Biomarkers, 2010; 15(2): 183–190

informa
healthcare

ORIGINAL ARTICLE

Identification of broadly discriminatory tissue biomarkers of synovitis with binary and multicategory receiver operating characteristic analysis

A. Ogdie[1], J. Li[2], L. Dai[3,7], M. E. Paessler[5], X. Yu[6,7], C. Diaz-Torne[4,7], M. Akmatov[9], H. R. Schumacher[1,7], and F. Pessler[8,9]

[1]Division of Rheumatology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA, [2]Department of Statistics and Applied Probability, National University of Singapore, Singapore, [3]Second Affiliated Hospital, Sun Yat-sen University, Guangzhou, China, [4]Hospital de la Santa Creu I Sant Pau, Barcelona, Spain, [5]Department of Pathology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA, [6]Traditional Chinese Medicine-Western Medicine Hospital, Cangzhou, Hebei, China, [7]Division of Rheumatology Philadelphia VA Medical Center, PA, USA, [8]Klinik und Poliklinik für Kinder- und Jugendmedizin, Technische Universität Dresden, Germany, and [9]Helmholtz Centre for Infection Research, Braunschweig, Germany

Abstract

Immunohistochemical synovial tissue biomarkers are used increasingly to classify arthropathies, study their pathogenesis, and to measure disease activity in clinical trials. We have used receiver operating characteristic (ROC) analysis to quantify the discriminatory abilities of markers for common inflammatory cells (subintimal CD15, CD68, CD3, CD20, CD38, and lining CD68), proliferating cells (Ki-67) and blood vessels (von Willebrand factor, vWF) among inflammatory (chronic septic arthritis, early arthritis and rheumatoid arthritis (RA)) and degenerative arthropathies (osteoarthritis (OA) and orthopedic arthropathies) and normal synovium. Six of the eight markers distinguished accurately between RA and the degenerative arthropathies (area under the curve (AUC) 0.91–0.97), whereas subintimal CD68 (AUC 0.92) and Ki-67 (AUC 0.87) distinguished best between OA and normal synovium. Fold differences in mean expression correlated only modestly with AUCs ($r^2 = 0.44$). Multicategory ROC analysis ranked Ki-67, subintimal CD68, and CD15 as discriminating best among all six sample groups, and thus identified them as the most broadly applicable markers.

Keywords: *Computational biology; gene expression; growth factors/cytokines/inflammatory mediators*

Introduction

Traditionally, the difference in the expression of a synovial tissue marker between two specimen cohorts is assessed by determining the ratio of mean or median expression values between the two cohorts, combined with a statistical probability, *p* (e.g. 'marker A is x-fold more highly expressed in cohort 1 than in cohort 2'). However, depending on the relative fold differences in mean expression and the spread of the data, this method may not be ideal. For instance, in spite of a large difference between the mean expression values of two arthropathies, there may be a high degree of overlap in the individual values between the groups. Likewise, smaller differences between two means may be significant if there is little overlap between the data of each group. Receiver operating characteristic (ROC) curve analysis may be a valuable alternate method to assess synovial tissue markers because it is based on the degree of overlap in values between two cohorts of

Address for Correspondence: Frank Pessler, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany. Tel.: +49 531 6181-1112. Fax: +49 351 6181-1199. E-mail: frank.pessler@helmholtz-hzi.de

(Received 24 August 2009; revised 13 October 2009; accepted 13 October 2009)

ISSN 1354-750X print/ISSN 1366-5804 online © 2010 Informa UK Ltd
DOI: 10.3109/13547500903411095

specimens. Moreover, it expresses the discriminatory ability of the marker in question as a single value that falls into a defined standardized range between 0.5 and 1.0 and can thus be compared with results from other tests or markers.

ROC analysis originates from the field of signal theory and was originally used to study the performance of radar operators. It has since been applied in medicine to assess the accuracy of diagnostic tests (Zweig 1988, Zweig & Campbell 1993, Zweig et al. 1992, Faraggi & Reiser 2002). First, sensitivities and specificities are calculated from a set of test results which were derived in the attempt to distinguish between two outcomes (diagnoses). The ROC curve is then generated by plotting sensitivity (true positive rate) on the *y*-axis and 1-specificity (false positive rate) along the *x*-axis. The area under this curve (AUC) is a direct measure of the discriminatory value of the test, with a perfect test having an AUC of 1, a test with an AUC of 0.5 having no discriminatory value, and a test with an AUC <0.5 being negatively associated with the outcome in question. ROC analysis has traditionally been used to evaluate the ability of one test to distinguish between two outcomes (binary ROC analysis). However, a recently reported extension of this method, termed ROC analysis with multiple classes and multiple tests ('multicategory ROC analysis') (Li & Fine 2008, Li & Zhou 2009), makes it possible to evaluate the discriminatory ability of a test to distinguish among multiple outcomes (diagnoses). In this method, the hypervolume under the ROC manifold (HUM) is analogous to the AUC in conventional ROC analysis and summarizes the accuracy of a test in the simultaneous discrimination among multiple diagnoses. The HUM of a test in the simultaneous differentiation among K classes corresponds to acceptance of the null hypothesis when HUM = ≤1/K!. A higher HUM thus designates a more accurate test.

We have recently used conventional (binary) ROC analysis to address specific questions of synovial biomarker expression (Pessler et al. 2008c, d). In a related article, we used binary and multicategory ROC analysis to define the diagnostic accuracy of a three-component synovitis score and its components (E. Slansky et al., submitted for publication). Using both binary and multicategory ROC analysis, we have now evaluated eight immunohistochemical synovial biomarkers in their abilities to differentiate among several inflammatory and non-inflammatory (degenerative) arthropathies and normal synovium. We have also compared results of ROC analysis with fold differences in mean expression of the same markers among the same arthropathies and do not find a strong correlation between the two methods, suggesting that they reveal substantially different aspects of synovial biomarker expression.

Methods

Synovial tissue specimens were obtained by closed needle biopsy or surgically at the time of arthroplasty. The following specimen cohorts were studied: non-inflamed control specimens ($n = 22$), consisting of biopsies from healthy volunteers ($n = 11$) and patients with non-inflammatory knee pain ($n = 10$, two biopsies being available from one patient), which were indistinguishable from the normal specimens by a variety of histological and immunohistochemical markers (Diaz-Torne et al. 2007); rheumatoid arthritis (RA) with active disease despite disease-modifying antirheumatic drug (DMARD) treatment ($n = 28$, 21 needle biopsies and seven surgical specimens); early undifferentiated arthritis (duration <12 months, $n = 10$, all needle biopsies); chronic (disease duration >4 weeks) septic arthritis (SeA) proven by positive bacterial culture ($n = 11$, consisting of the specimens used in a previous study (Pessler et al. 2008c) plus two new specimens, all surgically and arthroscopically obtained specimens); non-inflammatory orthopedic arthropathies (Orth.A, $n = 23$, consisting of femur fracture, $n = 3$; avascular necrosis of the femur, $n = 3$; meniscus and/or ligament injury, $n = 13$; and plica syndrome, n = 4, all surgically and arthroscopically obtained specimens) (Pessler et al. 2008b); and osteoarthritis (OA, $n = 31$, 18 needle biopsies and 13 arthroscopically and surgically obtained specimens). General antigen preservation was tested by staining for von Willebrand Factor (vWF). Tissues were fixed in formalin and embedded in paraffin according to standard practice. Sections (5 µm thick) were immunostained on a semiautomated staining system (Ventana Benchmark; Vantana, Tucson, AZ, USA) for CD3 (T cells, antibody clone PS1), CD20 (B cells, L-26), CD15 (neutrophilic granulocytes, MMA), CD38 (plasma cells, SPC32), CD68 (macrophages, Kp1), Ki-67 (proliferating cells, K-2) (Pessler et al. 2008d), and vWF (vascular endothelium, polyclonal rabbit IgG). All antibodies were prediluted preparations from Ventana, except the anti-CD38 IgG, which was purchased from Novocastra (Newcastle Upon Tyne, UK). The numbers of positive staining cells or vessels per high-power field (hpf, 400x, corresponding to 0.159 mm^2) were determined to a depth of 1 hpf (diameter 0.454 mm) in five to 15 fields per specimen by manual counting as described previously (Diaz-Torne et al. 2007, Pessler et al. 2008a). Only fields with clearly recognizable lining and subintimal vasculature were included. The density of CD68+ lining cells was defined as the number of CD68+ cells in the surface layer of the intima per 400x field (diameter 0.454 mm) with the specimen oriented such that the lining traversed the maximum diameter (centre) of the microscopic field. Due to frequent loss of lining tissue

in SeA, CD68+ lining cells were not quantified in these specimens.

SPSS biostatistical software version 15.0 (SPSS Inc., Chicago, IL, USA) was used for conventional ROC analysis. Paired comparisons were set up such that the presumably more inflamed arthropathy was the state variable. This corresponded to the following hierarchy of presumed inflammation: SeA>RA>early arthritis>OA>Orth.A>normal tissue. For example, a marker that would be positively associated with a diagnosis of RA in the comparison RA versus OA would have an AUC of >0.5. The trade-off value was defined as the expression density along the ROC curve at which the sum of sensitivity plus specificity was maximal. The Youden index corresponds to:

$$(\text{sensitivity[fraction]} + \text{specificity[fraction]}) - 1$$

at the trade-off value, with an index equal to 1 identifying a test with optimal accuracy (Youden 1950). Fold differences in mean expression (ratios of means) were calculated by dividing the mean expression density of a marker in one arthropathy by its mean expression density in a second arthropathy, with statistical significance determined with the Mann–Whitney *U* test. Multicategory ROC analysis was performed using code written in the R software environment for biostatistical computing (http://www.r-project.org/) as described (Li & Fine 2008).

Results

Quantitative expression of the markers

Figure 1 shows expression densities of all markers in the specimen cohorts. Overall, densities were highest in SeA, RA and early arthritis, intermediate or near normal in OA and Orth.A, and lowest in the normal controls. As expected, the RA specimens were rich in all inflammatory cell types including cells of humoral immunity (CD20 and CD38, Figure 1B), whereas the SeA cohort contained the highest mean density of CD15+ cells (Figure 1C) (Pessler et al. 2008c). Although densities of all markers were somewhat lower in early arthritis than in RA, none of these differences were statistically significant. As reported previously (Pessler et al. 2008a, b), the normal specimens contained low densities of CD68+ and CD3+ cells (Figure 1A), but practically no evidence of CD15+ cells or humoral immunity, while

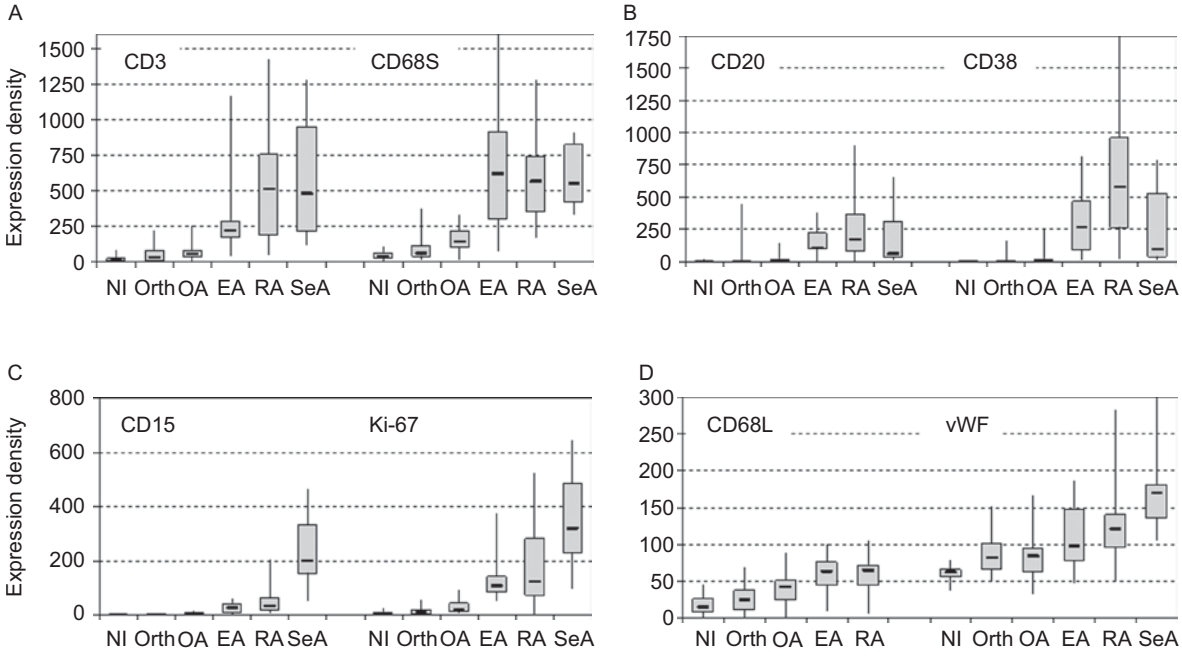

Figure 1. Expression profiles of immunohistochemical synovial biomarkers. Expression densities were determined by manual counting under 400x magnification and expressed as positive staining subintimal cells mm^{-2} (CD15, CD3, CD20, CD38, CD68S and Ki-67), positive staining subintimal vessels mm^{-2} (vWF), or positive staining cells mm^{-1} in the surface layer of the intima (CD68L). Boxes represent the 25th to 75th percentiles and horizontal lines the median; upper and lower whiskers represent the maximal and minimal values, respectively. Note that y-axes have different scales in the four panels. (A) CD3 and CD68S. (B) CD20 and CD38. (C) CD15 and Ki-67. (D) CD68L and vWF. CD68L, CD68 in the surface layer of the lining; CD68S, CD68 in the subintima; Nl, histologically normal specimens from healthy volunteers and individuals with non-inflammatory knee pain (n=22); Orth, non-inflammatory orthopaedic arthropathies (n=23); OA, osteoarthritis (n=31); EA, early undifferentiated arthritis of <12 months duration (n=10); RA, rheumatoid arthritis (n=28); SeA, chronic septic arthritis with symptoms >4 weeks (n=11); vWF, von Willebrand factor.

the Orth.A cohort exhibited a mild degree of synovitis, intermediate between the normal controls and OA. Consistent with our previous results (Pessler et al. 2008c), a gradual increase of vWF expression (vascular density; Figure 1D) and a logarithmic increase of Ki-67+ proliferating cells (Figure 1C) were noted from the normal specimens over OA and RA to SeA. There also was an enrichment of CD68+ cells in the surface layer of the lining in the inflammatory arthropathies and, to a lesser extent, in OA (Figure 1D). For each marker, the values shown in Figure 1 were then used to calculate the fold differences in mean expression for selected pairs of specimen groups (Table S1, see online version of this article). These reached maximum values of >1000:1 when CD38 expression was compared between the inflamed arthropathies and normal controls, which was partially due to the extremely low expression of this marker in normal synovium (Pessler et al. 2008a).

Trade-off values and Youden index

The trade-off value is the value possessing the optimal compromise between sensitivity and specificity for the distinction between two diagnostic outcomes. The Youden index expresses the diagnostic accuracy of a test at this trade-off value in a single value between 0 and 1. For instance, comparing RA and OA using subintimal CD68, a trade-off value of 345.5 positive staining cells mm^{-2} possessed a sensitivity of 75% and specificity of 100% for the diagnosis of RA, yielding a Youden index of 0.75. Overall, values for the Youden index ranged from 0.14 to 1.0 (mean 0.64, median 0.63, SD 0.28), with the optimal value of 1 resulting mostly from comparisons between the inflamed arthropathies and normal tissue, but also in the comparison SeA vs OA when CD15, CD20 or Ki-67 were used. The lowest values (<0.25) were observed in comparisons among the inflamed arthropathies, particularly SeA or RA vs early arthritis, but also between Orth.A and OA or normal tissue (Table S1, see online version of this article).

Conventional (binary) ROC analysis: general findings

Figure 2 illustrates ROC curves for subintimal CD68 (a well-established synovial tissue marker of RA (Haringman et al. 2005, Kraan et al. 1999)) in selected comparisons in which it possessed perfect (AUC = 1), high (AUC = 0.92) or no (AUC = 0.55) discriminatory ability. AUCs for all paired comparisons are summarized in Table 1, and the corresponding confidence intervals and *p*-values in the supplemental table (Table S1, see online version of this aticle). AUCs ranged from 0.28 to 1.0 (mean 0.80, median 0.83, SD 0.19) and had a significant discriminatory power in 64 of the 92 (70%) comparisons, as defined by the 95% confidence interval not crossing over 0.5 and a *p*-value of <0.05. Values for Youden index and AUC correlated strongly ($r^2 = 0.89$). All markers accurately differentiated the normal controls from the inflammatory arthropathies, i.e. early arthritis (AUC range 0.78–1.0, Youden index range 0.72–1.0), RA (AUC range 0.93–1.0, Youden index range 0.84–1.0) and SeA (AUC range 0.99–1.0, Youden index 1.0). Likewise, most markers were able to distinguish the inflammatory arthropathies well from the less inflamed ones (i.e. orthopedic arthropathies and osteoarthritis), but had little or no discriminatory ability when comparing pairs within the highly inflamed (SeA, RA and early arthritis) or less inflamed (OA and Orth.A) groups. CD15, Ki-67 and vWF were notable exceptions in that they had good discriminatory abilities for the comparisons between SeA and RA or early arthritis.

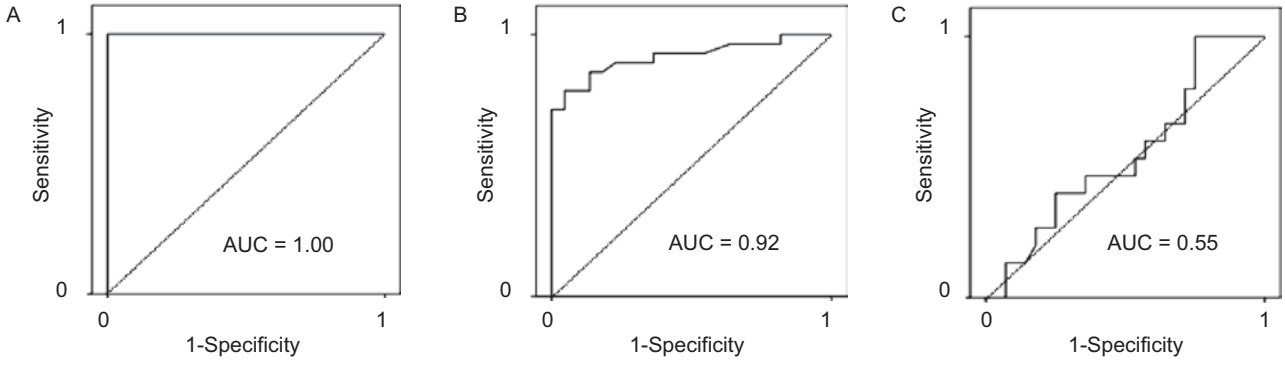

Figure 2. Receiver operating characteristic (ROC) curves illustrating the greatly variable diagnostic ability of subintimal CD68 (CD68S) in three different scenarios. ROC curves were generated with SPSS 15.0 biostatistical software. (A) CD68S in the differentiation of rheumatoid arthritis (RA) vs normal synovium, illustrating a marker with perfect discriminatory ability; area under the curve (AUC) = 1 (95% confidence interval (CI) 1.00–1.00; $p < 0.001$). (B) CD68S, osteoarthritis (OA) vs normal tissue, a marker of high accuracy; AUC = 0.92 (95% CI 0.84–1.00; $p < 0.001$). (C) CD68S, septic arthritis (SeA) vs RA, a marker with no discriminatory ability for this comparison; AUC = 0.55 (95% CI 0.33–0.76; $p = 0.58$).

Table 1. Summary of areas under the ROC curves (AUCs).[a]

	AUCs							
Comparison	CD15	CD20	CD38	CD3	CD68S	CD68L	Ki-67	vWF
Orth.A vs normal	0.71	0.51	0.78[b]	0.65	0.68[b]	0.61	0.60	0.81[b]
OA vs normal	0.79[b]	0.64	0.78[b]	0.75[b]	0.92[b]	0.79[b]	0.84[b]	0.76[b]
OA vs Orth.A	0.62	0.60	0.56	0.57	0.75[b]	0.71[b]	0.75[b]	0.49
Early vs normal	0.99[b]	0.78[b]	1.00[b]	0.98[b]	0.99[b]	0.89[b]	1.00[b]	0.89[b]
Early vs OA	0.88[b]	0.74[b]	0.91[b]	0.90[b]	0.84[b]	0.67	0.95[b]	0.66
RA vs normal	1.00[b]	0.97[b]	1.00[b]	0.99[b]	1.00[b]	0.94[b]	1.00[b]	0.93[b]
RA vs OA	0.96[b]	0.91[b]	0.97[b]	0.95[b]	0.96[b]	0.73[b]	0.92[b]	0.76[b]
RA vs early	0.63	0.64	0.68	0.69[b]	0.48	0.54	0.57	0.55
SeA vs normal	1.00[b]	0.99[b]	1.00	1.00[b]	1.00[b]	n/d	1.00[b]	1.00[b]
SeA vs OA	1.00[b]	0.87[b]	0.89[b]	0.97[b]	1.00[b]	n/d	1.00[b]	0.95[b]
SeA vs early	0.97[b]	0.56	0.45	0.71	0.54	n/d	0.90[b]	0.83[b]
SeA vs RA	0.92[b]	0.38	0.28	0.53	0.56	n/d	0.77[b]	0.84[b]

[a]Confidence intervals (CI), trade-off values, sensitivities, specificities, Youden indices and ratios of mean expression are listed in the Table S1 (see online version of this article). [b]AUC diagnostically significant in that the 95% CI does not cross 0.5.
Early, early undifferentiated arthritis (duration <12 months); n/d, not determined; OA, osteoarthritis; Orth.A, orthopaedic arthropathies; RA, rheumatoid arthritis; ROC, receiver operating characteristic; SeA, chronic septic arthritis (>4 weeks duration).

Six markers differentiated RA from OA with high AUCs ranging from 0.91 to 0.97. Likewise, most markers easily differentiated early arthritis from normal tissue or OA. On the other hand, the highest AUC in the comparison RA versus early arthritis was only 0.69 (marker CD3, Youden index 0.5), while the remaining AUCs ranged between 0.48 and 0.68 (Youden index range 0.2–0.38) and had lower confidence intervals crossing below the 0.5 line. Underscoring the similarities between OA and Orth.A, the best markers for the distinction between these two cohorts, CD68 and Ki-67, had AUCs of only 0.75 (Youden index 0.51 and 0.35), while the remaining AUCs ranged between 0.49 and 0.71, had confidence intervals crossing <0.5, and were thus non-discriminatory. Most markers displayed some discriminatory power separating the less inflamed arthropathies from normal tissue. For instance, all eight markers had statistically significant AUCs >0.5 for the discrimination between OA and normal controls, with subintimal CD68 possessing by far the highest value (AUC 0.92, Youden index 0.74). On the other hand, only vWF, CD38 and subintimal CD68 could discriminate between Orth.A and normal controls, with vWF having the highest AUC (0.81, Youden index 0.56).

Discriminatory abilities of specific markers

CD15 appeared to have the best overall discriminatory ability in that it possessed AUCs of ≥0.90 for seven of the ten paired comparisons for which results with all eight markers (except lining CD68, which could not be determined in the SeA specimens) were available. Consistent with the high frequency of neutrophilic granulocytes in chronically infected joints (Pessler et al. 2008c), CD15 was the best marker for the distinction between SeA and RA or early arthritis, possessing AUCs of 0.92 and 0.97, respectively, and Youden indices of 0.76 and 0.87. It also had high AUCs for the comparisons between RA and the normal or OA cohorts but could not discriminate between early arthritis and RA. Subintimal CD68, an established synovial tissue marker of RA (Haringman et al. 2005, Kraan et al. 1999), discriminated well between RA and OA (AUC 0.96, Youden index 0.75), between OA and normal tissue (for which it had the highest AUC of all the markers), and, to a lesser extent, also between early arthritis and OA. In contrast, AUCs close to the non-discriminatory value of 0.5 were observed for comparisons among the inflammatory arthropathies, i.e. RA, early arthritis and SeA. CD68+ cells in the lining surface (CD68L) separated RA and early arthritis well from normal tissue, and also had a relatively good, statistically significant discriminatory ability between OA and normal tissue (AUC 0.79, Youden index 0.44). CD3 was the only marker that could distinguish between early arthritis and RA, albeit with a low AUC of 0.69. As suggested pathogenetically, it was highly expressed in RA and early arthritis and differentiated these accurately from OA and normal controls. The results of the markers for humoral immunity, CD20 and CD38, resembled each other, although CD38 tended to have higher AUCs than CD20. Both discriminated well between the inflammatory arthropathies and normal controls or, to a lesser extent, OA. Both distinguished OA from normal tissue with intermediate AUCs, but only CD38 separated Orth.A from normal tissue. CD38 was also the only marker that was inversely related to a diagnosis in that it had a statistically significant AUC <0.5 for the comparison SeA vs RA (the AUC of CD20 for this comparison was <0.5, too, but the 95% confidence interval crossed above this value and was thus

not significant). Consistent with our previous results (Pessler et al. 2008d), Ki-67 distinguished well between RA and the less inflamed specimens. In addition, it had good discriminatory abilities for the comparisons SeA vs RA (AUC 0.77, Youden index 0.41), SeA vs early arthritis (AUC 90.0, Youden index 0.62) and OA vs normal synovium (AUC 0.84, Youden index 0.53). vWF separated the inflamed arthropathies accurately from normal tissue. In addition, it was the best marker for the distinction between Orth.A and normal synovium (AUC 0.81, Youden index 0.56) and discriminated well between the two most highly inflamed arthropathies, SeA vs RA.

Ranking the markers with multicategory ROC analysis

Multicategory ROC analysis (Li & Fine 2008, Li & Zhou 2009) was then used to establish a hierarchy of the markers according to their discriminatory abilities in all of the diagnostic comparisons (Table 2). Considering the numbers of tests and classes analysed, the individual HUMs as well as the rank order were statistically significant (Li & Fine 2008). The rank order by HUM values identified Ki-67, subintimal CD68, and CD15 as the markers with the best overall discriminatory abilities. The lowest HUM resulted for CD20, followed by vWF. Importantly, the HUM of the weakest marker (CD20) was still markedly higher than the HUM of the null hypothesis, indicating that even CD20 had a substantial overall discriminatory ability.

Comparing ROC curve AUCs and fold differences in mean expression

We then asked whether AUCs correlated with the ratios in mean expression of the same paired comparisons. Linear regression using the original linear values revealed no correlation ($r^2 = 0.051$). As the ratios of mean expression spread over a 1000-fold range, they were also transformed on a $\log_{10}$ scale. Regression with these log-transformed values revealed a modest positive correlation between AUCs and expression ratios with an r^2 value of 0.44 (Figure 3).

Table 2. Rank of markers by multicategory ROC analysis.

Rank	Marker	Hypervolume under the ROC manifold (HUM)[a]
1	Ki-67	0.106
2	CD68S	0.090
3	CD15	0.087
4	CD3	0.068
5	CD38	0.059
6	vWF	0.040
7	CD20	0.017
8	Null hypothesis[b]	0.0014

[a]HUMs were calculated using code written in the R environment for statistical calculations as described in (Li & Fine 2008). [b]Hypothetical marker with no discriminatory value and a HUM of the null hypothesis (1/K!, i.e. 1/6! = 0.0014).
CD68S, subintimal CD68. vWF, von Willebrand factor.

Discussion

We used ROC analysis to explore the usefulness of eight immunohistochemical markers in the differentiation and classification of synovial tissue specimens. While haematoxylin and eosin (H&E) stained specimens can be used to measure the degree of inflammation (e.g. using the synovitis score according to Krenn and colleagues (Krenn et al. 2002, 2006)), or to identify pathognomonic histological features (Schumacher & Kulka 1972, Schumacher et al. 2008), the use of diligently quantified immunohistochemical parameters may result in a higher sensitivity to differences and may thus yield more precise results. Along these lines, prior studies have demonstrated the value of subintimal CD68 in diagnosing RA (Kraan et al. 1999) and predicting and monitoring response to therapy in RA (Haringman et al. 2005, Jahangier et al. 2006). The present ROC analysis supports those results fully in that subintimal CD68 distinguished RA accurately from OA or normal tissue. It also lends further support to subintimal CD68 as a broadly applicable synovial tissue marker insofar as it had high AUCs for other comparisons, most notably in delineating OA from normal tissue, for which it was the best marker. Moreover, multicategory ROC analysis ranked it among the three best overall markers. Previous studies have shown enrichment of CD68+ cells in the

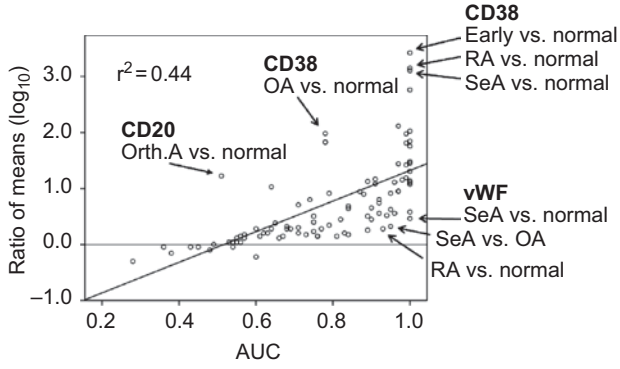

Figure 3. Correlation between ratios of means and area under curve (AUC) values. Linear regression analysis was performed with the SPSS 15.0 biostatistical software, using the values for ratios of means and AUCs (see Table S1 in online version of this article and Table 1) for the same paired comparisons. The ratio of means was calculated by dividing the mean expression density of the marker in the state variable by its mean expression density in the test variable. The arrows indicate selected discrepancies between ratios of means and AUCs, some of which are discussed in the text.

synovial lining in a subset of OA patients (Pessler et al. 2008a, Oehler et al. 2002). The moderately high AUC of 0.79 in the comparison OA versus normal controls suggests that this may be a relatively specific feature that could potentially be included in evaluating OA synovia. The association of increased lining CD68 with OA, compared with other 'non-inflammatory' arthropathies, is also supported by the observation that, although the AUC for the comparison OA versus Orth.A was only moderate with 0.71, the specificity for a diagnosis of OA was 91% at the trade-off value. The present analysis also highlights other markers that were underappreciated in the past. We have previously shown that, as expected pathogenetically, CD15 is highly expressed in chronic septic arthritis (Pessler et al. 2008c) compared with RA. The present ROC analysis revealed that it is also specifically associated with a diagnosis of RA or – to a somewhat lesser extent – early arthritis compared with OA or normal tissue, suggesting that CD15+ cells may play pathogenetic roles in RA and early arthritis. Increased angiogenesis has been documented well in inflammatory arthropathies and even OA (Pessler et al. 2008c, Maruotti et al. 2006). Consistent with this, vWF had high AUCs for several comparisons between the inflammatory arthropathies and less inflamed or normal tissue. Notably, it discriminated best between normal and Orth.A tissue (suggesting increased angiogenesis even in this mildly inflamed cohort), and multicategory ROC analysis revealed that it had a better overall discriminatory ability than CD20.

A general advantage of ROC analysis is that commonly used diagnostic parameters such as sensitivities, specificities, Youden index and accuracy (sensitivity + specificity/2; not included in the present analysis) can be calculated. Moreover, the AUC values fall into a defined range and can thus be compared with values from other tests or specimen cohorts. Whereas synovial tissue markers have not been evaluated with ROC analysis in other specimen cohorts, other rheumatological studies have employed this method. For instance, in a study on the ability of common blood and synovial fluid markers to diagnose septic arthritis, the best marker was synovial fluid leukocyte count with an AUC of 0.75 and a Youden index of 0.38 (Li et al. 2007). In one recent study of the ability of anti-CCP antibodies to diagnose RA, the AUC in the distinction between patients with RA ($n=120$) and individuals with other rheumatic diseases ($n=71$) or healthy controls ($n=50$) was 0.83 (Yang et al. 2007). Higher AUCs and Youden indexes were often observed in our present study, highlighting the good discriminatory power of the immunohistochemical markers for some of the comparisons.

There was only a modest correlation between AUCs and ratios in mean expression, suggesting that the two parameters measure different outcomes in some situations. This is best exemplified by CD38 and vWF, two markers located at opposite ends of the spectrum of mean expression ratios. CD38 had the highest absolute fold differences in mean expression in seven of 12 comparisons but had the single best AUC only in one case. On the other hand, the fold differences in mean vWF expression were small, with the greatest being 3.0 for the comparison SeA versus normal. Nonetheless, it achieved an AUC of 1 for this comparison and an AUC of 0.81 for the comparison Orth.A versus normal, for which the fold difference in mean expression was only 1.4. Thus, ROC curve analysis may reveal substantial differences in expression that may not be evident from comparing the difference between two means. It may therefore be useful in characterizing tissue markers that possess either unusually high or low differences in mean expression.

Multicategory ROC analysis allowed us to rank the markers according to their overall discriminatory abilities. Thus, it identified Ki-67 as the best overall marker for the distinctions chosen for the present study and, in addition to the well-established subintimal CD68, highlighted CD15 as a potentially valuable synovial biomarker for inflammatory arthropathies. While the choice of these specimen cohorts does not fully reflect the prevalence of arthropathies encountered in clinical practice (for instance, the spondyloarthropathies were not included), it does include RA and early undifferentiated arthritis, i.e. two common inflammatory arthropathies, and overall a mixture of inflammatory, degenerative and normal specimens. While the HUMs of the present multicategory ROC analysis were statistically significant considering the numbers of tests and classes included in the analysis, the HUM for a test with perfect discriminatory ability among all of the specimen cohorts would be 1.0 (Li & Fine 2008). Considering the similarities among the SeA, RA and early arthritis cohorts and between OA and Orth.A, this value probably cannot be achieved with any of the immunohistochemical markers available to date. Indeed, this notion clearly demonstrates that the discriminatory abilities of the markers used here are still far from optimal. Multicategory ROC analysis could now be used to identify combinations of two or more markers with better overall discriminatory abilities (higher HUMs) than the single markers. Likewise, there were several difficult differentiations, e.g. RA vs early arthritis and OA vs Orth.A. In these situations, too, combinations of two or more markers would have higher discriminatory abilities, and multicategory ROC analysis would be an appropriate method to identify the optimal combination (Li & Fine 2008).

The present study highlights the use of ROC analysis in synovial tissue classification and underscores the value of

immunohistological markers in making a tissue diagnosis in certain scenarios. We believe that another powerful application of ROC curve analysis might be in clinical trials where synovial biopsies are available. For instance, it should lend itself well to identifying markers that predict a treatment response or correlate with it, e.g. by discriminating between responders and non-responders. It should also be possible to refine this approach further and use multicategory ROC analysis to identify the best combination of markers for this or similar distinctions, or to identify the best marker for the discrimination among multiple clinical outcomes.

Acknowledgements

We thank Gilda Clayburne for expert technical assistance, Irene Loglisci, Roxann Thompson, Valice Matthews and John Romano (Philadelphia VA Medical Center, Department of Pathology) for help with histology, Lan X. Chen (Presbyterian Hospital, Philadelphia, PA) for helpful discussion and Angela Rösen-Wolff (University Children's Hospital, Technical University of Dresden, Germany) for helpful discussion and a critical reading of the manuscript. Support is gratefully acknowledged through National Institutes of Health Training Grant T32-CA 09140, a 2007-8 American College of Rheumatology/Research and Education Foundation Amgen Pediatric Rheumatology Research Award, and International Collaboration Seed Grant SGP 08/A01 from the German Federal Ministry for Education and Science (all to F.P.). The study was approved by the institutional review boards of the Philadelphia VA Medical Center, the University of Pennsylvania and the 2nd Affiliated Hospital of Sun Yat-sen University (Guangzhou, China).

***Declaration of interest*:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Diaz-Torne C, Schumacher HR, Yu X, Gomez-Vaquero C, Dai L, Chen LX, Clayburne G, Einhorn E, Sachdeva RM, Singh JA, Pessler F. (2007). Absence of histologic evidence of synovitis in patients with Gulf War veterans' illness with joint pain. *Arthritis Rheum (Arthr Care Res)* 57:1316–23.

Faraggi D, Reiser B. (2002). Estimation of the area under the ROC curve. *Stat Med* 21:3093–106.

Haringman J, Gerlag D, Zwinderman A, Smeets T, Kraan M, Baeten D, Mcinnes I, Bresnihan B, Tak P. (2005). Synovial tissue macrophages: sensitive biomarkers for response to treatment in patients with rheumatoid arthritis. *Ann Rheum Dis* 64:834–8.

Jahangier ZN, Jacobs JW, Kraan MC, Wenting MJ, Smeets TJ, Bijlsma JW, Lafeber FP, Tak PP. (2006). Pretreatment macrophage infiltration of the synovium predicts the clinical effect of both radiation synovectomy and intra-articular glucocorticoids. *Ann Rheum Dis* 65:1286–92.

Kraan MC, Haringman JJ, Post WJ, Versendaal J, Breedveld FC, Tak PP. (1999). Immunohistological analysis of synovial tissue for differential diagnosis in early arthritis. *Rheumatology (Oxford)* 38:1074–80.

Krenn V, Morawietz L, Burmester G, Kinne R, Mueller-Ladner U, Muller B, Al E. (2006). Synovitis score: discrimination between chronic low-grade and high-grade synovitis. *Histopathology* 49:358–64.

Krenn V, Morawietz L, Haupl T, Neidel J, Peterson I, Konig A. (2002). Grading of chronic synovitis – a histopathological grading system for molecular and diagnostic pathology. *Pathol Res Pract* 198:317–25.

Li J, Fine JP. (2008). ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies. *Biostatistics* 9:566–76.

Li J, Zhou XH. (2009). Nonparametric and semiparametric estimation of the three way receiver operating characteristic surface. *J Stat Planning Int* 139:4133–42.

Li SF, Cassidy C, Chang C, Gharib S, Torres J. (2007). Diagnostic utility of laboratory tests in septic arthritis. *Emerg Med J* 24:75–7.

Maruotti N, Cantatore FP, Crivellato E, Vacca A, Ribatti D. (2006). Angiogenesis in rheumatoid arthritis. *Histol Histopathol* 21:557–66.

Oehler S, Neureiter D, Meyer-Scholten C, Aigner T. (2002). Subtyping of osteoarthritic synoviopathy. *Clin Exp Rheumatol* 20:633–40.

Pessler F, Chen L, X, Dai L, Gomez-Vaquero C, Diaz-Torne C, Paessler ME, Scanzello C, Cakir N, Einhorn E, Schumacher HR. (2008a). A histomorphometric analysis of synovial biopsies from individuals with Gulf War Veterans' Illness and joint pain compared to normal and osteoarthritis synovium. *Clin Rheumatol* 27:1127–34.

Pessler F, Dai L, Diaz-Torne C, Gomez-Vaquero C, Paessler ME, Zheng DH, Einhorn E, Range U, Scanzello C, Schumacher HR. (2008b). The synovitis of 'non-inflammatory' orthopaedic arthropathies: a quantitative histological and immunohistochemical analysis. *Ann Rheum Dis* 67:1184–7.

Pessler F, Dai L, Diaz-Torne C, Ogdie A, Gomez-Vaquero C, Paessler ME, Einhorn E, Chen LX, Schumacher HR. (2008c). Increased angiogenesis and cellular proliferation as hallmarks of the synovium in chronic septic arthritis. *Arthritis Rheum* 59:1137–46.

Pessler F, Ogdie A, Diaz-Torne C, Dai L, Yu X, Einhorn E, Gay S, Schumacher HR. (2008d). Subintimal Ki-67 as a synovial tissue biomarker for inflammatory arthropathies. *Ann Rheum Dis* 67:162–7.

Schumacher HR, Chen LX, Pessler F. (2008). Synovial biopsy in the evaluation of nonrheumatic systemic diseases causing arthritis. *Curr Opin Rheumatol* 20:61–5.

Schumacher HR Jr, Kulka JP. (1972). Needle biopsy of the synovial membrane –experience with the Parker–Pearson technique. *N Engl J Med* 286:416–19.

Yang CL, Zhou YR, Zhou F, Li J. (2007). [Diagnostic value of serum anti-cyclic citrullinated peptide antibodies in patients with rheumatoid arthritis]. *Nan Fang Yi Ke Da Xue Xue Bao* 27:1277–9.

Youden WJ. (1950). Index for rating diagnostic tests. *Cancer* 3:32–5.

Zweig MH. (1988). Evaluation of the clinical accuracy of laboratory tests. *Arch Pathol Lab Med* 112:383–6.

Zweig MH, Broste SK, Reinhart RA. (1992). ROC curve analysis: an example showing the relationships among serum lipid and apolipoprotein concentrations in identifying patients with coronary artery disease. *Clin Chem* 38:1425–8.

Zweig MH, Campbell G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39:561–77.