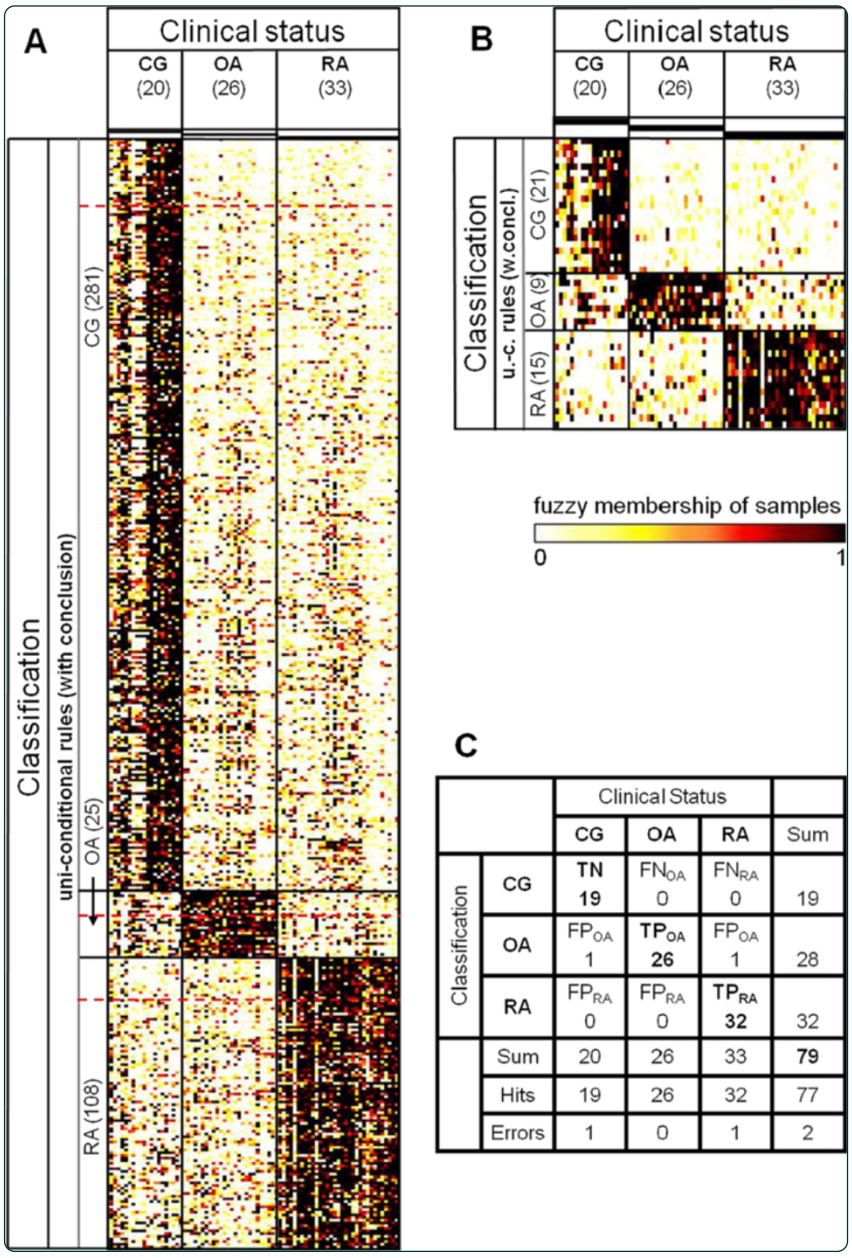
arthritis research&therapy



Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation

Woetzel *et al.*

BioMed Central

Woetzel *et al. Arthritis Research & Therapy* 2014, **16**:R84
http://arthritis-research.com/content/16/2/R84

arthritis research&therapy

RESEARCH ARTICLE Open Access

Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation

Dirk Woetzel[1], Rene Huber[2,3], Peter Kupfer[4], Dirk Pohlers[2,5], Michael Pfaff[1,6], Dominik Driesch[1], Thomas Häupl[7], Dirk Koczan[8], Peter Stiehl[9], Reinhard Guthke[4] and Raimund W Kinne[2*]

Abstract

Introduction: Discrimination of rheumatoid arthritis (RA) patients from patients with other inflammatory or degenerative joint diseases or healthy individuals purely on the basis of genes differentially expressed in high-throughput data has proven very difficult. Thus, the present study sought to achieve such discrimination by employing a novel unbiased approach using rule-based classifiers.

Methods: Three multi-center genome-wide transcriptomic data sets (Affymetrix HG-U133 A/B) from a total of 79 individuals, including 20 healthy controls (control group - CG), as well as 26 osteoarthritis (OA) and 33 RA patients, were used to infer rule-based classifiers to discriminate the disease groups. The rules were ranked with respect to Kiendl's statistical relevance index, and the resulting rule set was optimized by pruning. The rule sets were inferred separately from data of one of three centers and applied to the two remaining centers for validation. All rules from the optimized rule sets of all centers were used to analyze their biological relevance applying the software Pathway Studio.

Results: The optimized rule sets for the three centers contained a total of 29, 20, and 8 rules (including 10, 8, and 4 rules for 'RA'), respectively. The mean sensitivity for the prediction of RA based on six center-to-center tests was 96% (range 90% to 100%), that for OA 86% (range 40% to 100%). The mean specificity for RA prediction was 94% (range 80% to 100%), that for OA 96% (range 83.3% to 100%). The average overall accuracy of the three different rule-based classifiers was 91% (range 80% to 100%). Unbiased analyses by Pathway Studio of the gene sets obtained by discrimination of RA from OA and CG with rule-based classifiers resulted in the identification of the pathogenetically and/or therapeutically relevant interferon-gamma and GM-CSF pathways.

Conclusion: First-time application of rule-based classifiers for the discrimination of RA resulted in high performance, with means for all assessment parameters close to or higher than 90%. In addition, this unbiased, new approach resulted in the identification not only of pathways known to be critical to RA, but also of novel molecules such as serine/threonine kinase 10.

Introduction

Rheumatoid arthritis (RA) and osteoarthritis (OA) are the most common forms of arthritis [1]. In spite of different pathogeneses, these arthritides exhibit phenotypic similarities and overlapping cellular and molecular characteristics [1,2]. RA is a progressive, chronically inflammatory, destructive joint disease of still unknown etiology, perpetuated by an invasive synovial membrane (also known as pannus tissue) [3]. Various activated or semi-transformed cell types in the synovial membrane (monocytes/macrophages, osteoclasts, T cells and B cells, dendritic cells and endothelial cells, synovial fibroblasts) contribute to the development and progression of RA by secretion of proinflammatory cytokines and tissue-degrading proteases [4,5]. Similarly, OA is characterized by progressive destruction of cartilage and bone and dysregulation of synovial function [6]. OA arises from the damage of articular cartilage

* Correspondence: raimund.w.kinne@med.uni-jena.de
[2]Experimental Rheumatology Unit, Department of Orthopedics, Jena University Hospital, Waldkrankenhaus Rudolf Elle, Klosterlausnitzer Straße 81, 07607 Eisenberg, Germany
Full list of author information is available at the end of the article

induced by physical injury and is subsequently influenced by a variety of intrinsic (for example, genetic, cellular, or immunologic) factors [7]. The OA synovial membrane also shows an inflammatory component, although clearly less pronounced than in RA [2,7].

Compatible with these similarities, the synovial tissue of OA and RA patients contains mesenchymal precursor cells and attempts to regenerate damaged cartilage and subchondral bone in the adult organism. In contrast to fetal healing, however, the synovial tissue may require inflammation to sustain and control the fibroproliferation [8].

Although these overlapping features have led to the development of pharmacological or surgical therapies effective in both diseases [9-12], the similarities at the same time impede a reliable discrimination of the two arthritides. Diagnostic methods classically include radiography [13], histopathological assessment of synovitis [14], detection of rheumatic nodules, selected laboratory values such as rheumatoid factor and citrullinated peptides [15,16], and evaluation of the patients' individual and family history [17]. Recently, an improved ultrasound-based scoring system has also been proposed [18]. In general, American College of Rheumatology criteria for RA [15,19] or for OA [16] are often used for diagnostic purposes, although they were originally intended as classification criteria, for example, for the comparison of cohorts in different clinical studies [20]. However, an appropriate discrimination of RA and OA is particularly difficult at later stages of the diseases, and the recent revision of the respective criteria has not significantly improved their diagnostic capability [20]. For instance, the presence of rheumatoid factor as a marker for RA has been questioned due to its high variability and should be replaced by the level of anti-citrullinated protein antibodies [21].

An easier discrimination of different forms of arthritis has been attempted by molecular approaches, in particular, disease-specific gene expression profiling. These attempts have partially focused on the expression of selected candidate molecules with a known influence on the respective diseases; for example, type I interferon family members [22,23], tumor necrosis factor superfamily and bone morphogenetic protein family members [24], citrullinated synovial proteins [25], and proteases such as metalloproteinases or cathepsins [26]. Although these studies have indicated the existence of individual or combined biomarkers for RA, the validity of this approach has not been universal. Some of the studies have succeeded in discriminating RA from normal controls, but not from other arthritides, while other studies have successfully discriminated RA from other forms of arthritis (such as spondylopathy or psioriatic arthritis), but not from OA [24].

In parallel to candidate gene analyses, broader, unbiased genome-wide gene expression profiles [27] have been used to identify disease-specific signatures and hidden biomarkers in rheumatology with microarray-based methods [28]. This has been applied to discriminate early versus late RA [29] and to discriminate RA versus OA [30,31]. In addition, differentially expressed genes have been successfully used to predict the response of RA patients to therapeutic approaches, for example, the capability of certain (type I interferon-responsive) genes to predict rituximab nonresponders [32] and anti-tumor necrosis factor nonresponders [33] or to define homogeneous subgroups within a heterogeneous disease such as RA [22]. However, most studies were not designed to identify gene expression patterns as a potential diagnostic tool, but rather to elucidate the underlying transcriptional networks [34]. The validity of the identified genes as markers for RA or OA was generally also not validated in replication cohorts. Finally, differentially expressed respectively regulated genes or pathways common to RA and OA remain a major challenge [30].

These obstacles may be overcome using microarray data from several analytic centers to identify sets of differentially expressed genes for the reliable diagnosis of different arthritides. For this purpose, bioinformatic methods suitable to process and interpret the large amounts of high-dimensional data, and also algorithms for the identification of rules concerning the expression of disease-specific genes, are of utmost importance [35].

In personalized medicine and theranostics, the generation of decision rules is a well-established method for the design of clinical decision support systems and/or for the discovery of relevant relationships among pathogenetically relevant genes in large databases [36,37]. This approach is intended to identify strong rules using different measures of so-called interestingness, for example, specificity for a certain disease entity. To select interesting rules from the set of all possible rules, constraints on various measures of significance can be used, such as thresholds on support and confidence. In our hands [38], the relevance index introduced by Kiendl and coworkers [39-43] is able to generate robust rule sets with high predictive strength from data of high dimension (for example, number of genes) but of low sample number. A deterministic decision rule R_r is defined by 'IF P_r (y) THEN C_r', where P_r describes a premise evaluating the observations y (that is, the enhanced expression of a given gene) and C_r is the set of possible conclusions (for example, the prediction of a disease status of a given individual). In the present work, C_r is a categorical variable defined by the set of three clinical states {'CG' – control group, 'RA' – rheumatoid arthritis, 'OA' – osteoarthritis} and each premise P_r is defined by the expression of only one gene (uniconditional rules).

This rule-oriented approach may represent a more suitable alternative to the widely used identification of

differentially expressed genes to generate a sorted list of candidate genes of interest. The approach thus combines three major advantages: i) by avoiding the application of differentially expressed genes, it is more robust in its discriminative capacity to data heterogeneity among different donors or patients; ii) due to separate normalization and independent rule set generation, it is capable of eliminating center-specific effects, thus yielding higher sample sizes in study cohorts; and iii) cross-validation among different clinical centers is possible, independently of individual differentially expressed genes.

In this study, three multicenter genome-wide transcriptomic datasets from 79 individuals were used to infer rule-based classifiers to discriminate RA, OA, and healthy controls. The rule sets were inferred separately from one center and were applied to the other centers for validation. This novel approach resulted in high performance (close to 90% for specificity, sensitivity, and accuracy) for the discrimination of RA. Unbiased analysis of the biological relevance of the underlying rules by Pathway Studio (Elsevier, Munich, Germany) and gene enrichment analysis succeeded in identifying pathways with pathogenetic or therapeutic relevance in RA.

Materials and methods

Patients

Synovial membrane samples were obtained either from postmortem joints/traumatic joint injury cases (control group (CG); $n = 15$ and $n = 5$, respectively) or from RA/OA patients (all Caucasian) upon joint replacement/synovectomy at the Jena University Hospital, Chair of Orthopedics, Waldkrankenhaus 'Rudolf Elle', Eisenberg, Germany ($n = 33$, dataset 'Jena'), at the Department of Orthopedics/Institute of Pathology/Department of Rheumatology and Clinical Immunology, Charité-Universitätsmedizin Berlin ($n = 30$, dataset 'Berlin'), and at the Department of Orthopedics/Institute of Pathology, University of Leipzig ($n = 16$, dataset 'Leipzig'). After removal, tissue samples were frozen and stored at −70°C.

The study was approved by the respective ethics committees (Jena University Hospital: Ethics Committee of the Friedrich Schiller University Jena at the Medical Faculty; Charité-Universitätsmedizin Berlin: Charité Ethics Committee; and University of Leipzig: Ethics Committee at the Medical Faculty of the University of Leipzig) and informed patient consent was obtained. RA patients were classified according to the American College of Rheumatology criteria valid in the sample assessment period [15], OA patients were classified according to the respective criteria for OA [16]. The patients/donors were assigned to one of the three terms (categorical values): 'CG', 'RA', or 'OA' (for clinical characteristics of the donors/patients, see Table 1).

Data

Data for 79 patients/donors were obtained from three clinical groups located in Jena, Berlin, and Leipzig, respectively, as presented in Table 2.

Isolation of total RNA

Tissue homogenization, total RNA isolation, and treatment with RNase-free DNase I (Qiagen, Hilden, Germany) were performed as described previously [44].

Microarray analysis

Gene expression was analyzed using HG-U133 A/B RNA microarrays (Affymetrix, Santa Clara, CA, USA) for the datasets 'Jena', 'Berlin', and 'Leipzig' – a total of 79 microarrays. Labeling of RNA probes, hybridization, and washing were carried out according to the supplier's instructions. Microarrays were analyzed by laser scanning (Gene Scanner; Hewlett-Packard, Palo Alto, CA, USA).

Pre-processing of microarray data

Gene expression data were pre-processed by MAS5.0 (Affymetrix Microarray Suite). The data are accessible through Gene Expression Omnibus series [GSE:55235] (Haeupl; 'Berlin' data), [GSE:55584] (Stiehl; 'Leipzig' data), and [GSE:55457] (Kinne; 'Jena' data).

For the study group 'Jena_all', all probe sets independent of their Affymetrix 'present call' were used for further analysis. For the study groups 'Jena', 'Berlin', 'Leipzig', and 'Total', further analyses were restricted to those genes qualified by a 'present call' in all samples of the respective study group (as calculated by MAS 5.0). The data were separately normalized for the three different study groups 'Jena', 'Berlin', and 'Leipzig' by dividing the gene expression signals for a given gene i and sample/patient j by the median over all probe sets in this sample and were subsequently logarithmized ($\log_2$), yielding the values y_{ij}. By performing completely independent normalization and rule set generation (see Rule set generation) in the three different clinical datasets, potential problems related to differences in sample preparation and wet laboratory conditions were avoided [45].

Clustering

The data were separately clustered for each probe set (gene) using a modified fuzzy C-means algorithm and two clusters. Here, the fuzzy C-means algorithm [46] was applied for the normalized and logarithmized ($\log_2$) gene expression data (y_{ij}) of a given gene for every patient belonging to the respective group (that is, 'Jena_all', 'Jena', 'Berlin', 'Leipzig', or 'Total') to estimate membership degrees (M_{ijk}) ranging from 0 to 1 for unequivocal assignment to one of the groups 'low' or 'high' gene expression. The centers (CT_{ik}; $CT_{i1} < CT_{i2}$) of the respective gene expression clusters (CL_{ik}, $k = 1$ for the cluster labeled 'low' and $k = 2$ for that labeled 'high') were also estimated.

Table 1 Clinical characteristics of the patients at the time of synovectomy/sampling

Patients (total number)	Gender (male/female)	Age (Years)	Disease duration (years)	RF (+/–)	ESR (mm/1 hour)	CRP[a] (mg/l)	Number of ARA-criteria (RA)	Concomitant medication (*n*)
Control group								
(*n* = 20)	15/5	54.7 ± 4.0	0.3 ± 0.3[b]	n.d.	n.d.	n.d.	n.a.	NSAIDs (*n* = 1)
			(n.d. = 13)					None (*n* = 7)
								(n.d. = 12)
Osteoarthritis								
(*n* = 26)	4/22	71.0 ± 1.4	7.0 ± 1.3	3/18	22.4 ± 2.7	5.3 ± 1.5	0.2 ± 0.1	NSAIDs (*n* = 16)
			(n.d. = 1)	(n.d. = 5)	(n.d. = 5)	(n.d. = 3)		None (*n* = 10)
Rheumatoid arthritis								
(*n* = 33)	8/25	57.0 ± 2.7	12.5 ± 2.0	21/7	42.7 ± 4.5	21.4 ± 4.1	5.2 ± 0.3	Prednisolone (*n* = 23)
				(n.d. = 7)	(n.d. = 10)	(n.d. = 3)		Methotrexate (*n* = 18)
								Sulfasalazine (*n* = 5)
								Chloroquine (*n* = 2)
								Leflunomide (*n* = 2)
								Cyclosporine (*n* = 1)
								Gold (*n* = 1)
								NSAIDs (*n* = 22)

For the parameters age, disease duration, ESR, CRP, and number of ARA criteria (RA), means ± standard error of the mean are given; for the remaining parameters, numbers are provided. ARA, American Rheumatism Association (now American College of Rheumatology); n.a., not applicable; n.d., not determined; NSAID, nonsteroidal anti-inflammatory drug; RA, rheumatoid arthritis; RF, rheumatoid factor; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein.
[a]Normal range, <5 mg /l.
[b]Disease duration in joint trauma patients.

Subsequently, a modified membership degree was used (M_{ijk}'; with $M_{ij1}' = 1$ and $M_{ij2}' = 0$ if $y_{ij} < CT_{i1}$; with $M_{ij1}' = 0$ and $M_{ij2}' = 1$ if $y_{ij} > CT_{i2}$; with $M_{ijk}' = M_{ijk}$ otherwise; that is, for all data in between the two centers).

Rule set generation

First, all uniconditional rules were generated independently for the three different clinical study groups 'Jena', 'Berlin', and 'Leipzig' using the formula 'IF the premise P_r is fulfilled THEN the conclusion C_r is reached'. The premise P_r is defined as follows: the expression of gene i belongs to either the cluster labeled 'low' (CL_{i1}) or the cluster labeled 'high' (CL_{i2}). The three possible conclusions (C_r; that is, in the present study the prediction of the clinical status) are 'CG' (that is, no 'RA', no 'OA'), 'RA', or 'OA'.

These rules were ranked using the relevance index RI_r introduced by Kiendl and others [39-43]. Here, a rule 'IF P_r THEN C_r' is ranked on the basis of RI_r. In this case, RI_r represents the normalized gap between the confidence interval of the conditional probability of the conclusion C_r under the premise P_r and the confidence interval of the (unconditional) probability of the conclusion C_r, as described in Additional file 1. The calculation of the confidence interval was done using a significance level $alpha_S$ with a default value 0.95, and a reduced $alpha_S$ for 'Jena', 'Berlin', and 'Leipzig' in order to generate a sufficient number (>3) of rules with $RI_r > 0$ for each conclusion ('CG', 'RA', or 'OA'). Next, it was checked and confirmed that $alpha_S > alpha_{S\text{ran-dom}}$, where at least one rule was generated for each of the three conclusions using original pre-processed gene expression values y_{ij}, and a random assignment to the individual conclusions ('CG', 'RA', and 'OA') in the training set.

Rule set pruning

As a result of the primary rule set generation, a ranked set of $r_{max}(C, S)$ rules was generated using the criterion $RI_r > 0$.

Table 2 Number of clinical samples and transcriptome datasets

Study group *S*	Control	Osteoarthritis	Rheumatoid arthritis	Total	Microarray platform[a]
'Jena'	10	10	13	33	Affymetrix HG-U133 A
'Berlin'	10	10	10	30	Affymetrix HG-U133 A
'Leipzig'	0	6	10	16	Affymetrix HG-U133 A
'Total'	20	26	33	79	

[a]From Affymetrix, Santa Clara, CA, USA.

Rule set pruning was then applied in order to minimize the numbers of both rules (r_{opt}) and 'Errors' (that is, false assignment to one of the three conclusions; for more detail see Application of the rule sets and Evaluation of a rule set). The number of rules in each rule set was optimized by greedy search with the following constraints: the numbers r_{opt}(C, S) have to be at least 4 for each conclusion and not higher than the double of the minimum number of rules in any of the respective rule sets for the three conclusions – that is, $r_{opt}(C, S) \geq 4$ and $r_{opt}(C, S) \leq 2^* \min_C(r_{max}(C, S))$.

The purpose of this step was also to generate rule sets with a balanced number of rules for the three conclusions.

Application of the rule sets

The rule sets for the different conclusions were then applied to each sample (patient) j by voting in order to achieve an individual prediction of its clinical status.

First, each rule 'IF P_r THEN C_r' with the premise P_r (P_r = 'the expression y of gene i is assigned to cluster k (i.e., "low" or "high")') was weighted by application of the aforementioned fuzzy membership degree ($W_{rj} = M_{ijk}$ ´(y_{ij})) to the sample j (see earlier Clustering). These membership weights (W_{rj}; range from 0 to 1, with 1 indicating an unequivocal prediction of the conclusion C_r) were visualized in a heat map for all samples (j) and all rules (Figures 1, 2, 3, 4, 5A,B of the respective study group).

Next, the weights W_j('CG'), W_j('OA'), and W_j('RA') for each individual sample j were calculated by summing up the respective membership weights (W_{rj}) over all rules (r) belonging to the rule set for a given conclusion.

Finally, the highest weight was used for the prediction of the clinical status of each sample (so-called 'defuzzification'):

$$C_{\text{predict}\_j} = \arg\max\left(W_j(\text{'}CG\text{'}), W_j(\text{'}OA\text{'}), W_j(\text{'RA'})\right)$$

This procedure is used for prediction of the clinical status in both the original training set (y_{ij}) from a given study group (for example, 'Jena') and all subsequently analyzed test sets from other study groups (for example, 'Berlin' and 'Leipzig').

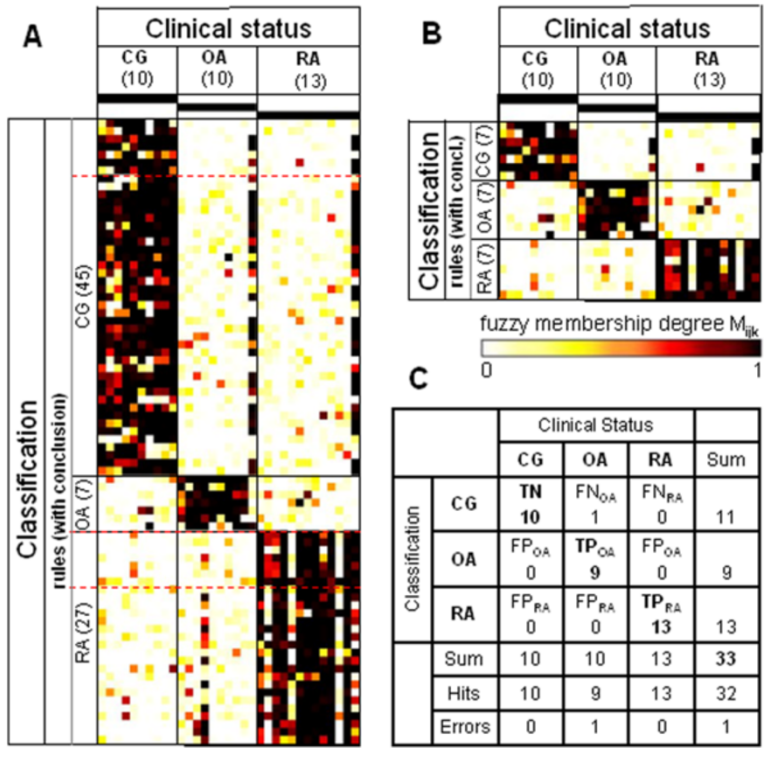


Figure 1 Heatmaps and confusion matrix for the study group 'Jena_all'. Data for the study group 'Jena_all' (that is, utilizing all probe sets) were obtained using the Jena patients (10 control group (CG), 10 osteoarthritis (OA), 13 rheumatoid arthritis (RA)) as the training set for the rule generation and re-applying the respective rules to the same dataset. **(A)** Heatmap of the membership weights applying all rules of the primary rule set (α = 0.95; 'CG', 45 rules; 'OA', seven rules; 'RA', 27 rules) for the prediction of the clinical status of the different samples; dashed red lines indicate the lower limits of the respective pruned lists of rules subsequently applied in **(B)**. **(B)** Heatmap of the membership weights applying pruned lists of rules (α = 0.95; 'CG', 'OA', and 'RA', seven top-ranked rules each) for optimized prediction of the clinical status of the different samples. **(C)** Confusion matrix for the rule set displayed in heatmap **(B)**. TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

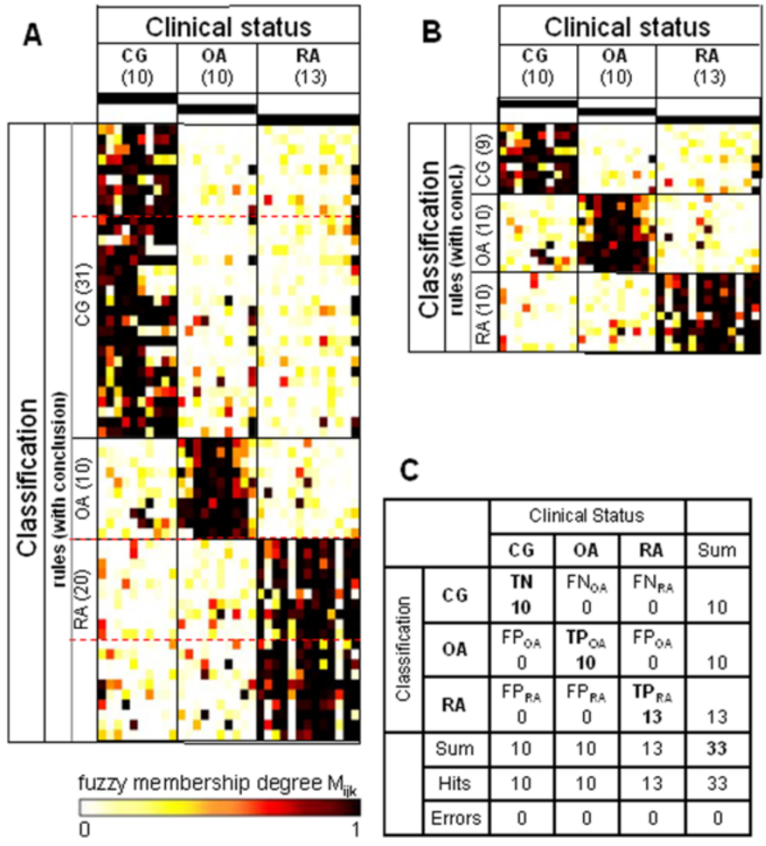


Figure 2 Heatmaps and confusion matrix for the study group 'Jena'. Data for the study group 'Jena' (that is, utilizing only the probe sets with MAS 5.0 present calls in all samples) were obtained using the Jena patients (10 control group (CG), 10 osteoarthritis (OA), 13 rheumatoid arthritis (RA)) as the training set for the rule generation and re-applying the respective rules to the same dataset. **(A)** Heatmap of the membership weights applying all rules of the primary rule set (α = 0.94; 'CG', 31 rules; 'OA', 10 rules; 'RA', 20 rules) for the prediction of the clinical status of the different samples; dashed red lines indicate the lower limits of the respective pruned lists of rules subsequently applied in **(B)**. **(B)** Heatmap of the membership weights applying pruned lists of rules (α = 0.94; 'CG', nine top-ranked rules; 'OA', 10 top-ranked rules; 'RA', 10 top-ranked rules) for optimized prediction of the clinical status of the different samples. **(C)** Confusion matrix for the rule set displayed in heatmap **(B)**. TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

Evaluation of a rule set

Comparing the predicted conclusions ($C_{\_predict\_j}$) with the observed clinical status (D_j), the numbers of true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*) were counted individually for the three states ('CG', 'OA', 'RA') to set up the confusion matrix. The sum of the *TP* and *TN* over the three states gives a number called 'Hits' and the sum of *FN* and *FP* a number called 'Errors'. The total sum ($n = TP + TN + FP + FN$) equals the number of samples.

The following measures were calculated to assess the quality of the classification:

Sensitivity for the classification of RA = TP_{RA} / (TP_{RA} + FN_{RA} + FP_{OA}); all values derived from the column clinical status RA in the respective confusion matrix

Sensitivity for the classification of OA = TP_{OA} / (TP_{OA} + FN_{OA} + FP_{RA}); all values derived from the column clinical status OA in the respective confusion matrix

Specificity for the classification of RA = TN_{RA} / (TN_{RA} + FP_{RA}); with $TN_{RA} = TN + FN_{OA} + TP_{OA} + FP_{OA}$ (latter value derived from the column clinical status CG) and with the value for FP_{RA} representing the sum of the two corresponding fields in the columns clinical status CG and OA of the respective confusion matrix

Specificity for the classification of OA = TN_{OA} / (TN_{OA} + FP_{OA}); with $TN_{OA} = TN + FN_{RA} + TP_{RA} + FP_{RA}$ (latter value derived from the column clinical status CG) and with the value for FP_{OA} representing the sum of the two corresponding fields in the columns clinical status CG and RA of the respective confusion matrix

Overall specificity (RA + OA) = $TN/(TN + FP_{OA} + FP_{RA})$; all values derived from the column clinical status CG in the respective confusion matrix

Accuracy = $(TN + TP_{OA} + TP_{RA})/n$

The sensitivities were calculated on the basis of the numbers from the corresponding columns of the

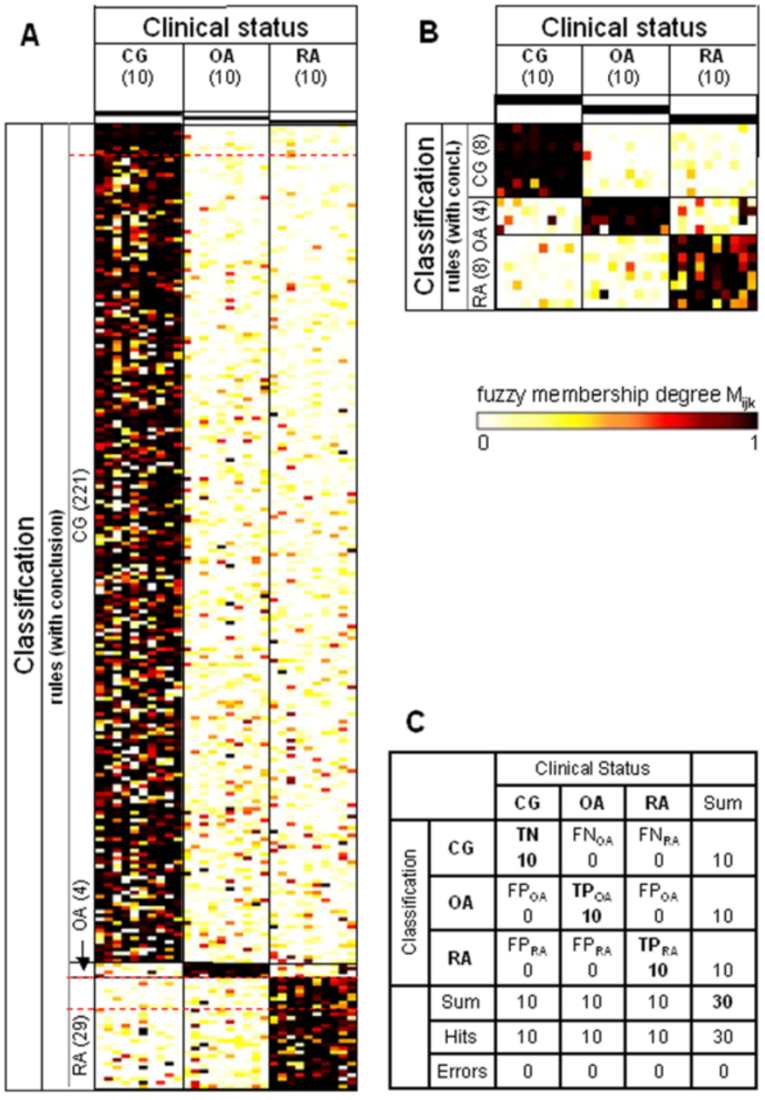


Figure 3 Heatmaps and confusion matrix for the study group 'Berlin'. Data were obtained using the study group 'Berlin' (10 control group (CG), 10 osteoarthritis (OA), 10 rheumatoid arthritis (RA)) as the training set for the rule generation and re-applying the respective rules to the same dataset. **(A)** Heatmap of the membership weights applying all rules of the primary rule set (α = 0.94; 'CG', 221 rules; 'OA', four rules; 'RA', 29 rules) for the prediction of the clinical status of the different samples; dashed red lines indicate the lower limits of the respective pruned lists of rules subsequently applied in **(B)**. **(B)** Heatmap of the membership weights applying pruned lists of rules (α = 0.94; 'CG', eight top-ranked rules; 'OA', four top-ranked rules; 'RA', eight top-ranked rules) for optimized prediction of the clinical status of the different samples. **(C)** Confusion matrix for the rule set displayed in heatmap **(B)**. TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

confusion matrix (see above). FN_{RA} represents the number of classifications as 'CG' if the ('true') clinical state was RA, and FN_{OA} the number of classifications as 'CG' if the ('true') clinical state was OA. For the study group 'Leipzig', which contains no control group ('CG'), FP_{RA} represents the misclassifications as 'RA', if the ('true') clinical status was OA, and FP_{OA} represents the misclassifications as 'OA', if the ('true') clinical status was RA.

Identification of biologically relevant molecules

Functional relations between the genes selected by the rule-based approach (total of 57) were screened using Pathway Studio (P9, version from 18 February 2013) following identification of synonyms in GeneCard (Weizmann Institute of Science, Rehovot, Israel. In addition, gene enrichment analysis was performed using the tool DAVID [47] to identify overrepresented GO-terms or KEEG pathways for the clinical states 'CG', 'OA', or 'RA' in the dataset 'Total'.

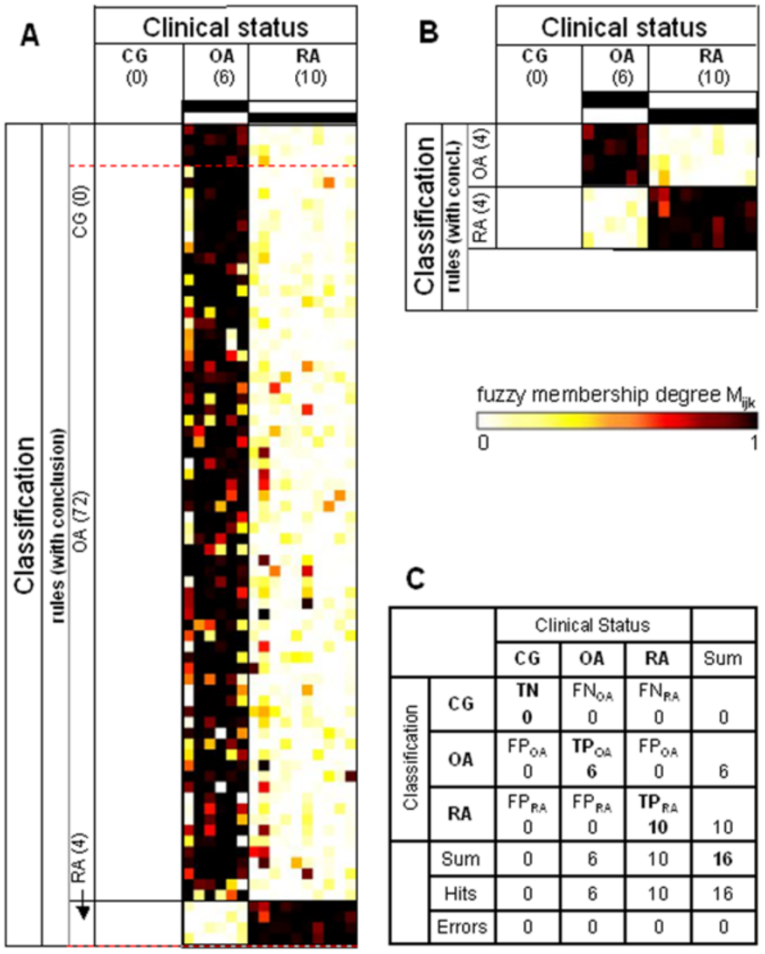


Figure 4 Heatmaps and confusion matrix for the study group 'Leipzig'. Data were obtained using the study group 'Leipzig' (0 control group (CG), six osteoarthritis (OA), 10 rheumatoid arthritis (RA)) as the training set for the rule generation and re-applying the respective rules to the same dataset. **(A)** Heatmap of the membership weights applying all rules of the primary rule set (α = 0.85; 'CG', zero rules; 'OA', 72 rules; 'RA', four rules) for the prediction of the clinical status of the different samples; dashed red lines indicate the lower limits of the respective pruned lists of rules subsequently applied in (B). **(B)** Heatmap of the membership weights applying pruned lists of rules (α = 0.85; 'CG', zero top-ranked rules; 'OA', four top-ranked rules; 'RA', four top-ranked rules) for optimized prediction of the clinical status of the different samples. **(C)** Confusion matrix for the rule set displayed in heatmap (B). TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

Results

In the first step, classifiers that discriminated between 'RA' patients, 'OA' patients, and healthy controls ('CG') were separately trained for each of the study groups and were subsequently applied (tested) for the other study groups not initially used for training.

Training of the classifiers

The significance level $alpha_S$ were set to the default value of 0.95 for 'Jena_all' ($n = 33$ patients/samples) and for 'Total' ($n = 79$). For the other study groups, $alpha_S$ was reduced to 0.94 for 'Jena' ($n = 33$) and 'Berlin' ($n = 30$) and to 0.85 for 'Leipzig' ($n = 16$), as described in Materials and methods. $alpha_S$ thus depended on both the sample size n and number m of considered probe sets (see below).

$alpha_{S\text{random}}$, for which at least one rule was randomly generated for each of the three conclusions, was between 0.01 and 0.10 smaller than the $alpha_S$ used for generation of the primary rule sets (see Additional file 2 and Materials and methods for details).

The training results obtained for the study group 'Jena_all' are shown in Figure 1. After primary rule generation, 45, seven, and 27 rules were obtained for the clinical states 'CG', 'OA', and 'RA', respectively (that is, the numbers r_{max} ('CG', 'Jena_all'), r_{max}('OA', 'Jena_all'), and r_{max}('RA', 'Jena_all')). The corresponding rule sets are listed in Additional file 3. For each rule ($r = 1, \ldots, r_{max}(C, \text{'Jena\_all'})$) and each sample (total of 33 patients; 10 CG, 10 OA, and 13 RA), the membership weight (W_r; calculated by the fuzzy membership degree) is displayed as a heat map in Figure 1A. After pruning, seven rules were selected for each

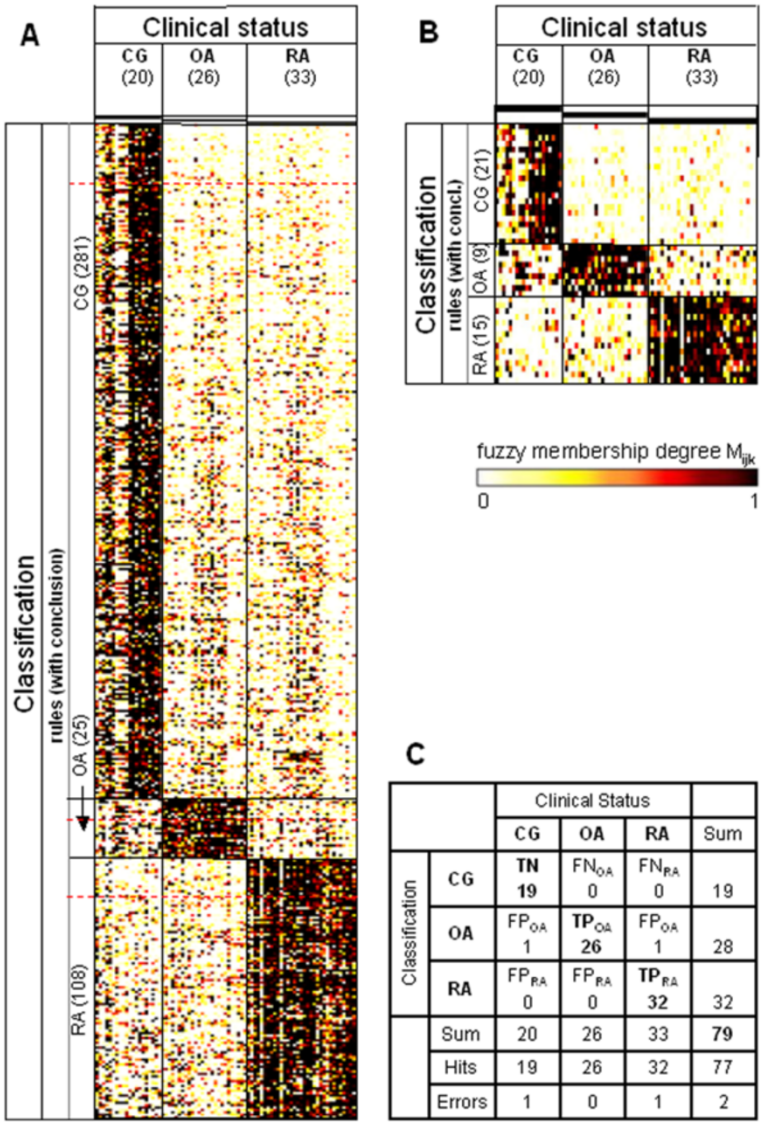


Figure 5 Heatmaps and confusion matrix for the study group 'Total'. Data were obtained using the study group 'Total' (pooled data from the three centers; 20 control group (CG), 26 osteoarthritis (OA), 33 rheumatoid arthritis (RA)) as the training set for the rule generation and re-applying the respective rules to the same dataset. **(A)** Heatmap of the membership weights applying all rules of the primary rule set (α = 0.95; 'CG', 281 rules; 'OA', 25 rules; 'RA', 108 rules) for the prediction of the clinical status of the different samples; dashed red lines indicate the lower limits of the respective pruned lists of rules subsequently applied in **(B)**. **(B)** Heatmap of the membership weights applying pruned lists of rules (α = 0.95; 'CG', 21 top-ranked rules; 'OA', nine top-ranked rules; 'RA', 15 top-ranked rules) for optimized prediction of the clinical status of the different samples. **(C)** Confusion matrix for the rule set displayed in heatmap **(B)**. TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

of the conclusions (Figure 1B). Figure 1C and Table 3 display the confusion matrix and quality parameters of the training results. Except for the sensitivity for OA (90%) and the accuracy (97%), all quality parameters reached 100%.

The following results are restricted to probe sets that were qualified by a 'present call' for all samples of the respective dataset. In the case of the dataset 'Jena', a number *m* of 7,768 probe sets was considered, for 'Berlin' 5,159 probe sets, for 'Leipzig' 8,539 probe sets, and for 'Total' 4,982 probe sets.

Using the reduced dataset for 'Jena', a total of 61 rules was generated (31 rules for 'CG', 10 rules for 'OA', and 20 rules for 'RA') as shown in Figure 2A. This primary rule set was pruned to a set of 29 rules, whose performance is displayed in Figure 2B,C. The rule set trained with the data of the study group 'Jena' and applied to the same dataset resulted in zero errors (Figure 2C) and an optimization of all quality parameters to 100% (Table 3).

The same type of analysis (application of 'present calls'; rule set training) was performed for the study

Table 3 Optimized number of pruned rules ($r_{opt}(C, S)$)[a] and assessment of training results

Study group *S*	'Jena_all'	'Jena'	'Berlin'	'Leipzig'	'Total'
Figure	1	2	3	4	5
Number of rules for 'CG'	7	9	8	0	21
Number of rules for 'OA'	7	10	4	4	9
Number of rules for 'RA'	7	10	8	4	15
Sensitivity for RA (%)	100	100	100	100	97
Sensitivity for OA (%)	90	100	100	100	100
Specificity for RA (%)	100	100	100	100	100
Specificity for OA (%)	100	100	100	100	96.2
Overall specificity (RA + OA) (%)	100	100	100	n.a.	95
Accuracy (%)	97	100	100	100	97.5

CG, control group; n.a., not applicable; OA, osteoarthritis; RA, rheumatoid arthritis. [a]See Additional file 3.

groups 'Berlin' and 'Leipzig' (Figures 3 and 4; summary in Table 3). Again, rule sets trained in and re-applied to the same dataset resulted in zero errors (Figures 3C and 4C). For the study group 'Leipzig', however, the overall specificity could not be estimated due to missing data in the control group ('CG'). Rule set training in the pooled 79 samples from the study groups 'Jena', 'Berlin', and 'Leipzig' (named study group 'Total') resulted in the rules displayed in Figure 5 and in only two errors (77 truly classified samples; Figure 5C).

Internal validation of pruned rule sets from the three clinical centers by leave-one-out cross-validation and bootstrapping resulted in acceptable error rates (see Additional file 2).

Testing of the classifiers

The classifiers separately trained in the study groups 'Jena', 'Berlin', and "Leipzig' (see Figures 2, 3 and 4) were next applied to the respective other study groups not used for training (Table 4). The average accuracy was approximately 91%, ranging from 80 to 100%. The mean

Table 4 Assessment of test results

Training set from study group	'Jena'	'Jena'	'Berlin'	'Berlin'	'Leipzig'	'Leipzig'
Test set from study group	'Berlin'	'Leipzig'	'Jena'	'Leipzig'	'Jena'	'Berlin'
Number of rules for 'CG'	9	9	8	8	0	0
Number of rules for 'OA'	10	10	4	4	4	4
Number of rules for 'RA'	10	10	8	8	4	4
Sensitivity for RA (%)	100	100	92.3	100	92.3	90
Sensitivity for OA (%)	40	100	90	83.3	100	100
Specificity for RA (%)	100	100	80	83.3	100	100
Specificity for OA (%)	100	100	91.3	100	92.3	90
Overall specificity (RA/OA) (%)	100	n.a.	60	n.a.	n.a.	n.a.
Accuracy (%)	80	100	81.8	93.8	95.6	95
Test samples	30	16	33	16	23	20
Hits for CG	10	0	6	0	0	0
Hits for OA	4	6	9	5	10	10
Hits for RA	10	10	12	10	12	9
Hits total	24	16	27	15	22	19
Errors for CG	0	0	4	0	0	0
Errors for OA	6	0	1	1	0	0
Errors for RA	0	0	1	0	1	1
Errors total	6	0	6	1	1	1

CG, control group; n.a., not applicable; OA, osteoarthritis; RA, rheumatoid arthritis.

sensitivity for the prediction of RA was 96%, ranging from 90 to 100%; and that for the prediction of 'OA' was 86%, ranging from 40 to 100%.

The number of 'Errors' for the prediction of RA was generally extremely small; in three cases ('Jena' → 'Berlin', 'Jena' → 'Leipzig', and 'Berlin' → 'Leipzig'), no errors were detected; in the remaining cases there was only one error each (1/13, 1/13, and 1/10, respectively).

For the remaining two clinical states (that is, 'CG' and 'OA') more errors were detected. In the case of 'Jena' → 'Berlin', six OA patients were misclassified as 'CG'; whereas in the case of 'Berlin' → 'Jena', three CG samples were misclassified as 'RA' and one CG sample as, OA in addition to one OA patient being misclassified as 'RA'.

Molecular interpretation of the obtained rule sets

The complete overlap of all rules (that is, premises and conclusion) resulting from the comparison of all study groups before pruning is shown in Additional files 3 and 4 (please note the cross-table listing of the overlapping genes in Table B of the sheet 'Rule Overlap among Data Sets' in Additional file 3).

If, for the purpose of identifying biologically relevant classifiers, the overlap analysis is focused on the three independent study groups 'Jena', 'Berlin', and 'Leipzig', a list of selected potential 'key' players can be extracted (Table 5).

Whereas no overlap between these groups was found for rules with the conclusion 'OA', remarkable overlap was found for the conclusions 'CG' and 'RA'.

The rule 'IF NFIL3 is highly expressed THEN CG' (with NFIL3 coding for the nuclear factor interleukin-3-regulated protein) was generated with high relevance from both the 'Jena' and the 'Berlin' datasets (ranked in third and fourth position, respectively; Table 5). In addition, the two genes MAT2A (methionine adenosyltransferase 2A) and TIPARP (2,3,7,8-tetrachlorodibenvzo-*p*-dioxin (TCDD)-inducible poly (ADP-ribose) polymerase) were identified in prominent rules for 'CG', each only present in the pruned rule set of one study group.

For the conclusion 'RA', the rules concerning the 'high' expression of the genes STAT1, GBP1, PLCG2, CSF2RB, and STK10 were highly ranked in pruned rule sets from different study groups. STAT1 (signal transducer and activator of transcription 1) was found in the pruned rule set 'Berlin' (rank 1), and GBP1 (interferon-inducible guanylate binding protein 1) in the pruned rule sets 'Jena' (rank 2) and 'Berlin' (ranks 2 and 8). PLCG2 (phospholipase c-gamma-2) was found in the pruned rule set 'Berlin' (rank 5), and STK10 (serine/threonine kinase 10) in the pruned rule set 'Jena' (rank 5).

Strikingly, the relevance of the rule 'IF CSF2RB is highly expressed THEN RA' (CSF2RB coding for the interleukin 3 receptor/granulocyte-macrophage colony stimulating factor 3 receptor, beta was supported by three different features: the rule was independently detected in the rule sets derived from all three centers ('Jena', 'Berlin', and 'Leipzig'); the rule occupied the highest rank (rank 1) in the rule set from 'Leipzig'; and its complementary rule 'IF *CSF2RB* is low THEN OA' was also detected in the rule set 'Leipzig' with rank 3 (see Additional file 3).

To address a potential pathogenetic role of the genes indicated in Table 5, their expression was compared among the three different clinical states (both

Table 5 Overlap between the three independent study groups

Gene symbol	Probe set name	Expression level	'Jena' rule rank	'Berlin' rule rank	'Leipzig' rule rank	'Total' rule rank
'CG'						
NFIL3	**203574_at**	**High**	**3**	**4**		1
JUND	203752_s_at	High	11	85		18
MAT2A	**200768_s_at**	**High**	**2**	83		5
TIPARP	**212665_at**	**High**	12	**8**		
LEPROTL1	202594_at	Low	27	127		113
'RA'						
STAT1	**M97935_3_at [200887_s_at]**	High	19	**1** (& 10)		2 (& 10)
GBP1	**202270_at [202269_x_at]**	**High**	**2**	**2 (& 8)**		5 (& 6)
PSMB9	204279_at	High	13	17		1
PLCG2	**204613_at**	**High**	14	**5**		4
LY75	205668_at	High	12	26		8
CSF2RB	**205159_at**	**High**	17	28	**1**	3
STK10	**40420_at**	**High**	**5**	21		12

The genes belonging to at least one of the pruned rule sets of the three independent study groups are highlighted in bold, genes/rules detected by two different probe sets are indicated by numbers in parentheses.

individually for the three different clinical centers and for the pooled study group 'Total' derived from all centers). In support of their relevance, all genes/rules characterizing 'CG' were significantly overexpressed in CG as compared with both RA and OA (Additional file 5) – with the exception of the gene/rule LEPROTL1 (leptin receptor overlapping transcript 1), which also showed significant differences, but with an opposite orientation (all $P \leq 0.05$; Mann Whitney U test).

Strikingly, all genes/rules identified for RA also appeared highly discriminative, as shown by a significant overexpression in RA in comparison with both CG and OA (P values between 10^{-11} and 0.05 for 41/42 comparisons; $P = 0.056$ for the remaining comparison; Additional file 5).

In addition to the analysis of the overlapping rules, all 57 rules generated from the different study groups after pruning – that is, 29 rules trained from the dataset 'Jena', 20 from 'Berlin', and eight from 'Leipzig' (highlighted in the complete rule set in Additional file 3) – were screened for functional relations using Pathway Studio following identification of synonyms in GeneCard.

Since for three Affymetrix probe sets no gene names were identified (see Additional file 6), only 54 genes were analyzed using Pathway Studio. The results of the Pathway Studio search for the conclusions 'CG' and 'RA' are shown in Additional files 7 and 8, respectively.

Again, no relations were found for the conclusion 'OA'. For 'RA', instead, three relations were found (Table 6). In addition to the well-known relation JAK2 → STAT1, which regards various cell types including fibroblasts (total of 70 references named by Pathway Studio), the relation STAT1 → GBP1 [48-50] and the relation JAK2 → CSF2RB [51-53] have only been addressed by a limited number of publications.

Please note that JAK2 is not contained in Table 5 since it was only detected in the rule set for 'RA' in the study group 'Jena' (rank 3).

Gene enrichment analysis for molecular interpretation of the obtained rule sets resulted in additional information. In CG, for example, there was low expression of genes involved in MHC class II antigen processing/presentation (Additional file 9, sheets 'CG Low BP' and 'CG Low KEGG'). In RA, in contrast, there was high expression of genes involved in immune response in general and leukocyte/T-cell/B-cell activation (Additional file 10, sheets 'RA High BP' and 'RA High KEGG'), as well as programmed cell death (Additional file 10, sheets 'RA High BP', 'RA High KEGG', and 'RA Low BP').

As already observed for the sensitivity and accuracy, as well as the rule overlap and molecular interpretation, OA patients were again more difficult to discriminate, as indicated by the almost complete absence of indicative GO terms or KEGG pathways in gene enrichment analysis (Additional file 11).

Table 6 Interactions between the premises/genes of the pruned rule sets generated from the 'Jena', 'Berlin', and 'Leipzig' data sets (total of 57 rules), as found by Pathway Studio and exemplified for the conclusion 'RA'

Relation	Type	Cell type	Number of references
JAK2 → STAT1	Promoter binding	Various	(70)
JAK2 → CSF2RB	Regulation	Hematopoietic cells	(3)
STAT1 → GBP1	Protein modification	Fibroblasts	(3)

For more details, see Additional file 8. Pathway Studio from Elsevier, Munich, Germany.

Discussion

In the present study, three multicenter, genome-wide transcriptomic datasets from a total of 79 individuals were used to infer rule-based classifiers to discriminate RA, OA, and healthy controls. In all cases, the rule sets were inferred separately from one of three centers and applied to the other centers for validation. This novel approach resulted in a high performance (close to 90% for specificity, sensitivity, and accuracy) for the discrimination of RA. Unbiased analysis of the biological relevance of the underlying rules by Pathway Studio resulted in the identification of pathways with known pathogenetic or therapeutic relevance in RA. In addition, serine/threonine kinase 10 (lymphocyte-oriented kinase) was identified as a novel molecule with a potential role in RA. Yet another novel contribution of the present study is the identification of molecules that identify normal synovial tissue, an aspect barely addressed to date.

New approach for the identification of discriminating genes and/or rules

A novel rule-based approach was used to identify genes (in combination with their expression status) suitable for the discrimination of the clinical states healthy controls ('CG'), 'OA', and 'RA'. This approach has the major advantage of skipping the identification of differentially expressed genes on the basis of fold changes and/or *t*-test or *U*-test analysis, a process highly sensitive to heterogeneity in the patient data and therefore often incapable of identifying relevant disease-specific genes.

The rule-based approach applied in the present study is based on the relevance index of Krone and Kiendl [40]; this relevance index has so far only been used for rule generation in electrical control engineering [41] or biotechnology [38]. In addition, there are only few examples for the application of this relevance index to omics data (for example [54]) and, to our knowledge, none for the application to data in the rheumatology field.

Rule set pruning, applied in order to minimize the numbers of both rules and 'Errors', was successfully used

to avoid overfitting and informative imbalance [55]. From our experience with heuristic rules, at least four rules per conclusion were required [38,55].

Quality parameters of the training results

For the datasets 'Jena', 'Berlin', and 'Leipzig', the values for disease-oriented sensitivity and specificity, overall specificity, and accuracy were all 100%. This high performance for the training of the classifiers was expected, but still shows that this approach is suitable for the analysis of gene expression data from synovial tissue.

Interestingly, the disease-specific sensitivity for OA in the 'Jena_all' dataset was only 90%, resulting in an accuracy of 97% (see Table 3), whereas the quality parameters in the 'Jena' dataset all reached 100%. This is probably due to the highly stringent approach of only using probe sets with a 'present call' in all samples, deliberately chosen to minimize false positives. This approach is further supported by reduced error rates in the internal validation of the "Jena' dataset in comparison with the 'Jena_all' dataset (see Additional file 2).

The results for the quality parameters in the largest possible dataset 'Total', containing 19 CG, 26 OA, and 32 RA, also proved highly satisfactory; that is, ≥95%. This further underlines the suitability of the relevance index approach for large-scale clinical studies with high numbers of RA and OA patients [27,30].

Quality parameters of the test results

The quality parameters of the test results for the prediction of RA were also highly satisfactory; that is, they showed a mean close to or higher than 90% for all assessment parameters (see Table 4). This shows that the real challenge of the present study – that is, the prediction of RA in test datasets independent of the training dataset – can be met with a high accuracy and may indeed contribute to the identification of biomarkers for RA.

Notably, the mean sensitivity and specificity for the prediction for OA was considerably lower than for RA, due to both misclassification of OA as 'CG' (six cases) or as 'RA' (two cases). This is consistent with the clinical problem of properly differentiating burnt-out, possibly more heterogeneous, OA with low inflammatory activity from normal controls on one hand, and active, highly inflammatory OA from RA on the other [1,2].

Molecular interpretation of the obtained rule sets

The number of studies aimed at identifying disease-specific signatures in rheumatology with microarray-based methods is limited [30,31,35,56-60]. Also, very few datasets addressing this question are publicly available and have been repeatedly used for bioinformatic analyses. In addition, with one exception [57], these studies have not analyzed matched multicenter datasets for rheumatic diseases. Finally, studies have resulted in the identification of numerous and heterogeneous biomarker genes or pathways with only limited overlap among the results of the different studies.

In the present study, in contrast, several rules were identified in more than one rule set generated in the three clinical centers; that is, five rules for the prediction of healthy controls (CG) and seven rules for the prediction of RA (see Table 5). Notably, a total of seven of these rules were represented not only in the primary rule set of the centers, but also in one or more of the respective pruned rule sets. Strikingly, no overlapping rules were observed for 'OA', again underlining the problem of properly differentiating OA from either CG or RA (see above for the Quality parameters of the test results).

In addition, automated analysis of interactions by Pathway Studio between the molecules identified in the union of all optimized rule sets (total of 57 rules; derived from three clinical centers with either two or three disease states) resulted in three interactions supported by at least three references; that is, JAK2 → STAT1 (70 references), STAT1 → GBP1 (three references) and JAK2 → CSF2RB (three references; see Table 6). Please note that JAK2 was only detected once at rank 3 in the 'Jena' rule set (see Additional file 3) and is therefore not listed in Table 5.

Rules for the prediction of healthy controls (CG)

The genes identified above as overexpressed in CG may represent a core set of markers of healthy tissue and reflect regulatory genes specifically involved in the down-regulation/prevention of rheumatic diseases (that is, OA or RA).

Nuclear factor interleukin-3-regulated protein

NFIL3 is a basic leucine transcription factor acting as a regulator of genes associated with acquired and innate immunity (for example, interleukin (IL)-3 and interferon-gamma (IFNγ) [61]) or with the inhibition of proliferation and senescence [62]. In particular, NFIL-3 negatively regulates IL-12 p40 in macrophages and dendritic cells [63,64] and suppresses TH2 cytokines [65], as well as the development and maturation of NK cells [66]. In addition, NFIL3 exhibits anti-apoptotic features [67]. In particular, the role of NFIL3 in limiting the production of proinflammatory IL-12 may explain its upregulation in the normal CG. On the other hand, its prominent influence on essential cellular features (for example, metabolism, growth, viability) points to a potential contribution to the pathogenesis of RA (and/or OA) in the case of dysregulated underexpression.

Jun D proto-oncogene

Members of the JUN and FOS families are known as immediate-early response proto-oncogenes, since they

are rapidly induced by various activating agents and, on the other hand, have a very short half-life (in the range of minutes for both mRNA and protein) [68]. As in the case of NFIL3, the transcription factor JunD also regulates genes involved in acquired and innate immunity [69], in proliferation and senescence [70], or in anti-apoptotic effects [71,72].

Individual JUN/FOS family molecules show different biological activities. Whereas C-JUN and C-FOS are known as activating proto-oncogenes with transforming activity [73,74], JUND also shows de-activating features [68,73,75-77]. The effects of AP-1 complexes composed of different JUN/FOS family members clearly depend on the local promoter context of genes driven by AP-1 (for example, MMP-1 [78,79]). JUND suppresses synovial fibroblast proliferation and even antagonizes Ras-mediated transformation of the fibroblasts [77], and thus its overexpression may exert a protective role in the synovial membrane of normal joints.

Methionine adenosyltransferase 2A

The importance of the overexpression of MAT2A in CG samples is presently unclear. This molecule is involved in the regulation of basic cellular functions, such as the synthesis of polyamines (thought to play a role in nucleic acid and protein synthesis) and developmental processes [80].

2,3,7,8-tetrachlorodibenzo-*p*-dioxin-inducible poly(ADP-ribose) polymerase (TIPARP)

Poly(ADP-ribosyl)ation physiologically contributes to the survival of damaged proliferating cells by immediate, DNA damage-dependent post-translational modification of histones and other proteins in the nucleus. By this process, poly(ADP-ribose) polymerases are involved in cellular functions such as proliferation and cell death. It is to be expected that the growing poly(ADP-ribose) polymerase superfamily may become the target of pharmacological strategies enhancing both antitumor efficacy and the treatment of a number of inflammatory and neurodegenerative disorders [81].

TiPARP (PARP-7) was originally identified by differential display as a TCDD-induced mRNA [82]. Although the exact function of TiPARP is presently unclear, its effects on T-cell function and its possible contribution to tumor promotion suggest a role also in the normal or arthritic synovial membrane [81].

Leptin receptor overlapping transcript-like 1

The leptin receptor overlapping transcript (also called OB-RGRP [83]) is one of the three members of a gene family [84,85]. Leptin receptor overlapping transcript molecules are small proteins of 131 to 140 amino acids, carrying four potential transmembrane domains.

LEPROTL1, a gene widely expressed in human tissues, including metabolic tissues such as muscle and liver [83,84,86], has an influence on growth, plasma insulin-like growth factor-1 levels, hepatic sensitivity to growth hormone, and cell-surface growth hormone or leptin receptor expression and leptin signaling [87,88].

The high importance of LEPROTL1in protein trafficking to the vacuole/lysosome of eukaryotic cells, a process initially regarded as pathogenetically relevant in RA [89-91], and in the downregulation of membrane protein levels suggests a phylogenetically conserved role for LEPROTL1 [85]. Since LEPROTL1 does not appear to act as a classic leptin receptor, its role in the physiology and pathophysiology of the synovial membrane is presently uncertain.

In the present dataset, the above-mentioned NFIL3, JUND, MAT2A, and TIPARP were indeed significantly overexpressed in the synovial membrane of CG as compared with both RA and OA (both individually for the three different clinical centers and for the pooled study group 'Total' derived from all centers; Additional file 5). Interestingly, overexpression of JUND (OA vs. RA) has not only been observed in synovial membranes, but also in proinflammatory synovial fibroblasts isolated from synovial tissue [92].

In contrast, LEPROTL1 was the only gene significantly underexpressed in the synovial membrane of CG as compared with both RA and OA, suggesting that this molecule may support inflammatory and/or degenerative joint diseases. Similarly to JUND, however, in an opposite direction, differential expression of LEPROTL1 was not only observed in synovial membranes, but also in resident synovial fibroblasts [92].

Rules for the prediction of rheumatoid arthritis

The genes overexpressed in RA synovial tissue (see Table 5) may represent biomarkers of RA and reflect processes specifically involved in the pathogenesis and/or progression of the disease. A disease specificity of the markers is strongly supported by their significant overexpression in RA, not only in comparison with CG but also with the 'disease' control OA (see Additional file 5). In the RA groups, genes especially associated with the regulation of immunologic processes appear to be suitable as disease-specific identifiers.

Signal transducer and activator of transcription 1

STAT1, a transcription factor regulating (amongst others) immunity-mediating genes, is known to be upregulated in RA patients [59,93]. In addition to other transcription factors (for example, NFKB or AP-1), STAT1 has long been regarded as a pivotal transcription factor involved in joint inflammation and destruction [60,94]. The identification of these key factors underlines the robustness of the present approach. This is further underlined by the fact

that the rule 'STAT1 high in RA' appears a total of five times in three different rule sets (rule set 'Jena', position 19; rule set 'Berlin', positions 1 and 10; rule set 'Total', positions 2 and 10; see Table 5 and Additional file 4 for details and the corresponding Affymetrix probe sets).

In addition, there was a reciprocal detection of the complementary rule 'IF STAT1 is low THEN OA' in the rule set 'Leipzig' with rank 12 (see Additional file 3).

Interferon-inducible guanylate binding protein 1

GBP1, a protein specifically binding guanylated nucleotides with potential effects on GTPases involved in signal transduction, has been already described as upregulated in RA versus OA synovial tissue [95]. Also, this factor is implicated in the pathogenesis of RA due to its upregulation by IFNγ [95,96]. As in the case of STAT1, this finding confirms that key mediators of rheumatoid inflammation have been identified in the present study. This is again further underlined by the fact that the rule 'GBP1 high in RA' appears a total of five times in three different rule sets (rule set 'Jena', position 2; rule set 'Berlin', positions 2 and 8; rule set 'Total', positions 5 and 6; see Table 5 and Additional file 4).

Proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2/low molecular mass protein 2)

The proteasomal subunit PSMB9 (also known as LMP2; see abbreviations) is involved in the regulation of proteolytic specificity, especially in response to IFN-γ, thus enabling the formation of immunoproteasomes and the generation of peptides presentable by MHC I molecules [97]. PSMB9 also enhances cytokine production (for example, tumor necrosis factor, IL-1β, IFNγ [98]). Indeed, this molecule shows a significant genetic association with RA in ethnic Han Chinese from Yunan [99] and is the target of autoimmune reactions in RA [100]. As for STAT1 and GBP1, the validity of the rule 'PSMB9 high in RA' is emphasized by the fact that it appears in three different rule sets (rule set 'Berlin', position 13; rule set 'Leipzig', position 17; rule set 'Total', position 1; see Table 5 and Additional file 4).

Phospholipase C-gamma-2

The function of members of the phospholipase C family is the hydrolysis of phospholipids into fatty acids and other lipophilic molecules. The family members are grouped into several subtypes and catalyze the hydrolysis of phosphatidylinositol 4,5-bisphosphate to inositol 1,4,5-trisphosphate and 1,2-diacylglycerol, which both have important second messenger functions. Phospholipase C-gamma is activated by phosphorylation in response to various growth factors or immune signals, is broadly expressed, and carries diverse biological functions in inflammation, cell growth, signaling/death, and maintenance of membrane phospholipids. Activating mutations in the PLCG2 gene have been shown to induce autoimmunity, inflammation, and/or inflammatory arthritis in murine models [101,102]. PLCG2 has already been recognized as an excellent discriminator of RA against other types of arthritis or autoimmune diseases [103] and appears to be significantly upregulated in RA synovial tissue as compared with the normal synovial membrane [104]. As for STAT1, GBP1, and PSMB9/LMP2, the validity of the rule 'PLCG2 high in RA' was emphasized by its appearance in three independently established rule sets (rule set 'Berlin', position 5; rule set 'Jena', position 14; rule set 'Total', position 4; see Table 5 and Additional file 4).

Lymphocyte antigen 75

Ly75, a member of the human macrophage mannose receptor family (also known as DEC205 or GP200-MR6), supports antigen presentation of dendritic cells [105] and mediates anti-proliferative as well as promaturational effects in B lymphocytes [106]. This molecule is apparently upregulated in monocytes derived from RA patients in comparison with those from normal donors [107], but its role in disease is currently unknown. Interestingly, however, single nucleotide polymorphisms of the Ly75 antigen belong to the three single nucleotide polymorphisms most significantly associated with type 2 diabetes mellitus, leaving open a possible role of Ly75 in inflammatory disease [108].

CSF2RB (interleukin 3 receptor/granulocyte macrophage colony stimulating factor 3 receptor, beta)

A most striking finding in the present study is the rule 'CSF2RB high in RA'. CSF2RB codes for a transmembrane protein and acts as a common receptor subunit (also known as common beta chain) for granulocyte–macrophage colony-stimulating factor (GM-CSF), IL-5, and IL-3, which play a preeminent role in inflammation and hematopoiesis [109,110]. One of the ligands of CSF2RB (that is, GM-CSF) has long been implicated in the pathogenesis of RA, and other rheumatic or autoimmune diseases [60,111-119]. This has recently led to the development of neutralizing therapeutic monoclonal antibodies specifically directed against the α-chain of the GM-CSF receptor, which have been successfully used for the treatment of RA [120-122].

Notably, the rule 'CSF2RB high in RA' appeared in the independently established rule sets of all analyzed cohorts (rule set 'Jena', position 17; rule set 'Berlin', position 28; rule set 'Leipzig', position 1; and, remarkably, rule set 'Total', position 3), again underling the validity of the completely unbiased procedure of rule set generation. As in the case of STAT1, there was a reciprocal detection of the complementary rule 'IF CSF2RB is low THEN OA' in the rule set 'Leipzig' with rank 3 (Additional file 3).

Serine/threonine kinase 10 (lymphocyte-oriented kinase)
STK10 is a member of the Ste20 family of serine/threonine protein kinases with similarity to several known polo-like kinase kinases [123], which associates with and phosphorylates polo-like kinase 1 and whose functional inactivation interferes with normal cell cycle progression. STK10 also negatively regulates IL-2 expression in T cells via the mitogen-activated protein kinase kinase 1 pathway [124]. Interestingly (and potentially relevant for RA), STK10 is involved in the regulation of cytoskeletal rearrangement through phosphorylation of the ezrin–radixin–moesin proteins [125], a process also strongly emphasized by a previous report [96] and by a relatively low expression of the respective genes in the gene enrichment analysis in the 'CG' group (see Additional file 9; sheet 'CG low BP'). In addition, STK10 potentiates dexamethasone-induced apoptosis [126] and may thus contribute to the dysregulation of apoptosis possibly involved in RA [127]. Finally, STK10 may play a role in autoimmune skin diseases [128], although a direct involvement of this molecule in arthritis has never been reported.

As in the case of rules for healthy control (CG), all genes used for the prediction of RA were indeed significantly overexpressed in the synovial membrane of RA as compared with both OA and CG (both individually for the three different clinical centers and for the pooled study group 'Total' derived from all centers; see Additional file 5). Interestingly, highly significant overexpression of CSF2RB (RA vs. OA; $P = 5.4 \times 10^{-6}$) was not only observed in synovial membranes, but also in proinflammatory synovial fibroblasts isolated from synovial tissue [92].

Finally, in combination with JAK2, one of the most influential rules in the 'Jena' RA group (position 3; high in RA), a subset of the genes (STAT1, GBP1, CSF2RB) can be combined in a JAK/STAT-dependent gene regulatory network [59,60,129-131]. This also indicates that the rules identifying RA patients in the present study are not generated randomly, but reflect a mechanistic relevance within the context of RA pathogenesis. Concerning JAK2, its concrete relevance in RA is stressed by the development of therapeutic approaches directed at the JAK pathway [129].

Overall, the present study confirmed the involvement of partially or well-known molecules/pathways in RA (for example, STAT1, GBP1, PLCG2, CSF2RB), but also identified molecules previously not associated with RA (for example, STK10). Also, to our knowledge, there are at present no reports on molecules/pathways positively identifying the clinical status 'CG' in general, and the NFIL-3 pathway in particular. Finally, the present study presents for the first time a 'unifying hypothesis' by addressing the overlap of the highly ranked rules/genes among different clinical centers and thus pinning down molecules of universal relevance in heterogeneous patient cohorts from different centers. This is also supported by the representation of the top 12 rules of the 'Total' dataset in the overlap table; that is, the largest independently analyzed cohort in the present study (total of 79 donors (patients).

Conclusions

In this study, three multicenter, genome-wide transcriptomic datasets were applied to infer rule-based classifiers/genes to discriminate RA, OA, and healthy controls, and were subsequently analyzed for their biological relevance using Pathway Studio and gene enrichment analysis. This novel approach resulted in a high performance for the discrimination of RA and the identification of factors with known pathogenetic or therapeutic relevance in RA (for example, STAT1, GBP1, IFNγ, GM-CSF, and its receptor CSF2RB, as well as JAK2, the latter pointing to a JAK/STAT-dependent gene regulatory network). This indicates that the rules identifying RA patients were not generated randomly, but reflect (disease-specific) key biomarkers with mechanistic relevance for RA pathogenesis and progression, some of them well established and already exploited for therapeutic purposes.

The present study contributes to focusing the diagnostic and therapeutic interest in RA on relevant and innovative molecules or pathways; for example, GM-CSF and its receptor CSF2RB. The fact that such known pathways have been identified in the present study for the prediction of RA suggests a high sensitivity and validity of the current approach. In addition, the present study for the first time addressed a multicenter cross-validation and may thus contribute to the identification of molecules with universal relevance in heterogeneous RA patient cohorts, possibly including the previously undescribed STK10.

At a molecular level, the biomarkers were significantly overexpressed in RA synovial tissue (mostly in the study groups from all three centers), not only in comparison with healthy controls, but also with the 'disease' control OA. In addition, significant overexpression was not limited to the synovial tissue as a whole, but also occurred in isolated synovial fibroblasts, a cell population regarded as highly important for chronic inflammatory RA.

In perspective, validation, refinement, and generalization of the present rule-based, discriminative procedure in a larger prospective cohort are necessary. The identified biomarkers may prove useful for diagnosis or differential diagnosis of RA patients (including potential subpopulations), as well as for stratification and monitoring of (responders and nonresponder) patients in routine or experimental clinical trials.

Additional files

Additional file 1: Calculation of the relevance index.

Additional file 2: Internal validation of rule sets.

Additional file 3: List of the 'complete primary rule sets' for all datasets, as well as the 'Rule overlap among data sets'. The data are displayed as either 'complete primary rule sets' with the pruned (optimized) rules highlighted in bold (Sheet 1) or as the 'Rule Overlap among Data Sets' (Sheet 2) with the rules/genes showing an overlap between the three independent study groups 'Jena', 'Berlin', and 'Leipzig' highlighted in grey. In both cases, the rules were generated as stated in Materials and methods ('Rule set generation') and the ranks of the individual rules in the respective dataset are indicated.

Additional file 4: Listing of the overlap among the different rule sets. The data are displayed as the 'Rule Overlap among Data Sets' including the gene names. The ranks of the individuals rules in the respective dataset are indicated and the rules/genes showing an overlap between the three independent study groups 'Jena', 'Berlin', and 'Leipzig' are highlighted in grey.

Additional file 5: Log-fold change and *P* values for differentially expressed genes. Log-fold change (log2 FC) and *P* values (Mann Whitney U test, red: $P \geq 0.05$) for the genes differentially expressed among patients with a different clinical status (genes significantly overexpressed in RA versus both CG and OA are highlighted in grey; see also Table 6).

Additional file 6: Genes (original symbols) and the synonyms used as input for the Pathway Studio 9 search for interactions among the genes.

Additional file 7: Interactions among the genes in the pruned rule sets (CG). Interactions found by Pathway Studio among the genes contained in the pruned rule sets of the 'Jena' and 'Berlin' datasets for the conclusion 'CG'.

Additional file 8: Interactions among the genes in the pruned rule sets (RA). Interactions found by Pathway Studio among the genes contained in the pruned rule sets of the 'Jena', 'Berlin' and 'Leipzig' datasets for the conclusion 'RA'.

Additional file 9: Gene enrichment analysis for molecular interpretation (CG). Gene enrichment analysis for molecular interpretation of the obtained rule set for the conclusion 'CG' using the GO terms biological process (BP) and molecular function (MF), as well as KEGG pathways. The analyses were performed separately for the 'CG' rules showing a high or low expression level. Category = type of term (GO term/KEEG pathway); Term = denomination of term (interesting terms highlighted in grey); Count = list hits; number of genes in the rule set belonging to the term in question; p value = EASE score (upper boundary of the distribution of Jackknife Fisher exact probabilities given the actual Count, List Total, Pop Hits, and Pop Total); Genes = gene symbols of included rules/genes; List Total = number of genes in the rule set (for high and low expression, respectively); Pop Hits = number of genes in the population background belonging to the specific term; Pop Total = number of genes in the population background; BH-adjusted p value = Benjamini-Hochberg adjusted *P* value (threshold $P \leq 0.05$ indicated by fat frame).

Additional file 10: Gene enrichment analysis for molecular interpretation (RA). Gene enrichment analysis for molecular interpretation of the obtained rule set for the conclusion 'RA' using the GO terms biological process (BP) and molecular function (MF), as well as KEGG pathways. The analyses were performed separately for the 'RA' rules showing a high or low expression level. In the case of 'RA' rules showing a low expression level, there were only results for the GO terms BP and MF. Category = type of term (GO term/KEEG pathway); Term = denomination of term (interesting terms highlighted in grey); Count = list hits; number of genes in the rule set belonging to the term in question; p value = EASE score (upper boundary of the distribution of Jackknife Fisher exact probabilities given the actual Count, List Total, Pop Hits, and Pop Total); Genes = gene symbols of included rules/genes; List Total = number of genes in the rule set (for high and low expression, respectively); Pop Hits = number of genes in the population background belonging to the specific term; Pop Total = number of genes in the population background; BH-adjusted p value = Benjamini-Hochberg adjusted *P* value (threshold $P \leq 0.05$ indicated by fat frame).

Additional file 11: Gene enrichment analysis for molecular interpretation (OA). Gene enrichment analysis for molecular interpretation of the obtained rule set for the conclusion 'OA' using the GO terms biological process (BP) and molecular function (MF), as well as KEGG pathways. The analyses were performed separately for the 'OA' rules showing a high or low expression level. There were only results for the GO term MF in 'OA' rules showing a high expression level. Category = type of term (GO term/KEEG pathway); Term = denomination of term; Count = list hits; number of genes in the rule set belonging to the term in question; p value = EASE score (upper boundary of the distribution of Jackknife Fisher exact probabilities given the actual Count, List Total, Pop Hits, and Pop Total); Genes = gene symbols of included rules/genes; List Total = number of genes in the rule set (for high and low expression, respectively); Pop Hits = number of genes in the population background belonging to the specific term; Pop Total = number of genes in the population background; BH-adjusted p value = Benjamini-Hochberg adjusted *P* value.

Abbreviations

CG: control group; C_r: conclusion of the *r*th rule; CSF2RB: interleukin 3 receptor/granulocyte-macrophage colony stimulating factor 3 receptor, beta; GBP1: interferon-inducible guanylate binding protein 1; GM-CSF: granulocyte–macrophage colony-stimulating factor; IFNγ: interferon-gamma; IL: interleukin; JUND: jun D proto-oncogene; LEPROTL1: leptin receptor overlapping transcript-like 1; LY75: lymphocyte antigen 75; MAT2A: methionine adenosyltransferase 2A; NFIL3: nuclear factor interleukin-3-regulated protein; OA: osteoarthritis; PLCG2: phospholipase C-gamma-2; P_r: premise of the *r*th rule; PSMB9/LMP2: proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)/low molecular mass protein 2; RA: rheumatoid arthritis; RI: relevance index; STAT1: signal transducer and activator of transcription 1; STK10: serine/threonine kinase 10 (lymphocyte-oriented kinase); TCDD: 2,3,7,8-tetrachlorodibenzo-*p*-dioxin.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

DW, PK, MP, RG, and DD performed the bioinformatic analysis, contributed to the design of the study, and participated in the writing and finalization of the manuscript. RWK, RG, PS, RH, and TH contributed to the design of the study and participated in the layout, writing, and finalization of the manuscript. RH, DP, TH, PS, DK, and RWK designed or performed the experiments and participated in writing and finalization of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the German Federal Ministry of Education and Research (BMBF FKZ 0315719A and FKZ 0315719B; ERASysBio PLUS; LINCONET).

Author details

[1]BioControl Jena GmbH, Wildenbruchstraße 15, 07745 Jena, Germany. [2]Experimental Rheumatology Unit, Department of Orthopedics, Jena University Hospital, Waldkrankenhaus Rudolf Elle, Klosterlausnitzer Straße 81, 07607 Eisenberg, Germany. [3]Institute of Clinical Chemistry, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany. [4]Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Beutenbergstraße 11a, 07745 Jena, Germany. [5]Present address: Center of Diagnostics GmbH, Chemnitz Hospital, Flemmingstr. 2, 09116 Chemnitz, Germany. [6]Department of Medical Engineering and Biotechnology, University of Applied Sciences Jena, Carl-Zeiss-Promenade 2, 07745 Jena, Germany. [7]Department of Rheumatology and Clinical Immunology, Charite-Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany. [8]Institute of Immunology, University of Rostock, Schillingallee 68, 18057 Rostock, Germany. [9]Institute of Pathology, University of Leipzig, Liebigstraße 24, 04103 Leipzig, Germany.

Received: 12 July 2013 Accepted: 10 March 2014
Published: 1 April 2014

References

1. Murphy G, Nagase H: **Reappraising metalloproteinases in rheumatoid arthritis and osteoarthritis: destruction or repair?** *Nat Clin Pract Rheumatol* 2008, **4:**128–135.
2. de Lange-Brokaar BJ, Ioan-Facsinay A, van Osch GJ, Zuurmond AM, Schoones J, Toes RE, Huizinga TW, Kloppenburg M: **Synovial inflammation, immune cells and their cytokines in osteoarthritis: a review.** *Osteoarthritis Cartilage* 2012, **20:**1484–1499.
3. Choy E: **Understanding the dynamics: pathways involved in the pathogenesis of rheumatoid arthritis.** *Rheumatology (Oxford)* 2012, **51:**v3–v11.
4. Firestein GS: **Evolving concepts of rheumatoid arthritis.** *Nature* 2003, **423:**356–361.
5. Isaacs JD: **The changing face of rheumatoid arthritis: sustained remission for all?** *Nat Rev Immunol* 2010, **10:**605–611.
6. Rousseau JC, Delmas PD: **Biological markers in osteoarthritis.** *Nat Clin Pract Rheumatol* 2007, **3:**346–356.
7. Haseeb A, Haqqi TM: **Immunopathogenesis of osteoarthritis.** *Clin Immunol* 2013, **146:**185–196.
8. Reines BP: **Is rheumatoid arthritis premature osteoarthritis with fetal-like healing?** *Autoimmun Rev* 2004, **3:**305–311.
9. Schiff M, Peura D: **HZT-501 (DUEXIS®; ibuprofen 800 mg/famotidine 26.6 mg) gastrointestinal protection in the treatment of the signs and symptoms of rheumatoid arthritis and osteoarthritis.** *Expert Rev Gastroenterol Hepatol* 2012, **6:**25–35.
10. McCormack PL: **Celecoxib: a review of its use for symptomatic relief in the treatment of osteoarthritis, rheumatoid arthritis and ankylosing spondylitis.** *Drugs* 2011, **71:**2457–2489.
11. Ravi B, Escott B, Shah PS, Jenkinson R, Chahal J, Bogoch E, Kreder H, Hawker G: **A systematic review and meta-analysis comparing complications following total joint arthroplasty for rheumatoid arthritis versus for osteoarthritis.** *Arthritis Rheum* 2012, **64:**3839–3849.
12. Beasley J: **Osteoarthritis and rheumatoid arthritis: conservative therapeutic management.** *J Hand Ther* 2012, **25:**163–171.
13. Hashizume K, Nishida K, Fujiwara K, Kadota Y, Nakahara R, Ezawa K, Inoue H, Ozaki T: **Radiographic measurements in the evaluation and classification of elbow joint destruction in patients with rheumatoid arthritis.** *Clin Rheumatol* 2010, **29:**637–643.
14. Krenn V, Morawietz L, Burmester GR, Kinne RW, Mueller-Ladner U, Muller B, Häupl T: **Synovitis score: discrimination between chronic low-grade and high-grade synovitis.** *Histopathology* 2006, **49:**358–364.
15. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger TA Jr, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp JT, Wilder RL, Hunder GG: **The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis.** *Arthritis Rheum* 1988, **31:**315–324.
16. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, Christy W, Cooke TD, Greenwald R, Hochberg M, Howell D, Kaplan D, Koopman W, Longley S III, Mankin H, McShane DJ, Medsger T Jr, Meenan R, Mikkelsen W, Moskowitz R, Murphy W, Rothschild B, Segal M, Sokoloff L, Wolfe F: **Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association.** *Arthritis Rheum* 1986, **29:**1039–1049.
17. Ross C: **A comparison of osteoarthritis and rheumatoid arthritis: diagnosis and treatment.** *Nurse Pract* 1997, **22:**20–28.
18. Kunkel GA, Cannon GW, Clegg DO: **Combined structural and synovial assessment for improved ultrasound discrimination of rheumatoid, osteoarthritic, and normal joints: a pilot study.** *Open Rheumatol J* 2012, **6:**199–206.
19. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, Combe B, Costenbader KH, Dougados M, Emery P, Ferraccioli G, Hazes JM, Hobbs K, Huizinga TW, Kavanaugh A, Kay J, Kvien TK, Laing T, Mease P, Ménard HA, Moreland LW, Naden RL, Pincus T, Smolen JS, Stanislawska-Biernat E, Symmons D, *et al*: **2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative.** *Ann Rheum Dis* 2010, **69:**1580–1588.
20. Kennish L, Labitigan M, Budoff S, Filopoulos MT, McCracken WA, Swearingen CJ, Yazici Y: **Utility of the new rheumatoid arthritis 2010 ACR/EULAR classification criteria in routine clinical care.** *BMJ Open* 2012, **2:**e001117.
21. van der Linden MP, Batstra MR, Bakker-Jonges LE, Foundation for Quality Medical Laboratory Diagnostics, Detert J, Bastian H, Scherer HU, Toes RE, Burmester GR, Mjaavatten MD, Kvien TK, Huizinga TW, van der Helm-van Mil AH: **Toward a data-driven evaluation of the 2010 American College of Rheumatology/European League Against Rheumatism criteria for rheumatoid arthritis: is it sensible to look at levels of rheumatoid factor?** *Arthritis Rheum* 2011, **63:**1190–1199.
22. van der Pouw Kraan TC, van Baarsen LG, Rustenburg F, Baltus B, Fero M, Verweij CL: **Gene expression profiling in rheumatology.** *Methods Mol Med* 2007, **136:**305–327.
23. Lübbers J, Brink M, van de Stadt LA, Vosslamber S, Wesseling JG, van Schaardenburg D, Rantapää-Dahlqvist S, Verweij CL: **The type I IFN signature as a biomarker of preclinical rheumatoid arthritis.** *Ann Rheum Dis* 2013, **72:**776–780.
24. Grcevic D, Jajic Z, Kovacic N, Lukic IK, Velagic V, Grubisic F, Ivcevic S, Marusic A: **Peripheral blood expression profiles of bone morphogenetic proteins, tumor necrosis factor-superfamily molecules, and transcription factor Runx2 could be used as markers of the form of arthritis, disease activity, and therapeutic responsiveness.** *J Rheumatol* 2010, **37:**246–256.
25. Mutlu N, Bicakcigil M, Tasan DA, Kaya A, Yavuz S, Ozden AI: **Comparative performance analysis of 4 different anti-citrullinated protein assays in the diagnosis of rheumatoid arthritis.** *J Rheumatol* 2009, **36:**491–500.
26. Kido A, Pap G, Kawate K, Roessner A, Takakura Y: **Disease-specific expression patterns of proteases in synovial tissues.** *Pathol Res Pract* 2007, **203:**451–456.
27. Bhattacharya S, Mariani TJ: **Array of hope: expression profiling identifies disease biomarkers and mechanism.** *Biochem Soc Trans* 2009, **37:**855–862.
28. van Baarsen LG, Bos CL, van der Pouw Kraan TC, Verweij CL: **Transcription profiling of rheumatic diseases.** *Arthritis Res Ther* 2009, **11:**207.
29. Lequerré T, Bansard C, Vittecoq O, Derambure C, Hiron M, Daveau M, Tron F, Ayral X, Biga N, Auquit-Auckbur I, Chiocchia G, Le Loët X, Salier JP: **Early and long-standing rheumatoid arthritis: distinct molecular signatures identified by gene-expression profiling in synovia.** *Arthritis Res Ther* 2009, **11:**R99.
30. Yi CQ, Ma CH, Xie ZP, Cao Y, Zhang GQ, Zhou XK, Liu ZQ: **Comparative genome-wide gene expression analysis of rheumatoid arthritis and osteoarthritis.** *Genet Mol Res* 2013, **12:**3136–3145.
31. Li G, Han N, Li Z, Lu Q: **Identification of transcription regulatory relationships in rheumatoid arthritis and osteoarthritis.** *Clin Rheumatol* 2013, **32:**609–615.
32. Raterman HG, Vosslamber S, de Ridder S, Nurmohamed MT, Lems WF, Boers M, van de Wiel M, Dijkmans BA, Verweij CL, Voskuyl AE: **The interferon type I signature towards prediction of non-response to rituximab in rheumatoid arthritis patients.** *Arthritis Res Ther* 2012, **14:**R95.
33. Stuhlmüller B, Häupl T, Hernandez MM, Grützkau A, Kuban RJ, Tandon N, Voss JW, Salfeld J, Kinne RW, Burmester GR: **CD11c as a transcriptional biomarker to predict response to anti-TNF monotherapy with adalimumab in patients with rheumatoid arthritis.** *Clin Pharmacol Ther* 2010, **87:**311–321.
34. Glocker MO, Guthke R, Kekow J, Thiesen HJ: **Rheumatoid arthritis, a complex multifactorial disease: on the way toward individualized medicine.** *Med Res Rev* 2006, **26:**63–87.
35. Sha N, Vannucci M, Brown PJ, Trower MK, Amphlett G, Falciani F: **Gene selection in arthritis classification with large-scale microarray expression profiles.** *Comp Funct Genomics* 2003, **4:**171–181.
36. Quinlan JR: **Induction of decision trees.** *Mach Learn* 1986, **1:**81–106.
37. Simon S, Guthke R, Kamradt T, Frey O: **Multivariate analysis of flow cytometric data using decision trees.** *Front Microbio* 2012, **3:**114.
38. Guthke R, Schmidt-Heck W, Pfaff M: **Knowledge acquisition and knowledge based control in bioprocess engineering.** *J Biotechnol* 1998, **65:**37–46.
39. Troschke SO: **Kennzahlen der regelbasierten Modellierung in Experten systemen. Ein Ansatz zur Bewertung von Unsicherheit bei der automatischen Erzeugung von Produktionsregeln**. In *Diploma thesis.* Chair of Electrical Control Engineering, University of Dortmund, Germany; 1992.
40. Krone A, Kiendl H: **Automatic generation of positive and negative rules for two-way fuzzy controllers.** In: *Proceedings of the Second European Congress on Intelligent Techniques and Soft Computing, EUFIT '94.* Aachen (Germany) 1994, 438–447.
41. Krabs M, Kiendl H: **Anwendungsfelder der automatischen Regelgenerierung mit dem ROSA Verfahren.** *Automatisierungstechnik* 1995, **43:**269–276.

42. Jessen H, Slawinski T: **Test and rating strategies for data based rule generation**. In *Computational Intelligence, Sonderforschungsbereich 531*, Paper No. CI-39/98. Dortmund: German National Library of Science and Technology (TIB), Hannover, Germany; 1998. http://hdl.handle.net/10068/240405.
43. Kiendl H, Krause P, Schauten D, Slawinski T: **Data-based fuzzy modeling for complex applications**. In *Advance in Computational Intelligence: Theory and Practice (Natural Computing Series)*. Edited by Schwefel H-P, Wegener I, Weinert KD. Springer: Heidelberg, Germany; 2003:46–77.
44. Huber R, Kunisch E, Glück B, Egerer R, Sickinger S, Kinne RW: **Comparison of conventional and real-time RT-PCR for the quantitation of jun protooncogene mRNA and analysis of junB mRNA expression in synovial membranes and isolated synovial fibroblasts from rheumatoid arthritis patients.** *Z Rheumatol* 2003, **62:**378–389.
45. Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C: **Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods.** *PLoS One* 2011, **6:**e17238.
46. Bezdek JC, Pal SK: *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. New York: IEEE Press; 1992.
47. Huang DW, Sherman BT, Tan Q, Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA: **DAVID Bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.** *Nucleic Acids Res* 2007, **35:**W169–W175.
48. Briken V, Ruffner H, Schultz U, Schwarz A, Reis LF, Strehlow I, Decker T, Staeheli P: **Interferon regulatory factor 1 is required for mouse Gbp gene activation by gamma interferon.** *Mol Cell Biol* 1995, **15:**975–982.
49. Ni Z, Karaskov E, Yu T, Callaghan SM, Der S, Park DS, Xu Z, Pattenden SG, Bremner R: **Apical role for BRG1 in cytokine-induced promoter assembly.** *Proc Natl Acad Sci USA* 2005, **102:**14611–14616.
50. Snyder M, He W, Zhang JJ: **The DNA replication factor MCM5 is essential for Stat1-mediated transcriptional activation.** *Proc Natl Acad Sci USA* 2005, **102:**14539–14544.
51. Zhao Y, Wagner F, Frank SJ, Kraft AS: **The amino-terminal portion of the JAK2 protein kinase is necessary for binding and phosphorylation of the granulocyte-macrophage colony-stimulating factor receptor beta c chain.** *J Biol Chem* 1995, **270:**13814–13818.
52. Rane SG, Reddy EP: **JAKs, STATs and Src kinases in hematopoiesis.** *Oncogene* 2002, **21:**3334–3358.
53. Reddy EP, Korapati A, Chaturvedi P, Rane S: **IL-3 signaling and the role of Src kinases, JAKs and STATs: a covert liaison unveiled.** *Oncogene* 2000, **19:**2532–2547.
54. Reichelt O, Müller J, von Eggeling F, Driesch D, Wunderlich H, Schubert J, Gröne HJ, Stein G, Ott U, Junker K: **Prediction of renal allograft rejection by urinary protein analysis using ProteinChip arrays (surface-enhanced laser desorption/ionization time-of-flight mass spectrometry).** *Urology* 2006, **67:**472–475.
55. Driesch D, Wötzel D, Guthke R, Pfaff M: **Fuzzy cluster and fuzzy rule cancer status prediction based on gene expression data.** In *Proceedings of the 4th International Workshop on Biosignal Interpretation.* Edited by Cerutti S. Como, Italy: Schattauer, Stuttgart, Germany; 2002:7–10.
56. Ruschpler P, Lorenz P, Eichler W, Koczan D, Hänel H, Scholz R, Melzer C, Thiesen HJ, Stiehl P: **High CXCR3 expression in synovial mast cells associated with CXCL9 and CXCL10 expression in inflammatory synovial tissues of patients with rheumatoid arthritis.** *Arthritis Res Ther* 2003, **5:**R241.
57. Biswas S, Manikandan J, Pushparaj PN: **Decoding the differential biomarkers of Rheumatoid arthritis and Osteoarthritis: a functional genomics paradigm to design disease specific therapeutics.** *Bioinformation* 2011, **6:**153–157.
58. Xue F, Zhang C, He Z, Ding L, Xiao H: **Analysis of critical molecules and signaling pathways in osteoarthritis and rheumatoid arthritis.** *Mol Med Rep* 2013, **7:**603–607.
59. Yoshida S, Arakawa F, Higuchi F, Ishibashi Y, Goto M, Sugita Y, Nomura Y, Niino D, Shimizu K, Aoki R, Hashikawa K, Kimura Y, Yasuda K, Tashiro K, Kuhara S, Nagata K, Ohshima K: **Gene expression analysis of rheumatoid arthritis synovial lining regions by cDNA microarray combined with laser microdissection: up-regulation of inflammation-associated STAT1, IRF1, CXCL9, CXCL10, and CCL5.** *Scand J Rheumatol* 2012, **41:**170–179.
60. van der Pouw Kraan TC, van Gaalen FA, Kasperkovitz PV, Verbeet NL, Smeets TJ, Kraan MC, Fero M, Tak PP, Huizinga TW, Pieterman E, Breedveld FC, Alizadeh AA, Verweij CL: **Rheumatoid arthritis is a heterogeneous disease: evidence for differences in the activation of the STAT-1 pathway between rheumatoid tissues.** *Arthritis Rheum* 2003, **48:**2132–2145.
61. Zhang W, Zhang J, Kornuc M, Kwan K, Frank R, Nimer SD: **Molecular cloning and characterization of NF-IL3A, a transcriptional activator of the human interleukin-3 promoter.** *Mol Cell Biol* 1995, **15:**6055–6063.
62. Monnier V, Iché-Torres M, Rera M, Contremoulins V, Guichard C, Lalevée N, Tricoire H, Perrin L: **dJun and Vri/dNFIL3 are major regulators of cardiac aging in *Drosophila*.** *PLoS Genet* 2012, **8:**e1003081.
63. Smith AM, Qualls JE, O'Brien K, Balouzian L, Johnson PF, Schultz-Cherry S, Smale ST, Murray PJ: **A distal enhancer in Il12b is the target of transcriptional repression by the STAT3 pathway and requires the basic leucine zipper (B-ZIP) protein NFIL3.** *J Biol Chem* 2011, **286:**23582–23590.
64. Kobayashi T, Matsuoka K, Sheikh SZ, Elloumi HZ, Kamada N, Hisamatsu T, Hansen JJ, Doty KR, Pope SD, Smale ST, Hibi T, Rothman PB, Kashiwada M, Plevy SE: **NFIL3 is a regulator of IL-12 p40 in macrophages and mucosal immunity.** *J Immunol* 2011, **186:**4649–4655.
65. Kashiwada M, Cassel SL, Colgan JD, Rothman PB: **NFIL3/E4BP4 controls type 2 T helper cell cytokine expression.** *EMBO J* 2011, **30:**2071–2082.
66. Kamizono S, Duncan GS, Seidel MG, Morimoto A, Hamada K, Grosveld G, Akashi K, Lind EF, Haight JP, Ohashi PS, Look AT, Mak TW: **Nfil3/E4bp4 is required for the development and maturation of NK cells in vivo.** *J Exp Med* 2009, **206:**2977–2986.
67. Cowell IG: **E4BP4/NFIL3, a PAR-related bZIP factor with many roles.** *Bioessays* 2002, **24:**1023–1029.
68. Shaulian E, Karin M: **AP-1 in cell proliferation and survival.** *Oncogene* 2001, **20:**2390–2400.
69. Kogut MH, Genovese KJ, He H, Kaiser P: **Flagellin and lipopolysaccharide up-regulation of IL-6 and CXCLi2 gene expression in chicken heterophils is mediated by ERK1/2-dependent activation of AP-1 and NF-kappaB signaling pathways.** *Innate Immun* 2008, **14:**213–222.
70. Weitzman JB, Fiette L, Matsuo K, Yaniv M: **JunD protects cells from p53-dependent senescence and apoptosis.** *Mol Cell* 2000, **6:**1109–1119.
71. Mineva ND, Rothstein TL, Meyers JA, Lerner A, Sonenshein GE: **CD40 ligand-mediated activation of the de novo RelB NF-kappaB synthesis pathway in transformed B cells promotes rescue from apoptosis.** *J Biol Chem* 2007, **282:**17475–17485.
72. Zerbini LF, de Vasconcellos JF, Czibere A, Wang Y, Paccez JD, Gu X, Zhou JR, Libermann TA: **JunD-mediated repression of GADD45α and γ regulates escape from cell death in prostate cancer.** *Cell Cycle* 2011, **10:**2583–2591.
73. Schutte J, Viallet J, Nau M, Segal S, Fedorko J, Minna J: **Jun-B inhibits and c-fos stimulates the transforming and trans-activating activities of c-jun.** *Cell* 1989, **59:**987–997.
74. Morita Y, Kashihara N, Yamamura M, Okamoto H, Harada S, Kawashima M, Makino H: **Antisense oligonucleotides targeting c-fos mRNA inhibit rheumatoid synovial fibroblast proliferation.** *Ann Rheum Dis* 1998, **57:**122–124.
75. White LA, Brinckerhoff CE: **Two activator protein-1 elements in the matrix metalloproteinase-1 promoter have different effects on transcription and bind Jun D, c-Fos, and Fra-2.** *Matrix Biol* 1995, **14:**715–725.
76. Castellazzi M, Spyrou G, La Vista N, Dangy JP, Piu F, Yaniv M, Brun G: **Overexpression of c-jun, junB, or junD affects cell growth differently.** *Proc Natl Acad Sci U S A* 1991, **88:**8890–8894.
77. Wakisaka S, Suzuki N, Saito N, Ochi T, Sakane T: **Possible correction of abnormal rheumatoid arthritis synovial cell function by jun D transfection in vitro.** *Arthritis Rheum* 1998, **41:**470–481.
78. Bakiri L, Matsuo K, Wisniewska M, Wagner EF, Yaniv M: **Promoter specificity and biological activity of tethered AP-1 dimers.** *Mol Cell Biol* 2002, **22:**4952–4964.
79. Cuevas BD, Uhlik MT, Garrington TP, Johnson GL: **MEKK1 regulates the AP-1 dimer repertoire via control of JunB transcription and Fra-2 protein stability.** *Oncogene* 2005, **24:**801–809.
80. Tomasi ML, Ryoo M, Skay A, Tomasi I, Giordano P, Mato JM, Lu SC: **Polyamine and methionine adenosyltransferase 2A crosstalk in human colon and liver cancer.** *Exp Cell Res* 2013, **319:**1902–1911.
81. Amé JC, Spenlehauer C, de Murcia G: **The PARP superfamily.** *Bioessays* 2004, **26:**882–893.
82. Ma Q, Baldwin KT, Renzelli AJ, McDaniel A, Dong L: **TCDDinducible poly (ADP-ribose) polymerase: a novel response to 2,3,7,8-tetrachlorodibenzo-p-dioxin.** *Biochem Biophys Res Commun* 2001, **289:**499–506.
83. Bailleul B, Akerblom I, Strosberg AD: **The leptin receptor promoter controls expression of a second distinct protein.** *Nucleic Acids Res* 1997, **25:**2752–2758.
84. Huang Y, Ying K, Xie Y, Zhou Z, Wang W, Tang R, Zhao W, Zhao S, Wu H, Gu S, Mao Y: **Cloning and characterization of a novel human leptin receptor overlapping transcript-like 1 gene (LEPROTL1).** *Biochim Biophys Acta* 2001, **1517:**327–331.

85. Belgareh-Touze N, Avaro S, Rouille Y, Hoflack B, Haguenauer-Tsapis R: **Yeast Vps55p, a functional homolog of human obesity receptor gene-related protein, is involved in late endosome to vacuole trafficking.** *Mol Biol Cell* 2002, **13:**1694–1708.
86. Mercer JG, Moar KM, Hoggard N, Strosberg AD, Froguel P, Bailleul B: **B219/OB-R 5′-UTR and leptin receptor gene-related protein gene expression in mouse brain and placenta: tissue-specific leptin receptor promoter activity.** *J Neuroendocrinol* 2000, **12:**649–655.
87. Touvier T, Conte-Auriol F, Briand O, Cudejko C, Paumelle R, Caron S, Baugé E, Rouillé Y, Salles JP, Staels B, Bailleul B: **LEPROT and LEPROTL1 cooperatively decrease hepatic growth hormone action in mice.** *J Clin Invest* 2009, **119:**3830–3838.
88. Couturier C, Sarkis C, Séron K, Belouzard S, Chen P, Lenain A, Corset L, Dam J, Vauthier V, Dubart A, Mallet J, Froguel P, Rouillé Y, Jockers R: **Silencing of OB-RGRP in mouse hypothalamic arcuate nucleus increases leptin receptor signaling and prevents diet-induced obesity.** *Proc Natl Acad Sci USA* 2007, **104:**19476–19481.
89. Weissmann G: **The mediation of rheumatoid inflammation by lysosomes.** *Adv Clin Pharmacol* 1974, **6:**51–63.
90. Bitensky L, Butcher RG, Johnstone JJ, Chayen J: **Effect of glucocorticoids on lysosomes in synovial lining cells in human rheumatoid arthritis.** *Ann Rheum Dis* 1974, **33:**57–61.
91. Lockwood TD: **The lysosome among targets of metformin: new anti-inflammatory uses for an old drug?** *Expert Opin Ther Targets* 2010, **14:**467–478.
92. Wollbold J, Huber R, Pohlers D, Koczan D, Guthke R, Kinne RW, Gausmann U: **Adapted Boolean network models for extracellular matrix formation.** *BMC Syst Biol* 2009, **3:**77.
93. Ivashkiv LB, Hu X: **The JAK/STAT pathway in rheumatoid arthritis: pathogenic or protective?** *Arthritis Rheum* 2003, **48:**2092–2096.
94. Okamoto H, Cujec TP, Yamanaka H, Kamatani N: **Molecular aspects of rheumatoid arthritis: role of transcription factors.** *FEBS J* 2008, **275:**4463–4470.
95. Devauchelle V, Marion S, Cagnard N, Mistou S, Falgarone G, Breban M, Letourneur F, Pitaval A, Alibert O, Lucchesi C, Anract P, Hamadouche M, Ayral X, Dougados M, Gidrol X, Fournier C, Chiocchia G: **DNA microarray allows molecular profiling of rheumatoid arthritis and identification of pathophysiological targets.** *Genes Immun* 2004, **5:**597–608.
96. Kasperkovitz PV, Timmer TC, Smeets TJ, Verbeet NL, Tak PP, van Baarsen LG, Baltus B, Huizinga TW, Pieterman E, Fero M, Firestein GS, van der Pouw Kraan TC, Verweij CL: **Fibroblast-like synoviocytes derived from patients with rheumatoid arthritis show the imprint of synovial tissue heterogeneity: evidence of a link between an increased myofibroblast-like phenotype and high-inflammation synovitis.** *Arthritis Rheum* 2005, **52:**430–441.
97. Früh K, Yang Y: **Antigen presentation by MHC class I and its regulation by interferon gamma.** *Curr Opin Immunol* 1999, **11:**76–81.
98. Ebstein F, Kloetzel PM, Krüger E, Seifert U: **Emerging roles of immunoproteasomes beyond MHC class I antigen processing.** *Cell Mol Life Sci* 2012, **69:**2543–2558.
99. Yu L, Li Q, Lin J, Yu J, Li Q, Yi W, Sun H, Chu JY, Yang ZQ: **Association between polymorphisms of PSMB8, PSMB9 and TAP2 genes with rheumatoid arthritis in ethnic Han Chinese from Yunnan.** *Zhonghua Yi Xue Yi Chuan Xue Za Zhi (Chin Med Genet)* 2013, **30:**222–226.
100. Scheffler S, Kuckelkorn U, Egerer K, Dörner T, Reiter K, Soza A, Burmester GR, Feist E: **Autoimmune reactivity against the 20S-proteasome includes immunosubunits LMP2 (beta1i), MECL1 (beta2i) and LMP7 (beta5i).** *Rheumatology* 2008, **47:**622–626.
101. Yu P, Constien R, Dear N, Katan M, Hanke P, Bunney TD, Kunder S, Quintanilla-Martinez L, Huffstadt U, Schröder A, Jones NP, Peters T, Fuchs H, de Angelis MH, Nehls M, Grosse J, Wabnitz P, Meyer TP, Yasuda K, Schiemann M, Schneider-Fresenius C, Jagla W, Russ A, Popp A, Josephs M, Marquardt A, Laufs J, Schmittwolf C, Wagner H, Pfeffer K *et al*: **Autoimmunity and inflammation due to a gain-of-function mutation in phospholipase C gamma 2 that specifically increases external Ca2+ entry.** *Immunity* 2005, **22:**451–465.
102. Abe K, Fuchs H, Boersma A, Hans W, Yu P, Kalaydjiev S, Klaften M, Adler T, Calzada-Wack J, Mossbrugger I, Rathkolb B, Rozman J, Prehn C, Maraslioglu M, Kametani Y, Shimada S, Adamski J, Busch DH, Esposito I, Klingenspor M, Wolf E, Wurst W, Gailus-Durner V, Katan M, Marschall S, Soewarto D, Wagner S, de Angelis MH: **A novel N-ethyl-N-nitrosourea-induced mutation in phospholipase Cγ2 causes inflammatory arthritis, metabolic defects, and male infertility in vitro in a murine model.** *Arthritis Rheum* 2011, **63:**1301–1311.
103. Marco de Leon J: *Gene Expression Profiling of Multiple Autoimmune Diseases*, University of Minnesota Epidemiology Microform. Ann Arbor, MI: ProQuest LLC; 2008.
104. Array express: *Transcription profiling of human synovial samples from patients with osteoarthritis, rheumatoid arthritis vs controls treated with various drug regimes to characterise RA at the molecular level and to uncover key pathomechanisms*. Hinxton, Cambridge CB10 1SD, United Kingdom: The EMBL-European Bioinformatics Institute Wellcome Trust Genome Campus. [http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-1919/]
105. Kato M, Neil TK, Clark GJ, Morris CM, Sorg RV, Hart DN: **cDNA cloning of human DEC-205, a putative antigen-uptake receptor on dendritic cells.** *Immunogenetics* 1998, **47:**442–450.
106. McKay PF, Imami N, Johns M, Taylor-Fishwick DA, Sedibane LM, Totty NF, Hsuan JJ, Palmer DB, George AJ, Foxwell BM, Ritter MA: **The gp200-MR6 molecule which is functionally associated with the IL-4 receptor modulates B cell phenotype and is a novel member of the human macrophage mannose receptor family.** *Eur J Immunol* 1998, **28:**4071–4083.
107. Array express: *E-GEOD-38351 - The multifaceted balance of TNF-a and type I / II interferon responses in SLE and RA: how monocytes manage the impact of cytokines.*. Hinxton, Cambridge CB10 1SD, United Kingdom: The EMBL-European Bioinformatics Institute Wellcome Trust Genome Campus. [http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-38351/]
108. Greenawalt DM, Sieberts SK, Cornelis MC, Girman CJ, Zhong H, Yang X, Guinney J, Qi L, Hu FB: **Integrating genetic association, genetics of gene expression, and single nucleotide polymorphism set analysis to identify susceptibility Loci for type 2 diabetes mellitus.** *Am J Epidemiol* 2012, **176:**423–430.
109. Wang X, Lupardus P, Laporte SL, Garcia KC: **Structural biology of shared cytokine receptors.** *Annu Rev Immunol* 2009, **27:**29–60.
110. Hansen G, Hercus TR, McClure BJ, Stomski FC, Dottore M, Powell J, Ramshaw H, Woodcock JM, Xu Y, Guthridge M, McKinstry WJ, Lopez AF, Parker MW: **The structure of the GM-CSF receptor complex reveals a distinct mode of cytokine receptor activation.** *Cell* 2008, **134:**496–507.
111. Alvaro-Gracia JM, Zvaifler NJ, Firestein GS: **Cytokines in chronic inflammatory arthritis. V. Mutual antagonism between interferon-gamma and tumor necrosis factor-alpha on HLA-DR expression, proliferation, collagenase production, and granulocyte macrophage colony-stimulating factor production by rheumatoid arthritis synoviocytes.** *J Clin Invest* 1990, **86:**1790–1798.
112. Alvaro-Gracia JM, Zvaifler NJ, Firestein GS: **Cytokines in chronic inflammatory arthritis. IV. Granulocyte/macrophage colony-stimulating factor-mediated induction of class II MHC antigen on human monocytes: a possible role in rheumatoid arthritis.** *J Exp Med* 1989, **170:**865–875.
113. Xu WD, Firestein GS, Taetle R, Kaushansky K, Zvaifler NJ: **Cytokines in chronic inflammatory arthritis. II. Granulocyte–macrophage colony-stimulating factor in rheumatoid synovial effusions.** *J Clin Invest* 1989, **83:**876–882.
114. Wang Y, Thomson CA, Allan LL, Jackson LM, Olson M, Hercus TR, Nero TL, Turner A, Parker MW, Lopez AL, Waddell TK, Anderson GP, Hamilton JA, Schrader JW: **Characterization of pathogenic human monoclonal autoantibodies against GM-CSF.** *Proc Natl Acad Sci U S A* 2013, **110:**7832–7837.
115. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A: **Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases.** *Genes Immun* 2013, **14:**67–82.
116. Tenti S, Correale P, Conca R, Pastina P, Fioravanti A: **Occurrence of Sjögren syndrome in a long-term survivor patient with metastatic colon carcinoma treated with GOLFIG regimen.** *J Chemother* 2012, **24:**245–246.
117. Cook AD, Pobjoy J, Steidl S, Dürr M, Braine EL, Turner AL, Lacey DC, Hamilton JA: **Granulocyte-macrophage colony-stimulating factor is a key mediator in experimental osteoarthritis pain and disease development.** *Arthritis Res Ther* 2012, **14:**R199.
118. Zhang W, Cong XL, Qin YH, He ZW, He DY, Dai SM: **IL-18 upregulates the production of key regulators of osteoclastogenesis from fibroblast-like synoviocytes in rheumatoid arthritis.** *Inflammation* 2013, **36:**103–109.
119. Hughes-Austin JM, Deane KD, Derber LA, Kolfenbach JR, Zerbe GO, Sokolove J, Lahey LJ, Weisman MH, Buckner JH, Mikuls TR, O'Dell JR, Keating RM, Gregersen PK, Robinson WH, Holers VM, Norris JM: **Multiple cytokines and chemokines**

are associated with rheumatoid arthritis-related autoimmunity in first-degree relatives without rheumatoid arthritis: Studies of the Aetiology of Rheumatoid Arthritis (SERA). *Ann Rheum Dis* 2013, **72:**901–907.
120. Minter RR, Cohen ES, Wang B, Liang M, Vainshtein I, Rees G, Eghobamien L, Harrison P, Sims DA, Matthews C, Wilkinson T, Monk P, Drinkwater C, Fabri L, Nash A, McCourt M, Jermutus L, Roskos L, Anderson IK, Sleeman MA: **Protein engineering and preclinical development of a GM-CSF receptor antibody for the treatment of rheumatoid arthritis.** *Br J Pharmacol* 2013, **168:**200–211.
121. Nair JR, Edwards SW, Moots RJ: **Mavrilimumab, a human monoclonal GM-CSF receptor-α antibody for the management of rheumatoid arthritis: a novel approach to therapy.** *Expert Opin Biol Ther* 2012, **12:**1661–1668.
122. Burmester GR, Weinblatt ME, McInnes IB, Porter D, Barbarash O, Vatutin M, Szombati I, Esfandiari E, Sleeman MA, Kane CD, Cavet G, Wang B, Godwood A, Magrini F, EARTH Study Group: **Efficacy and safety of mavrilimumab in subjects with rheumatoid arthritis.** *Ann Rheum Dis* 2013, **72:**1445-1452.
123. Kuramochi S, Moriguchi T, Kuida K, Endo J, Semba K, Nishida E, Karasuyama H: **LOK is a novel mouse STE20-like protein kinase that is expressed predominantly in lymphocytes.** *J Biol Chem* 1997, **272:**22679–22684.
124. Tao L, Wadsworth S, Mercer J, Mueller C, Lynn K, Siekierka J, August A: **Opposing roles of serine/threonine kinases MEKK1 and LOK in regulating the CD28 responsive element in T-cells.** *Biochem J* 2002, **363:**175–182.
125. Belkina NV, Liu Y, Hao JJ, Karasuyama H, Shaw S: **LOK is a major ERM kinase in resting lymphocytes and regulates cytoskeletal rearrangement through ERM phosphorylation.** *Proc Natl Acad Sci U S A* 2009, **106:**4707–4712.
126. Fukumura K, Yamashita Y, Kawazu M, Sai E, Fujiwara S, Nakamura N, Takeuchi K, Ando M, Miyazono K, Ueno T, Ozawa K, Mano H: **STK10 missense mutations associated with anti-apoptotic function.** *Oncol Rep* 2013, **30:**1542–1548.
127. Korb A, Pavenstädt H, Pap T: **Cell death in rheumatoid arthritis.** *Apoptosis* 2009, **14:**447–454.
128. Yamamoto N, Honma M, Suzuki H: **Off-target serine/threonine kinase 10 inhibition by erlotinib enhances lymphocytic activity leading to severe skin disorders.** *Mol Pharmacol* 2011, **80:**466–475.
129. Seavey MM, Dobrzanski P: **The many faces of Janus kinase.** *Biochem Pharmacol* 2012, **83:**1136–1145.
130. Malemud CJ: **Differential activation of JAK enzymes in rheumatoid arthritis and autoimmune disorders by pro-inflammatory cytokines: potential drug targets.** *Int J Infereron Cytokine Mediator Res* 2010, **2:**97–111.
131. O'Shea JJ, Plenge R: **JAK and STAT signaling molecules in immunoregulation and immune-mediated disease.** *Immunity* 2012, **36:**542–550.

doi:10.1186/ar4526
Cite this article as: Woetzel *et al.*: Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. *Arthritis Research & Therapy* 2014 **16**:R84.