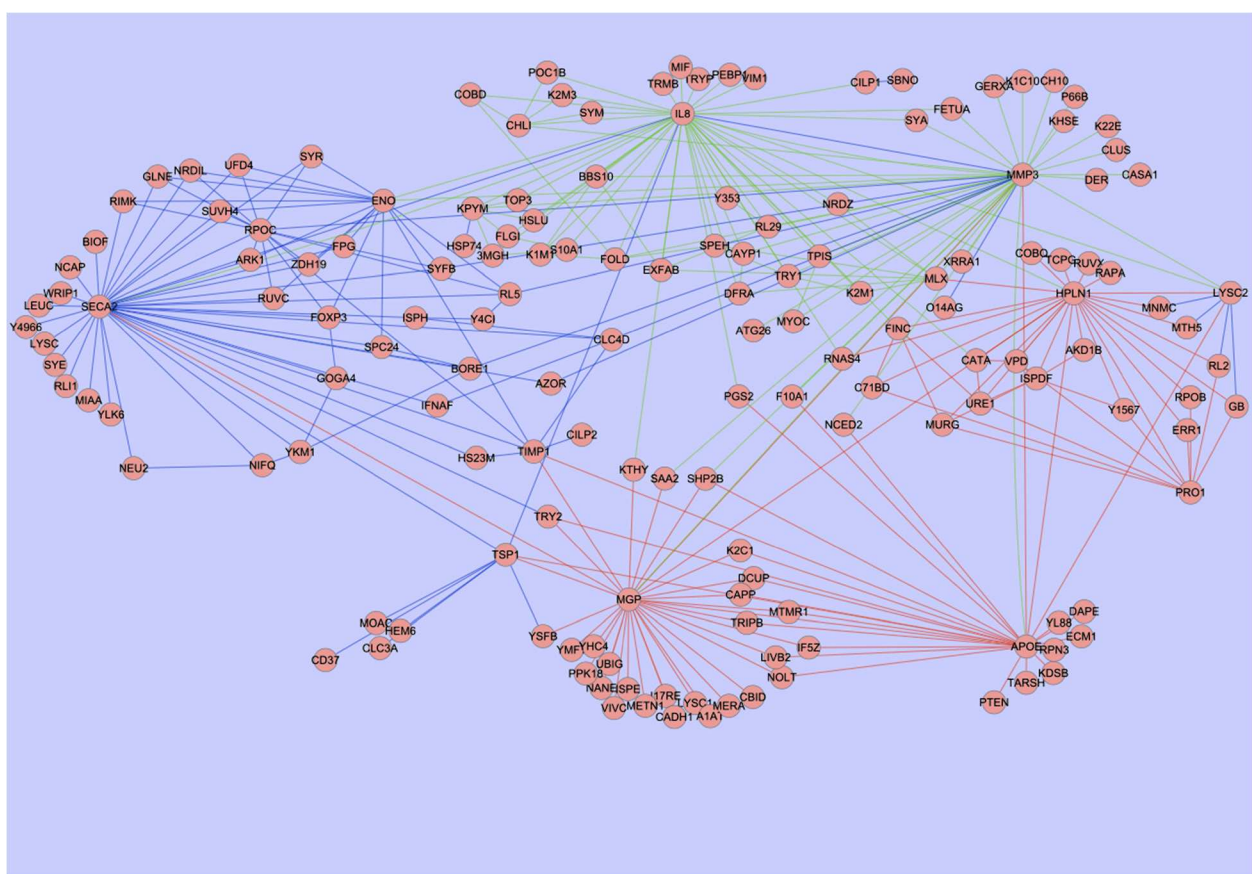
Figure 3 Protein interaction network generated from the top 100 BioHEL protein pairs for the ProteinProphet canine articular cartilage dataset. The most frequently used protein pairs for IL-1β in blue, carprofen in red and carprofen + IL-1β in green.

Table 6 Most connected proteins identified from the ProteinProphet protein pairs network

Most connected proteins	Description
interleukin 8 (IL-8)	A chemotactic factor known to attract neutrophils, basophils, and T-cells and is involved in neutrophil activation. IL-8 is released from a number of cell types in response to an inflammatory stimulus [46].
Matrix metalloproteinase-3 (MMP-3)	MMP-3 can degrade fibronectin, laminin, some gelatins, various collagens and cartilage proteoglycans [47,48].
Apolipoprotein E (APOE)	APOE mediates the binding, internalization, and catabolism of lipoprotein particles [49,50].
Matrix gla protein (MGP)	MGP interacts with the matrix of bone and cartilage and is thought to act as an inhibitor of bone formation [51,52].
Hyaluronan and proteoglycan link protein 1 (HPLN1)	HPLN1 stabilizes the aggregates of proteoglycan monomers with hyaluronic acid in the extracellular cartilage matrix [53].

cerebrospinal fluid and urine or marginally less complex samples like the secretome [60] with high abundant proteins saturating the MS/MS with higher signal levels. Proteins, such as COMP, are highly abundant in the cartilage and hinder identification of less abundant proteins relevant to biological processes. Sample preparation techniques such as proteome fractionation and deglycosylation should enable the identification of less abundant proteins and therefore more information could be uncovered using these techniques.

As described, additional analyses were performed on a number of top ranking proteins identified by these methods. However, further analysis is required to investigate the individual proteins highlighted and other proteins in the networks. This includes both laboratory-based experiments to confirm the presence of individual proteins and their levels within different sample types, and further literature and pathway analyses to mine relevant previously identified information.

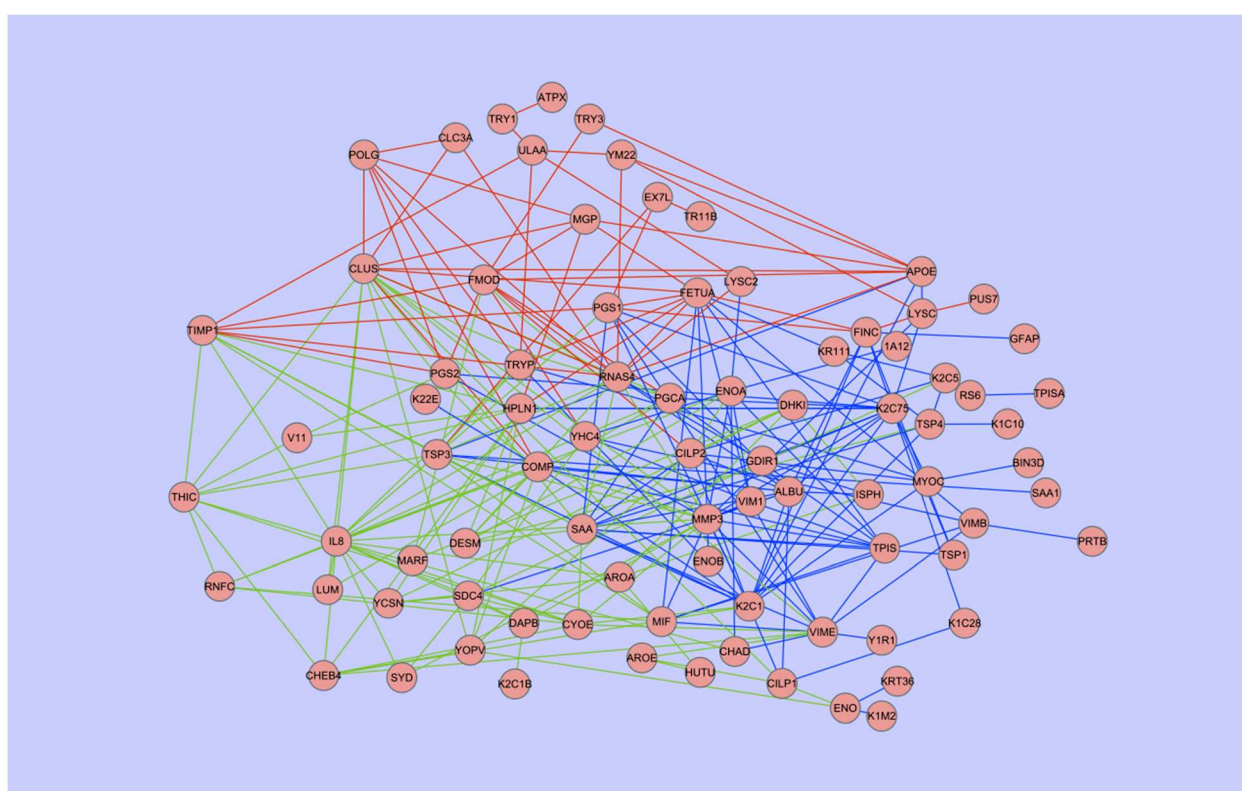Due to the nature of the machine learning methods used, it would be more suitable to analyze larger datasets

Figure 4 Protein interaction network generated from the top 100 BioHEL protein pairs for the emPAI canine articular cartilage dataset.
The most frequently used protein pairs for IL-1β in blue, carprofen in red and carprofen + IL-1β in green.

and therefore future work should include a similar study on a larger scale, with more replicate samples and a larger number of animals.

Conclusions

This study involved bioinformatic analysis of high throughput proteomic data generated using an explant model of cytokine-induced articular cartilage degradation. The approach described in this paper may be used in future studies for identification of early structural changes in cartilage and for drug testing, and screening of novel anti-inflammatory compounds from natural products. Extending our previous work with explant models of articular cartilage, bioinformatics techniques were applied to high throughput proteomics data to identify proteins suitable for use as exploratory biomarkers. This combination of laboratory-based and computational methods has provided results, which experimental techniques alone could not have provided. This proteomic and bioinformatics study has detected a number of established ECM proteins, including MMP-3, IL-8 and MGP, and therefore has shown the application of these bioinformatics tools is suitable for this purpose

and could be applied to proteomics data from other areas.

Endnotes

[a]http://www.legislation.gov.uk/ukpga/1986/14/contents
[b]http://www.legislation.gov.uk/ukpga/1966/36

Additional files

Additional file 1: Table S1. Proteins identified by Mascot in the control (untreated) samples with corresponding Mascot scores. The Mascot score is a probability based score, used to determine the significance of a protein match. The higher the score the less likely it is that the protein match occurred by random.

Additional file 2: Table S2. Proteins identified by Mascot in the IL-1β treated samples with their corresponding Mascot scores.

Additional file 3: Table S3. Proteins identified by Mascot in the carprofen treated samples with their corresponding Mascot scores.

Additional file 4: Table S4. Proteins identified by Mascot in the samples treated with a combination of carprofen and IL-1β and their corresponding Mascot scores.

Additional file 5: Figure S1. SDS-PAGE protein profile of secretome from dog one. a) control (1,2,3,4), IL-1β (5,6,7,8) b) control (1,2,3,4), carprofen (5,6,7,8) c) control (1,2,3,4), IL-1β + carprofen (5,6,7,8). Molecular weight markers (M) (in kDa) were Bio-Rad Precision Plus unstained standards.

Additional file 6: Figure S2. SDS-PAGE protein profile of secretome from dog two. a) control (1,2,3,4), IL-1β (5,6,7,8) b) control (1,2,3,4), carprofen (5,6,7,8) c) control (1,2,3,4), IL-1β + carprofen (5,6,7,8). Molecular weight markers (M) (in kDa) were Bio-Rad Precision Plus unstained standards. Lanes 1 – 8 each contain 14-μg protein. Lane 9 contains blank loading buffer control. Arrows indicate differences in protein bands between sample sets. Gels were silver stained.

Abbreviations
APOE: Apolipoprotein E; COMP: Cartilage oligomeric matrix protein; DMEM: Dulbecco's modified eagle medium; DTT: Dithiothreitol; ECM: Extracellular matrix; emPAI: Exponentially modified protein abundance index; HPLC: High-performance liquid chromatography; HPLN1: Hyaluronan and proteoglycan link protein 1; IL-1β: Interleukin 1 beta; IL-8: Interleukin-8; MGP: Matrix gla protein; MMP: Matrix metalloproteinase; MMP-3: Matrix metalloproteinase-3 (stromelysin-1); MS: Mass spectrometry; OA: Osteoarthritis.

Competing interests

Authors' contributions
Read, edited and approved the final manuscript: ALS, KLH, JRS, DA, SL, JB, AM. Conceived and designed the experiments: ALS, JB, SL, DA, AM. AM and JB contributed equally to the study design and conception. Performed the experiments: KLH, ALS, JRS. Analyzed the data: ALS, JB, AM. Contributed reagents/materials/analysis tools: ALS, JB. Wrote the paper: ALS, JB, SL, AM. All authors read and approved the final manuscript.

Authors' information
Jaume Bacardit and Ali Mobasheri
The D-BOARD European Consortium for Biomarker Discovery, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom; http://cordis.europa.eu/projects/rcn/105314_en.html
http://ec.europa.eu/research/health/medical-research/severe-chronic-diseases/projects/d-board_en.html

Acknowledgments

Disclosures

Author details
[1]School of Biosciences, Faculty of Science, University of Nottingham, Sutton Bonington Campus, LE12 5RD Leicestershire, UK. [2]Musculoskeletal Research Group, School of Veterinary Medicine and Science, Faculty of Medicine and Health Science, University of Nottingham, Sutton Bonington Campus, LE12 5RD Leicestershire, UK. [3]Bruker UK Limited, CV4 9GH Coventry, UK. [4]WALTHAM® Centre for Pet Nutrition, Waltham-on-the-Wolds, Melton Mowbray, LE14 4RT Leicestershire, UK. [5]Proteomics Laboratory, School of Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD Leicestershire, UK. [6]School of Computer Science, University of Nottingham, Jubilee Campus, NG8 1BB Nottingham, UK. [7]The D-BOARD European Consortium for Biomarker Discovery, University of Nottingham, University Park, NG7 2RD Nottingham, UK. [8]School of Computing Science, Newcastle University, Claremont Tower, Newcastle-upon-Tyne, NE1 7RU UK. [9]Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis, Nottingham University Hospitals, NG7 2UH Nottingham, UK. [10]Arthritis Research UK Pain Centre, The University of Nottingham, Queen's Medical Centre, NG7 2UH Nottingham, UK. [11]Medical Research Council and Arthritis Research UK Centre for Musculoskeletal Ageing Research, The University of Nottingham, Queen's Medical Centre, NG7 2UH Nottingham, UK. [12]Center of Excellence in Genomic Medicine Research (CEGMR), King Fahad Medical Research Center (KFMRC), King AbdulAziz University, 21589 Jeddah, Kingdom of Saudi Arabia. [13]Schools of Pharmacy and Life Sciences, University of Bradford, Richmond Road, BD7 1DP Bradford, UK. [14]Comparative Physiology, Medical Research Council-Arthritis Research UK Centre for Musculoskeletal Ageing Research, Arthritis Research UK Pain Centre, Arthritis Research UK Centre for Sport, Exercise, and Osteoarthritis, Faculty of Medicine and Health Sciences, The University of Nottingham, Sutton Bonington Campus, LE12 5RD Leicestershire, UK. [15]Faculty of Medicine and Health Sciences, The University of Nottingham, Sutton Bonington Campus, LE12 5RD Leicestershire, UK.

References
1. Lammi MJ, Häyrinen J, Mahonen A: **Proteomic analysis of cartilage- and bone-associated samples.** *ELECTROPHORESIS* 2006, **27**(13):2687–2701.
2. Eyre D: **Articular cartilage and changes in arthritis: collagen of articular cartilage.** *Arthritis Res* 2002, **4**(1):30–35.
3. Bian L, Angione SL, Ng KW, Lima EG, Williams DY, Mao DQ, Ateshian GA, Hung CT: **Influence of decreasing nutrient path length on the development of engineered cartilage.** *Osteoarthr Cartilage* 2009, **17**(5):677–685.
4. Newman AP: **Articular cartilage repair.** *Am J Sports Med* 1998, **26**(2):309–324.
5. Vaughan-Scott T, Taylor JH: **The pathophysiology and medical management of canine osteoarthritis.** *J S Afr Vet Assoc* 1997, **68**(1):21–25.
6. Buckwalter JA, Saltzman C, Brown T: **The impact of osteoarthritis: implications for research.** *Clin Orthop Relat R* 2004, **427**:S6–S15. 10.1097/01.blo.0000143938.0000130681.0000143939d.
7. Goldring MB: **Update on the biology of the chondrocyte and new approaches to treating cartilage diseases.** *Best Pract Res Clin Rheumatol* 2006, **20**(5):1003–1025.
8. Macrory L, Vaughan-Thomas A, Clegg P, Innes J: **An exploration of the ability of tepoxalin to ameliorate the degradation of articular cartilage in a canine in vitro model.** *BMC Vet Res* 2009, **5**(1):25.
9. Radin EL, Burr DB, Caterson B, Fyhrie D, Brown TD, Boyd RD: **Mechanical determinants of osteoarthrosis.** *Semin arthritis rheu* 1991, **21**(3, Supplement 2):12–21.
10. Sack KE: **Osteoarthritis. A continuing challenge.** *Western J Med* 1995, **163**(6):579–586.
11. Mobasheri A: **Osteoarthritis year 2012 in review: biomarkers.** *Osteoarthr Cartilage* 2012, **20**(12):1451–1464.
12. Bai JP, Bell R, Buckman S, Burckart GJ, Eichler HG, Fang KC, Goodsaid FM, Jusko WJ, Lesko LL, Meibohm B, et al: **Translational biomarkers: from preclinical to clinical a report of 2009 AAPS/ACCP biomarker workshop.** *AAPS J* 2011, **13**(2):274–283.
13. Kraus VB, Burnett B, Coindreau J, Cottrell S, Eyre D, Gendreau M, Gardiner J, Garnero P, Hardin J, Henrotin Y, et al: **Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis.** *Osteoarthr Cartilage* 2011, **19**(5):515–542.
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):1–19.
15. Bay-Jensen AC, Sondergaard BC, Christiansen C, Karsdal MA, Madsen SH, Qvist P: **Biochemical markers of joint tissue turnover.** *Assay Drug Dev Technol* 2010, **8**(1):118–124.
16. Bauer DC, Hunter DJ, Abramson SB, Attur M, Corr M, Felson D, Heinegård D, Jordan JM, Kepler TB, Lane NE, et al: **Classification of osteoarthritis biomarkers: a proposed approach.** *Osteoarthr and Cartilage* 2006, **14**(8):723–727.
17. Felson DT, Lohmander LS: **Whither osteoarthritis biomarkers?** *Osteoarthr cartilage/OARS, Osteoarthr Res Soc* 2009, **17**(4):419–422.
18. Moore RE, Kirwan J, Doherty MK, Whitfield PD: **Biomarker discovery in animal health and disease: the application of post-genomic technologies.** *Biomarker insights* 2007, **2**:185–196.

19. Ruiz-Romero C, Blanco FJ: **Proteomics role in the search for improved diagnosis, prognosis and treatment of osteoarthritis.** *Osteoarthr Cartilage* 2010, **18**(4):500–509.

20. Gharbi M, Deberg M, Henrotin Y: **Application for proteomic techniques in studying osteoarthritis: a review.** *Front Physiol* 2011, **2**:90.

21. Royce PM, Steinmann B: *Connective tissue and its heritable disorders: molecular, genetic, and medical aspects.* New York: John Wiley & Sons; 2002.

22. Buckwalter JA, Mankin HJ: **Articular cartilage: tissue design and chondrocyte-matrix interactions.** *Instructional course lectures* 1998, **47**:477–486.

23. Kiviranta I, Jurvelin J, Tammi M, SääMäunen A-M, Helminen HJ: **Weight bearing controls glycosaminoglycan concentration and articualr cartilage thickness in the knee joints of young beagle dogs.** *Arthritis & Rheumatism* 1987, **30**(7):801–809.

24. Dickinson SC, Vankemmelbebe MN, Buttle DJ, Rosenberg K, Heinegård D, Hollander AP: **Cleavage of cartilage oligomeric matrix protein (thrombospondin-5) by matrix metalloproteinases and a disintegrin and metalloproteinase with thrombospondin motifs.** *Matrix Biol* 2003, **22**(3):267–278.

25. Aigner T, Zien A, Hanisch D, Zimmer R: **Gene expression in chondrocytes assessed with use of microarrays.** *J Bone Joint Surg Am* 2003, **85-A**(Suppl 2):117–123.

26. Deutsch EW, Lam H, Aebersold R: **Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics.** *Physiol Genomics* 2008, **33**(1):18–25.

27. Kanehisa M, Bork P: **Bioinformatics in the post-sequence era.** *Nat Genet* 2003, **33**:305–310.

28. Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armañanzas R, Santafé G, Pérez A, *et al*: **Machine learning in bioinformatics.** *Brief Bioinform* 2006, **7**(1):86–112.

29. Kotsiantis SB: *Supervised machine learning: a review of classification techniques*, Proceedings of the 2007 conference on emerging artificial intelligence applications in computer engineering: real word AI systems with applications in eHealth, HCI, information retrieval and pervasive technologies. IOS Press: Emerging Artificial Intelligence Applications in Computer Engineering; 2007:3–24.

30. Fürnkranz J: **Separate-and-conquer rule learning.** *Artif Intell Rev* 1999, **13**(1):3–54.

31. Bacardit J, Burke E, Krasnogor N: **Improving the scalability of rule-based evolutionary learning.** *Memetic Computing* 2009, **1**:55–67.

32. Witten I, Frank E, Hall M: *Data mining: practical machine learning tools and techniques.* 3rd edition. San Francisco, CA: Morgan Kaufmann; 2011.

33. Clutterbuck AL, Smith JR, Allaway D, Harris P, Liddell S, Mobasheri A: **High throughput proteomic analysis of the secretome in an explant model of articular cartilage inflammation.** *J Proteomics* 2011, **74**(5):704–715.

34. Zwickl H, Traxler E, Staettner S, Parzefall W, Grasl-Kraupp B, Karner J, Schulte-Hermann R, Gerner C: **A novel technique to specifically analyze the secretome of cells and tissues.** *ELECTROPHORESIS* 2005, **26**(14):2779–2785.

35. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, **4**(9):1265–1272.

36. Rappsilber J, Ryder U, Lamond A, Mann M: **Large-scale proteomic analysis of the human spliceosome.** *Genome Res* 2002, **12**:1231–1245.

37. Keller A, Nesvizhskii AI, Kolker E, Aebersold R: **Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.** *Anal Chem* 2002, **74**(20):5383–5392.

38. Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Anal Chem* 2003, **75**(17):4646–4658.

39. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, Sun Z, Nilsson E, Pratt B, Prazen B, *et al*: **A guided tour of the trans-proteomic pipeline.** *Proteomics* 2010, **10**(6):1150–1159.

40. Kohavi R: *A study of cross-validation and bootstrap for accuracy estimation and model selection*, 14th International joint conference on artificial intelligence: 1995; 1995:1137–1145.

41. Urbanowicz RJ, Granizo-Mackenzie A, Moore JH: **An analysis pipeline with statistical and visualization-guided knowledge discovery for Michigan-style learning classifier systems.** *Computational Intelligence Magazine, IEEE* 2012, **7**(4):35–45.

42. Bassel GW, Glaab E, Marquez J, Holdsworth MJ, Bacardit J: **Functional network construction in arabidopsis using rule-based machine learning on large-scale data sets.** *The Plant Cell Online* 2011, **23**(9):3101–3116.

43. Glaab E, Bacardit J, Garibaldi JM, Krasnogor N: **Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data.** *PLoS ONE* 2012, **7**(7):e39932.

44. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498–2504.

45. Patra D, Sandell LJ: **Recent advances in biomarkers in osteoarthritis.** *Curr Opin Rheumatol* 2011, **23**(5):465–470.

46. Straubinger RK, Straubinger AF, Härter L, Jacobson RH, Chang YF, Summers BA, Erb HN, Appel MJ: **Borrelia burgdorferi migrates into joint capsules and causes an up-regulation of interleukin-8 in synovial membranes of dogs experimentally infected with ticks.** *Infect Immun* 1997, **65**(4):1273–1285.

47. Parks WC, Wilson CL, Lopez-Boado YS: **Matrix metalloproteinases as modulators of inflammation and innate immunity.** *Nat Rev Immunol* 2004, **4**(8):617–629.

48. Okada Y, Konomi H, Yada T, Kimata K, Nagase H: **Degradation of type IX collagen by matrix metalloproteinase 3 (stromelysin) from human rheumatoid synovial cells.** *FEBS Lett* 1989, **244**(2):473–476.

49. Luo CC, Li WH, Chan L: **Structure and expression of dog apolipoprotein A-I, E, and C-I mRNAs: implications for the evolution and functional constraints of apolipoprotein structure.** *J Lipid Res* 1989, **30**(11):1735–1746.

50. Mahley RW, Ji Z-S: **Remnant lipoprotein metabolism: key pathways involving cell-surface heparan sulfate proteoglycans and apolipoprotein E.** *J Lipid Res* 1999, **40**(1):1–16.

51. Zebboudj AF, Imura M, Boström K: **Matrix GLA protein, a regulatory protein for bone morphogenetic protein-2.** *J Biol Chem* 2002, **277**(6):4388–4394.

52. Dhore CR, Cleutjens JPM, Lutgens E, Cleutjens KBJM, Geusens PPM, Kitslaar PJEHM, Tordoir JHM, Spronk HMH, Vermeer C, Daemen MJAP: **Differential expression of bone matrix regulatory proteins in human atherosclerotic plaques.** *Arterioscler Thromb Vasc Biol* 2001, **21**(12):1998–2003.

53. Hardingham TE, Fosang AJ: **Proteoglycans: many forms and many functions.** *FASEB J* 1992, **6**(3):861–870.

54. Clutterbuck AL, Asplin KE, Harris P, Allaway D, Mobasheri A: **Targeting matrix metalloproteinases in inflammatory conditions.** *Curr Drug Targets* 2009, **10**(12):1245–1254.

55. Chandran V, Gladman DD: **Update on biomarkers in psoriatic arthritis.** *Curr Rheumatol Rep* 2010, **12**(4):288–294.

56. Keyszer G, Lambiri I, Nagel R, Keysser C, Keysser M, Gromnica-Ihle E, Franz J, Burmester GR, Jung K: **Circulating levels of matrix metalloproteinases MMP-3 and MMP-1, tissue inhibitor of metalloproteinases 1 (TIMP-1), and MMP-1/TIMP-1 complex in rheumatic disease. Correlation with clinical activity of rheumatoid arthritis versus other surrogate markers.** *J Rheumatol* 1999, **26**(2):251–258.

57. Rai MF, Sandell LJ: **Inflammatory mediators: tracing links between obesity and osteoarthritis.** *Crit Rev Eukaryot Gene Expr* 2011, **21**(2):131–142.

58. Nishiura H, Tanaka J, Takeya M, Tsukano M, Kambara T, Imamura T: **IL-8/NAP-1 is the major T-cell chemoattractant in synovial tissues of rheumatoid arthritis.** *Clin Immunol Immunopathol* 1996, **80**(2):179–184.

59. Goldring MB: **The role of cytokines as inflammatory mediators in osteoarthritis: lessons from animal models.** *Connect Tissue Res* 1999, **40**(1):1–11.

60. Righetti PG, Castagna A, Antonioli P, Boschetti E: **Prefractionation techniques in proteome analysis: the mining tools of the third millennium.** *Electrophoresis* 2005, **26**(2):297–319.