

## EXTENDED REPORT

# Technical validation of cDNA based microarray as screening technique to identify candidate genes in synovial tissue biopsy specimens from patients with spondyloarthritis

M Rihl, D Baeten, N Seta, J Gu, F De Keyser, E M Veys, J G Kuipers, H Zeidler, D T Y Yu

*Ann Rheum Dis* 2004;**63**:498–507. doi: 10.1136/ard.2003.008052

See end of article for authors' affiliations

Correspondence to:  
Dr M Rihl, Hannover  
Medical School (MH),  
Department of  
Rheumatology (OE 6850),  
Carl-Neuberg-Str. 1,  
30625 Hannover,  
Germany; Rihl.Markus@  
MH-Hannover.de

Accepted 18 August 2003

**Objectives:** To validate the use of cDNA based microarray on synovial biopsies by analysing the experimental variability due to amplification of RNA, reproducibility of the assay, heterogeneity of the tissue, and statistical analysis.

**Methods:** Total RNA was extracted from three spondyloarthritis (SpA) and three osteoarthritis (OA) synovial tissue biopsy specimens and from the peripheral blood mononuclear cells (PBMC) of four healthy donors. Exponential RNA amplification by SMART-PCR was compared with linear amplification. Reproducibility was tested by comparing different microarray systems and by performing duplicate experiments. Sample heterogeneity was assessed by comparing overall gene expression profiles, histopathology, and analysis of genes expressed in the synovium and normal PBMC. Statistical analysis using *t* test and Bonferroni adjustment was verified by permutation of class labels.

**Results:** Gene expression was concordant in 12/14 (86%) cytokine/chemokine genes between both microarrays and different RNA amplification systems. When one microarray system was used, expressed genes were 78–95% concordant in duplicate experiments. Gene expression profiles had a higher degree of similarity between SpA synovium than between PBMC or OA synovium despite clear histopathological differences between synovial samples. Comparison of SpA synovium with OA synovium and with PBMC yielded 11 and 18 expressed transcripts, respectively; six were shared in both comparisons. Permutations of SpA and OA samples yielded only one expressed gene in 19 comparisons.

**Conclusions:** These data provide evidence that microarrays can be used for analysis of synovial tissue biopsies with high reproducibility and low variability of the generated gene expression profiles.

Microarrays are increasingly used for random screening of mRNA transcripts from cells and tissues. The technology is based on the principle that the biology of an organism corresponds with its gene expression profile disclosing preferentially expressed genes by comparative analysis. It allows profiling and simultaneous monitoring of a vast number of genes in one single hybridisation experiment. An increasing body of evidence suggests that microarrays are a valuable tool in the study of complex diseases, even though one needs to keep in mind that the technique is merely a screening approach contributing to hypothesis driven research. One basic potential lies in the detection of differentially expressed genes that would not have been expected by the current knowledge base, eventually leading to generation of new hypotheses about mechanisms in physiology and pathology. Recently, it has been shown that molecular profiling of various cancer tissues by microarray provides better recognition of new disease entities and prediction of prognosis or response to treatment than standard clinical or histological methods.<sup>1–3</sup>

We have previously used microarray technology with high technical reproducibility to screen for potentially disease mediating mRNA transcripts in peripheral blood and synovial fluid mononuclear cells from patients with active spondyloarthritis (SpA).<sup>4–5</sup> Because the synovium is probably the primary site of inflammation within the arthritic joint,<sup>6</sup> it naturally is the relevant target structure for further study of specific disease mechanisms by microarray. However, to date there are no published reports of gene expression profiles from SpA synovial tissue biopsy specimens and very few

reports are available about rheumatoid arthritis synovium.<sup>7,8</sup> This is probably because the use of microarray on synovial tissue biopsy specimens raises a number of important questions about the technical aspects and the resulting scientific validity, even before starting to confirm microarray results by independent methods such as reverse transcriptase-polymerase chain reaction (RT-PCR) or immunohistochemistry. As clearly pointed out in a recent editorial,<sup>9</sup> these aspects include reproducibility of methods, sample heterogeneity, statistical analysis with thresholds of arbitrary nature, and the apprehension that comparison of different patients and heterogeneous tissue composition (resident versus migrated inflammatory cells) reflecting the individual disease process might account for a great deal of experimental variability. Moreover, a technical aspect of specific interest for small synovial tissue biopsy specimens as obtained by needle arthroscopy is the quality and quantity of RNA/cDNA and the requirement for its amplification, possibly causing a bias in amplification of high or low abundance mRNA molecules, leading to a major distortion of gene expression profiles.

This study aimed at exploring and validating the technical aspects of microarray for analysis of synovial tissue biopsies, using SpA as a prototype of chronic inflammatory arthritis.

**Abbreviations:** DMARDs, disease modifying antirheumatic drugs; HC, healthy controls; LPS, lipopolysaccharide; OA, osteoarthritis; PBMC, peripheral blood mononuclear cells; RT-PCR, reverse transcriptase-polymerase chain reaction; SMART, Switch Mechanism At the 5' end of the RNA Template; SpA, spondyloarthritis

Firstly, we analysed the reproducibility of the technique by studying exponential RNA amplification and the variability between different arrays on peripheral blood mononuclear cells (PBMC), as well as the run-to-run variability on synovial tissue biopsies. Secondly, we assessed the effect of sample heterogeneity on microarray results by analysing the gene expression profiles of SpA synovium, osteoarthritis (OA) synovium, and PBMC, and by comparing the results with the histopathology of the SpA and OA samples. Finally, we analysed the validity of the classical statistical approach (analysis of variance, Student's *t* test with Bonferroni adjustment) and the appropriateness of the chosen thresholds by permutation testings of data of the SpA and OA synovial tissue samples. This approach was undertaken to demonstrate that the particular analysis can yield a low number of false positive results and that an increase in the number of false positive genes inevitably occurs when less stringent thresholds are used.

## METHODS

### PBMC and synovial biopsy samples

PBMC from four healthy donors (HC) were separated by a standard Ficoll-Histopaque procedure. PBMC from one healthy male donor were resuspended in serum-free RPMI medium (Gibco Invitrogen, Carlsbad, CA) and incubated in a tissue culture plate for 20 minutes at room temperature. The adherent cell fraction contained more than 80% of cells with macrophage-like appearance as determined by inverted microscope and described earlier.<sup>5</sup> These cells were further subjected to incubation with 10 ng/ml of *Salmonella*-lipopolysaccharide (LPS; Sigma, St Louis, OH) for 4 hours. Exerting this stimulus ensured enhanced expression of a considerable number of genes specifically encoding for proinflammatory molecules to be analysed.<sup>10</sup> The model of LPS stimulated monocytes uses the subtractive potential of the microarray analysis, which enables determination of comparability of amplified versus unamplified cDNA and reproducibility of two different microarray assays.

Synovial tissue biopsy samples were obtained by needle arthroscopy in three patients with SpA and three with OA as previously described<sup>11</sup>; six small biopsy pieces of the synovial membrane were collected from each patient. All patients with SpA fulfilled the European Spondylarthropathy Study Group (ESSG) classification criteria.<sup>12</sup> These patients were chosen for analysis because they had active and early disease with an average time period from beginning of symptoms to the date of biopsy of 5.6 (range 3–8) months. They were all tested positive for the HLA-B27 allele. Patients had not received treatment with disease modifying antirheumatic drugs (DMARDs) at the time of biopsy. Both the synovial tissue biopsy specimens from the three patients with OA and the resting PBMC from four healthy donors were used as control samples (table 1 gives detailed characteristics of the patients and control subjects). All patients had given their informed consent, and the procedures followed were in accordance with the ethical standards of the University of California, Los Angeles.

### Synovial histopathology

Synovial tissue sections were available in two of three patients with SpA and two of three patients with OA. Histological and immunohistochemical analyses were performed as reported in a previous publication.<sup>6</sup> Briefly, the histological measures included synovial lining hyperplasia, degree of vascularity, degree of infiltration, presence of lymphoid aggregates, plasma cells, and polymorphonuclear cells. Immunohistochemical peroxidase staining (LSAB+kit, Dako, Glostrup, Denmark) was performed for T cells (anti-CD3, Dako), B cells (anti-CD20, Dako), macrophages

(anti-CD68, Dako), plasma cells (anti-CD138, Dako), and endothelial cells (anti-CD146, Dako). The parameters were scored semiquantitatively as described earlier.<sup>6</sup>

### RNA extraction and RNA/cDNA amplification

Before RNA extraction, synovial tissues were ground by a pellet pestle set and homogenised in denaturing medium (solution D; Stratagene, La Jolla, CA). PBMC/monocytes were resuspended in solution D. Total RNA was extracted by four serial applications of phenol:chloroform 5:1 (pH 4.5; Ambion, Austin, TX), followed by two precipitations in isopropanol at  $-80^{\circ}\text{C}$  for 1 hour and incubation with RQ1 DNase (Promega, Madison, WI) at  $37^{\circ}\text{C}$  for 1 hour. RNA was extracted again with phenol:chloroform and precipitated in 100% ethanol at  $-80^{\circ}\text{C}$  for 30 minutes. Total RNA from monocytes before and after stimulation with LPS was split into aliquots of 150 ng, 300 ng, and 5  $\mu\text{g}$  samples. The 150 ng and 300 ng aliquots were subjected to SMART-PCR (BD Biosciences Clontech, Palo Alto, CA) in a total reaction volume of 50  $\mu\text{l}$ . The technology (Switch Mechanism At the 5' end of the RNA Template) allows reverse transcription of small amounts of total RNA and subsequent amplification of the entire cDNA.<sup>13</sup> For our studies, conditions of the PCR reaction were determined as follows: after reverse transcription using cDNA-specific primer and SMART II-oligonucleotide for annealing, 10  $\mu\text{l}$  of single stranded cDNA dissolved in TE buffer (5 mM Tris, 1 mM EDTA, pH 7.5) was mixed with 32  $\mu\text{l}$  of molecular biology grade water, 5  $\mu\text{l}$  of  $10\times$  reaction buffer, 1  $\mu\text{l}$  of  $50\times$  dNTP, 1  $\mu\text{l}$  of specific PCR primer, and 1  $\mu\text{l}$  of  $50\times$  polymerase mix. After denaturation at  $95^{\circ}\text{C}$  for 2 minutes, one PCR reaction cycle was performed at  $95^{\circ}\text{C}$  for 1 minute,  $65^{\circ}\text{C}$  for 1 minute, and  $68^{\circ}\text{C}$  for 6 minutes. Termination mix (2  $\mu\text{l}$ ) was added after completion of the reaction. PCR product (5  $\mu\text{l}$ ) was subjected to gel electrophoresis (1.5% agarose in  $0.5\times$  TE buffer) followed by ethidium bromide staining. Cycle numbers of the PCR reaction were conservatively optimised in order to restrict the amplification reaction to the exponential phase, strictly avoiding any over- or underamplified templates. This was achieved by choosing a cycle number two cycles fewer than the one that showed no more increase in band intensity—that is, the start of the plateau phase.

### Microarrays

After amplification with the optimal cycle number, cDNA was purified using QIAquick PCR purification kit (Qiagen, Santa Clarita, CA) and mixed with the rediprime II dry beads random prime labelling system, followed by incubation with  $^{32}\text{P}$  at  $37^{\circ}\text{C}$  for 10 minutes (both Amersham Pharmacia Biotech, Piscataway, NJ). Probes were purified from unincorporated nucleotides using NucleoSpin purification kit (BD Biosciences Clontech), heat denatured for 5 minutes at  $99^{\circ}\text{C}$ , and hybridised for 20 hours at  $68^{\circ}\text{C}$  to the cDNA based nylon membrane Atlas Human 1.2 array containing 1185 genes (BD Biosciences Clontech). Microarray experiments of 5  $\mu\text{g}$  total RNA aliquots were performed using the standard linear procedure as described in a previous publication.<sup>5</sup> To validate findings generated with the Atlas array, we used another microarray system (cDNA based nylon membrane GeneFilters, GF211, containing 4136 genes, Research Genetics Invitrogen, Carlsbad, CA). Again 150 ng and 300 ng RNA aliquots were applied by using the described SMART-PCR amplification, whereas the 5  $\mu\text{g}$  sample was subjected to the linear conventional procedure according to the instructions of Research Genetics Invitrogen as outlined in the protocol. After exposure to a phosphor screen cassette for 2 days, all membranes were scanned through a STORM 860 scanner (Molecular Dynamics, Sunnyvale, CA). The software packages "AtlasImage 1.5" (BD Biosciences

**Table 1** Characteristics of patients and healthy subjects

Label/diagnosis	Sex	Age	DD (months)	SJC	HLA-B27	RF	ESR (mm/1st h)	CRP (mg/l)	Drug treatment
SpA1 (AS)	M	39	8	1	Pos	Neg	34	59.6	NSAID
SpA2 (AS)	M	44	6	2	Pos	Neg	21	8.2	NSAID
SpA3 (uSpA)	F	21	3	3	Pos	Neg	9	5.5	None
OA1	F	72	60	0	Neg	Pos	8	2.6	NSAID
OA2	F	65	12	1	Neg	NA	12	3.7	None
OA3	M	74	6	2	Neg	Neg	21	5.0	NSAID
HC1	M	38	–	0	NA	NA	NA	NA	None
HC2	M	32	–	0	Pos	NA	NA	NA	None
HC3	M	32	–	0	Neg	Neg	NA	NA	None
HC4	M	40	–	0	NA	NA	NA	NA	None

Patients with SpA all had active disease with disease duration from the beginning of symptoms to the time of biopsy of less than 1 year. All three patients with SpA and one healthy control were positive for the HLA-B27 allele.

DD, disease duration; SJC, swollen joint count; RF, rheumatoid factor; ESR, erythrocyte sedimentation rate; CRP, C reactive protein; AS, ankylosing spondylitis; uSpA, undifferentiated spondyloarthritis; NA, not assessed; NSAID, non-steroidal anti-inflammatory drug.

Clontech) and “Pathways 2.01” (Research Genetics) retrieved signal intensities of each gene. The local background intensity measured around each gene was then subtracted from all signal intensities. The local background intensities of all individual genes were subsequently averaged, resulting in the mean background intensity of a particular membrane. Because normalisation and all further calculations were performed with data from which the local background had already been subtracted, thresholds for analysis were derived from the mean background of all membranes used in the analysis. Microarray experiments from all synovial tissues and healthy PBMC used 150 ng of total RNA amplified by SMART-PCR as described above. Labelling and hybridisation of probes to Human Atlas 1.2 membranes were performed also as described above. For a complete list of genes see <http://www.clontech.com/atlas/genelists/index.shtml>.

### Data analysis and statistics

The G3PDH signal was used as indirect quality marker for the RNA used as only experiments were accepted that had a G3PDH signal lying within 1.5 times the standard deviation of all membranes evaluated. Normalisation was performed as previously described.<sup>4,5</sup> A gene was considered to be LPS induced when the signal intensity of the particular mRNA transcript was at least twice as high as the average background signal of all membranes used in that particular experiment, and at least a twofold increase compared with non-LPS stimulated monocytes was seen. Because we were screening for LPS induced genes, we preferred the sensitivity in these comparisons to be high and accordingly the chosen ratio cut off point was 2. However, it must be taken into account that a low threshold inevitably decreases specificity, risking a higher number of false positive genes.

For the comparison of synovial tissue samples with normal PBMC, classification of gene expression data was achieved by hierarchical cluster analysis, implementing pairwise similarity function (average linkage analysis, Cluster-TreeView<sup>14</sup>;) for clustering genes and between-groups linkage analysis with Pearson correlation (SPSS 10.0 for Windows) for clustering samples. Results of clustered samples were visualised by dendrogram, reflecting the degree of similarity between the various expression datasets. Before statistical testing, two criteria had to be fulfilled in this study: firstly, difference of mean signal intensities from each transcript in the two comparison groups (SpA *v* OA, SpA *v* HC) was required to be more than twofold higher than the background level of all membranes used. Secondly, signal intensity ratios of disease versus control (SpA *v* OA, SpA *v* HC) were required to be higher than 4 in order to achieve a more stringent cut off point for this particular analysis. However, it should be noted that thresholds for signal to

background ratio and signal increase in microarray studies are arbitrary. A signal increase four times higher than the control group can raise specificity by lowering the number of false positive genes, on the one hand, but will also lead to a lower sensitivity, on the other.

As for the generation of statistically valid data, genes that met the above mentioned criteria were subjected to Student's *t* test (two tailed, unpaired) in combination with Hartley's *F*<sub>max</sub> test for homogeneity of variance. Bonferroni adjustment of  $\alpha$  was performed in relation to the number of genes fulfilling the above criteria. A value of  $p < 0.05$  after Bonferroni correction was considered to be significant—that is,  $p < 0.0017$  in the comparison of SpA *v* OA synovium and  $p < 0.00063$  in the comparison of SpA synovium *v* healthy PBMC. We also calculated the *p* values for all possible rearrangements of signal intensity data from all samples in both of the comparisons (19 permutations in SpA *v* OA and 34 permutations in SpA *v* HC). To confirm the validity of our analysis, we reanalysed (difference of mean higher than twofold background signal; ratio—that is, signal increase higher than 4; analysis of variance, Student's *t* test, Bonferroni adjustment) the SpA and OA synovial tissue samples by shuffling the complete set of all signal intensity data (permutation of class labels). The same permutational analysis was performed with less stringent criteria (difference of mean higher than only onefold instead of twofold background signal; ratio higher than 2 instead of 4; analysis of variance, Student's *t* test without Bonferroni adjustment—that is, a *p* value  $< 0.05$  was considered to be significant) in order to test to what extent the number of false positively expressed genes would increase.

## RESULTS

### Validation and reproducibility of the exponential RNA amplification method

This experiment was performed to demonstrate that RNA aliquots with different amounts of starting material amplified either linearly or exponentially yield reproducible results.

Complementary DNA templates were generated by exponential RNA amplification of two different RNA aliquots (150 and 300 ng) from monocytes of one healthy subject before and after LPS stimulation and were then analysed by microarray. To test the global reproducibility of this technique, correlation matrices using Pearson correlation were calculated between the 150 ng and the 300 ng Atlas array datasets: coefficients were 0.94 and 0.99 for adherent PBMC before and after LPS stimulation, respectively, with both correlations being significant at the 0.01 level (see fig 1A for XY scatter plots). Because this global reproducibility appeared to be high, further analysis of the reproducibility of individual gene expression was performed. As a number of



cytokines and chemokines are known to be strongly up regulated after LPS stimulation of monocytes,<sup>10</sup> these experiments mainly focused on this particular family of mRNA transcripts. Cytokine/chemokine genes were found to be expressed on a threefold higher level than all other LPS induced genes. The 150 ng and 300 ng exponentially amplified samples and the 5 µg linearly amplified samples were analysed by two different arrays: the Atlas 1.2 array containing a total number of 68 cytokines/chemokines and the GF211 array containing a total number of 59 cytokines/chemokines, with 32 overlapping cytokines/chemokines between both arrays. Table 2 shows that with the Atlas array nine of 68 identical cytokines/chemokines were detected in both the 150 ng and the 300 ng RNA aliquot samples obtained by exponential amplification. Moreover, six of these genes were also detected in the conventionally prepared sample containing 5 µg of total RNA, whereas only one additional cytokine was detected in this sample but in neither of the SMART-PCR samples (IL1β). The same analysis using the GF211 array yielded similar results: eight of 59 cytokines/chemokines were detected in the two 150 ng and 300 ng SMART-PCR samples as well as in the conventionally prepared sample. Again, only two chemokines were detected in the 5 µg sample but not in the two SMART-PCR samples, while one cytokine (IL1β) was detected in the 5 µg sample and in only one of the two SMART-PCR samples.

In the SMART amplified RNA samples, seven LPS induced genes other than cytokine/chemokine genes were expressed in both the Atlas Array and the GF211 systems: metalloproteinase inhibitor 1, urokinase inhibitor, IEX-1L anti-death protein, guanine nucleotide binding protein, intercellular adhesion molecule-1, epidermal growth factor receptor, and protease inhibitor 19. Of those seven transcripts, only one transcript (IEX-1L anti-death protein) was detected in all samples irrespective of the mode of amplification and the array system used.

#### Variability between different microarrays

This experiment was performed to demonstrate that RNA aliquots hybridised to two different nylon based array systems yield reproducible results.

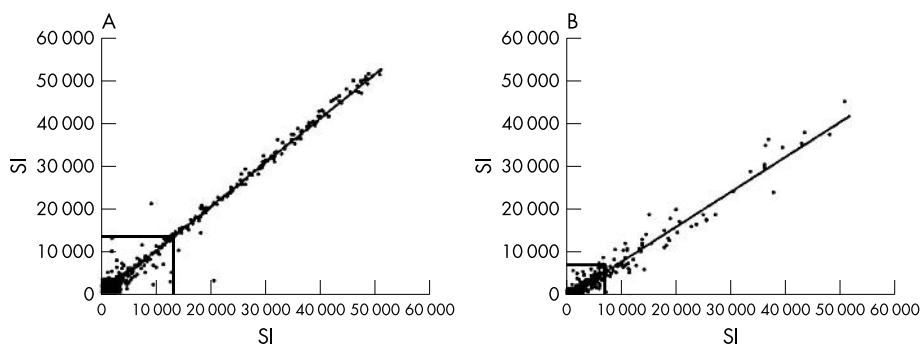
Using the same dataset (table 2), the three different samples were compared between the two microarray systems

focusing on the 32 cytokines/chemokines immobilised on both of the two nylon membranes used. In seven of the eight genes that were expressed in at least one of the six experiments performed, the results were concordant in at least five experiments: four cytokines/chemokines were expressed in all six experiments, two genes were expressed in five of all six experiments. For one cytokine (IL1β) and one chemokine (MCP-4), the expression profile was discordant. Based on these results, it was decided to use 150 ng of total RNA because this would allow synovial tissue biopsy samples with very low yields of RNA to be studied. The SMART-PCR and Atlas array were considered to be sufficiently sensitive for the desired approach and were therefore used for the following experiments.

#### Run-to-run variability of microarray on synovial tissue biopsies

Two RNA aliquots from two synovial tissue samples each were tested by duplicate experiments in order to demonstrate a low run-to-run variability.

Although exponential amplification of RNA appeared to be reproducible, microarray data obtained from synovial tissue biopsies might still be biased owing to variability of the microarray procedure itself rather than owing to variability of the RNA amplification. To examine this issue, SMART-PCR amplified cDNA templates from synovial tissue biopsy samples of two patients with ankylosing spondylitis were submitted to two microarray testings each. Correlation coefficients of signal intensities of all genes showed high values between the two datasets of both patient samples: 0.98 and 0.95, respectively, at a significance level of 0.01 (see fig 1B for XY scatter plots). Genes identified as being expressed in this analysis were defined as genes with a signal intensity higher than twice the background level of the membranes used. When we focused on individual genes, analysis of the first patient sample indicated 78 expressed transcripts in the first array and 64 expressed transcripts in the second array, with 61 expressed in both testings (61/78 (78%)). Among those 61 expressed genes, the highest ratio of expression values between duplicate experiments—that is, RNA aliquots from the identical sample, was 1.8. When the same analysis was carried out on the second patient sample, 21 expressed genes were found in both the first and the second array,



**Figure 1** XY scatter plots for signal intensity data as generated by Atlas 1.2 microarray from different RNA aliquots. (A) displays the XY scatter plot of signal intensity data as generated from LPS stimulated monocytes of a healthy donor. Two different RNA aliquots from this donor (150 and 300 ng) were amplified by SMART-PCR and subsequently assayed by the Human Atlas 1.2 microarray system. The two datasets yield reproducible signal intensities as can be seen by their almost identical distribution with very few outliers. Calculation showed that the correlation coefficient between the two datasets was 0.99 (significance level 0.01; Pearson correlation). The local background was subtracted from the signal intensity data, which were normalised. The bold lines depict the onefold background level at 13 400 (SI, signal intensity; arbitrary units). (B) Signal intensities from a duplicate experiment using identical aliquots from a synovial tissue biopsy sample are shown. Total RNA (150 ng) was amplified by SMART-PCR and hybridised to the Atlas 1.2 array membrane. In comparison with LPS stimulated monocytes, most genes are expressed on a lower level, which is typically found in synovial tissue samples. Accordingly, the onefold background level was lower at 6700. Also, the distribution of signal intensities was more scattered as compared with fig 1A but still shows a reasonably high reproducibility. Here, the correlation coefficient was 0.98 (significance level 0.01; Pearson correlation).

**Table 2** Gene expression profiles of cytokines/chemokines in LPS stimulated monocytes as assessed by different RNA amplification methods and array systems

Up regulated cytokines/ chemokines	Exponential		Linear	Exponential		Linear
	Atlas 1.2 150 ng	Atlas 1.2 300 ng	Atlas 1.2 5 µg	GF211 150 ng	GF211 300 ng	GF211 5 µg
8 of 32 genes immobilised on both Atlas 1.2 and GF211 arrays						
IL1 $\alpha$	+	+	-	+	+	+
IL1 $\beta$	-	-	+	+	-	+
IL6	+	+	-	+	+	+
<b>IL8</b>	+	+	+	+	+	+
<b>MIP-1<math>\alpha</math></b>	+	+	+	+	+	+
<b>MIP-1<math>\beta</math></b>	+	+	+	+	+	+
<b>MCP-1</b>	+	+	+	+	+	+
MCP-4	-	-	-	-	-	+
3 of 36 genes immobilised only on Atlas 1.2 array						
TNF $\alpha$	+	+	+	NA	NA	NA
MIP-2 $\alpha$	+	+	+	NA	NA	NA
IP-10	+	+	-	NA	NA	NA
3 of 27 genes immobilised only on GF211 array						
MIP-1 $\delta$	NA	NA	NA	-	-	+
GRO1	NA	NA	NA	+	+	+
MCP-3	NA	NA	NA	+	+	+

Different RNA aliquots were tested by exponential (150 ng and 300 ng, SMART-PCR) and linear (5 µg, conventional array procedure) RNA/cDNA amplification. Thus created cDNA probes were assayed by two different filter based nylon arrays (Atlas 1.2, BD Biosciences Clontech and GF211, Research Genetics Invitrogen) and compared. (+) indicates signal intensity ratio (before/after LPS stimulation) of  $>2$ , whereas (-) indicates a ratio of  $\leq 2$ . Genes in bold were found to be expressed (ratio  $>2$ ) in all of the six comparative experiments performed. Twelve of 14 (86%) genes were expressed in a concordant way—that is, either in both of the exponentially or in the linearly amplified samples. Two genes (IL1 $\beta$  and MCP-4) were expressed in a discordant manner. NA (not available) indicates that the particular gene was not immobilised on the microarray membrane.

whereas 20 genes were detected in both testings (20/21 (95%)). Here, the highest ratio of expression values between duplicate experiments was 1.5. These data indicate a low variability of gene expression as measured by the array and led us to perform only one experiment for each probe in the following study.

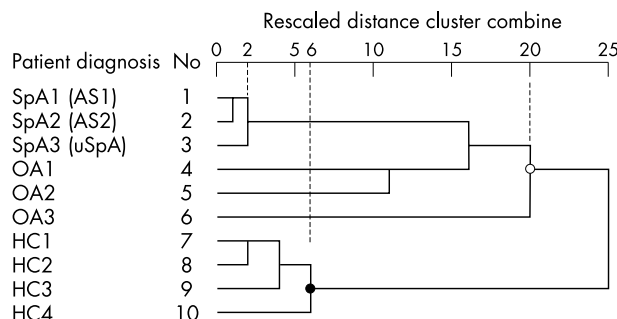
### Analysis of synovial tissue sample heterogeneity

Gene expression profiles of SpA synovial tissue samples were compared with two different control groups: (a) heterogeneous OA synovial tissue samples, and (b) PBMC from healthy donors, constituting a rather homogeneous cell population.

To study the impact of sample and/or patient heterogeneity on the microarray data, we analysed synovial tissue biopsy samples from three patients with SpA and three with OA as well as PBMC from four healthy controls. Organisation of gene expression data from synovial biopsies by hierarchical cluster analysis indicates the degree of similarity between samples as displayed by the dendrogram in fig 2. Thus, all six synovial samples were clustered under one node, whereas healthy PBMC were clustered separately under a different node. The degree of similarity was particularly high between the three patients with SpA, even higher than between the PBMC of four healthy controls. In contrast, the OA samples had a more heterogeneous expression profile.

Interestingly, histological analysis of synovial tissue samples from four of the six patients indicated a quite different profile. Both of the patients with SpA studied were histopathologically heterogeneous, with clear lining hyperplasia, hypervascularity, and inflammatory infiltration in SpA3 and a nearly normal synovium in SpA2 (table 3, fig 3). On the other hand, SpA2 resembled histologically OA2, whereas OA3 showed a different pattern of inflammatory cell infiltration, with mainly macrophages. Accordingly, the histopathological heterogeneity of the SpA as well as the OA samples contrasted with the homogeneity of the gene expression profiles within each group, suggesting that the

microarray data are not merely biased by sample heterogeneity within one disease group as assessed by histopathology.



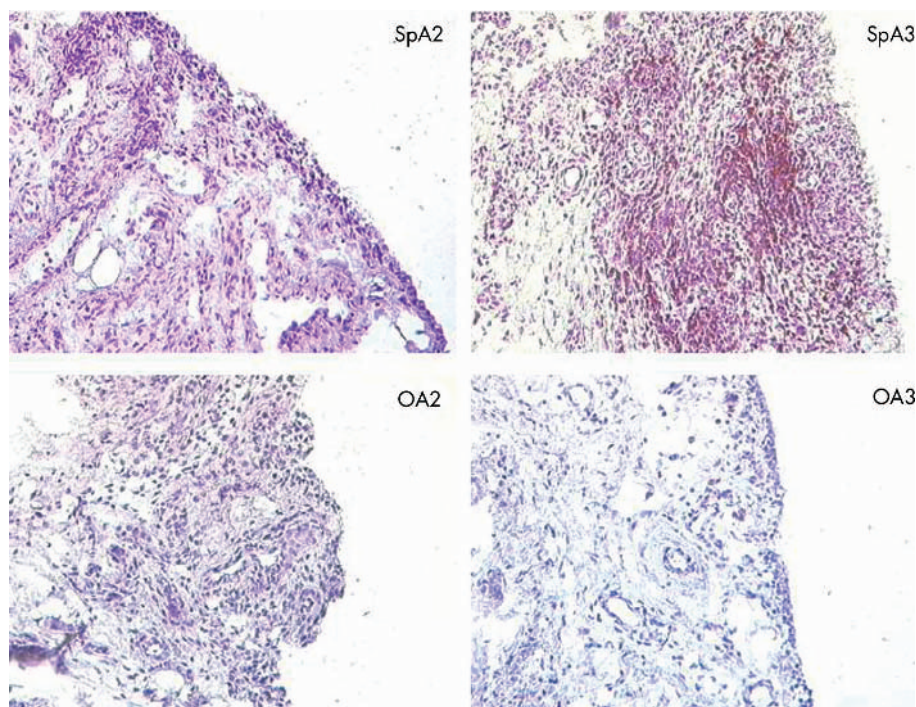
**Figure 2** Dendrogram depicting results of hierarchical cluster analysis from gene expression values of six synovial tissue samples and four normal PBMC. Hierarchical cluster analysis of gene expression data of all 10 samples as assessed by between-groups linkage analysis. The two columns on the left list the patient diagnoses and attribute numbers to the 10 samples used in this analysis. The horizontal scale ("rescaled distance cluster combine") uses arbitrary units from 0 to 25 in order to measure the similarity between the various samples that are clustered. The length of the multiple horizontal tree branches measured on the scale above reflects the degree of similarity between the various datasets. Each sample (horizontal line) is connected to the adjacent sample by a vertical line, which is extended by the thin dashed line and can be read off the scale. The three SpA synovial samples are classified as most similar among the three different groups (SpA, OA synovium, and normal PBMC) because their horizontal lines merge at 2 as compared with the four normal PBMC samples, which merge at 6. When all six synovial samples are taken together, they demonstrate the lowest degree of similarity with a vertical line merging at 20 on the scale above. Note the separation of samples (three SpA and three OA synovial tissue samples (SpA1–3, OA1–3) and four healthy PBMC (HC1–4)) into two different nodes (O) of the dendrogram. Homogeneity of gene expression pattern is highest among SpA samples (quantified as 2) and lowest among OA samples (quantified as 20).

Considering the high degree of global similarity of the microarray profiles of the three SpA synovial tissue samples, the influence of sample heterogeneity was further analysed by identifying genes that might be expressed by microarray assay in SpA synovium compared with OA synovium, on the one hand, and with PBMC on the other (table 4). When SpA and OA synovium were compared, 31 transcripts were found to have a difference of mean signal intensities higher than twice the background level (SpA *v* OA). Twenty nine of the 31 genes had a signal intensity ratio (SpA/OA) of more than 4. They were subjected to statistical testing. All data had equal variance. According to *t* test, 11 transcripts were found to be significantly up regulated in SpA when compared with OA (table 4, genes 1–11). Comparison of SpA with HC retrieved 86 genes with a difference twofold higher than the background signal (SpA *v* HC). Seventy nine of the 86 transcripts had a ratio (SpA/HC) of more than 4. Here, *t* test identified 18 mRNA transcripts which were significantly up regulated in SpA compared with HC (table 4, genes 1–6 and 12–23). Six genes were identical in both comparisons (6/11 (55%) and 6/18 (33%), respectively), indicating an important overlap of the microarray results when using either a heterogeneous sample such as OA synovium or a more homogeneous sample such as PBMC for comparison. Moreover, only one gene was expressed in OA synovium compared with PBMC, indicating that the increase in the level of gene expression in SpA synovium was not due solely to differences in the composition of the cell population. When average linkage analysis was used for clustering 1185 transcripts of the 10 arthritis samples, one homogeneous SpA cluster containing 16 genes was identified (fig 4), possibly suggesting similar function or regulation of transcripts present in this particular cluster.<sup>14</sup> Nine of those 16 genes were also identified by *t* test when comparing either SpA with OA or SpA with HC.

### Validation of the statistical approach by permutation of class labels

This test was performed in order to verify that the statistical analysis is specific.

The relatively small number of genes that were expressed in SpA synovium (11 and 18 out of 1185 genes for comparison with OA synovium and with PBMC, respectively) indicated that a commonly used approach for statistical analysis of microarray data, which was implemented in the present study, is quite stringent (difference of signal intensities higher than twofold of background, fourfold increase of signal intensities, Fmax/*t* test, Bonferroni adjustment). To further validate that the obtained results indeed reflect gene expression and are not random results due to multiple comparisons, analysis of three SpA and three OA datasets was tested by permutation of samples. As shown in table 5, three samples were randomly assigned to one group, independent of diagnosis, and subsequently tested for expressed genes compared with the three reciprocal samples. Of major interest, the comparison of three SpA samples with three OA samples yielded 11 expressed genes as reported above, whereas all 19 remaining possible combinations together yielded only one expressed gene. When the thresholds were chosen to be less stringent (difference of mean higher than onefold background signal, signal increase higher than 2, analysis of variance, Student's *t* test without Bonferroni adjustment—that is, a *p* value <0.05 was considered to be significant) the number of genes identified as being expressed in SpA compared with OA was 61 as compared with 11 previously. However, when permutation analysis was performed using the less stringent criteria, naturally the number of false positive genes in all other possible combinations of samples was seven as compared with one gene before.



**Figure 3** Microscopic picture of anti-CD3 staining on frozen synovial sections of two patients with SpA and two with OA (original magnification  $\times 160$ ). SpA3 shows clear signs of chronic inflammation with synovial lining hyperplasia, hypervascularity, and perivascular infiltration of lymphoid cells, whereas SpA2, OA2, and OA3 show only minimal signs of hypervascularity and inflammatory infiltration (see also table 3). Peroxidase staining against CD3 is strongly positive in SpA3, showing mainly a perivascular pattern, but is negative in all other patients.



**Table 3** Evaluation of various histological and immunohistochemical parameters of synovial tissue sections available from four of the six patients studied by microarray

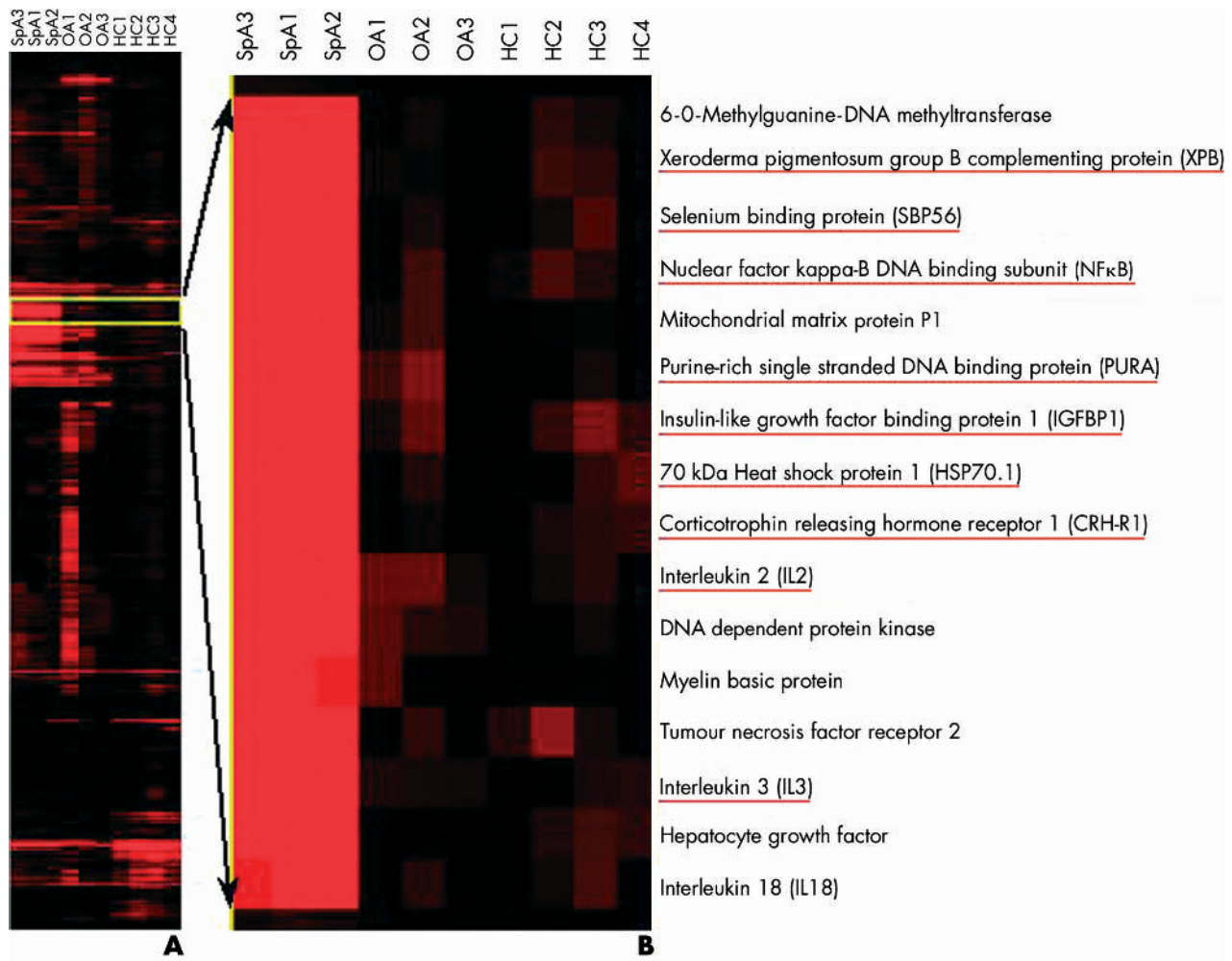
Patient	Lining (0-3)	Vasc (0-3)	Infiltr (0-3)	LA (0-1)	PC (0-1)	PMN (0-1)	CD3 (0-3)	CD20 (0-3)	CD68 (0-3)	CD138 (0-3)	CD146 (0-3)
SpA2	1	1	1	0	0	0	0	0	0	0	0
SpA3	2	2	2	0	0	0	2pp	1	1	1	2
OA2	1	1	1	0	0	0	0	0	0	0	1
OA3	1	1	0	0	0	0	0	0	2	0	2

Blinded scoring was performed according to criteria described previously.<sup>6</sup> Presence of lymphoid aggregates (LA), plasma cells (PC), and polymorphonuclear cells (PMN) was scored as either 1 (present) or 0 (absent). Hyperplasia of the lining layer (lining), degree of vascularisation (vasc), and infiltration (infiltr) as well as intensity of peroxidase staining for CD3, CD20, CD68, CD138, and CD146 were scored from 0 to 3 (pp, perivascular pattern).

## DISCUSSION

In this study we initially evaluated the technical reproducibility of the combined RNA amplification and microarray assay using both PBMC and synovial tissue biopsies. SMART-PCR used in combination with Atlas array has recently been shown to yield representative gene expression results when compared with the conventional linear procedure as published by a Clontech research group.<sup>15</sup> However, an independent group reported slight discrepancies in signal intensity ratios between SMART amplified probes and those

conventionally prepared.<sup>16</sup> Nevertheless, in that particular study a higher sensitivity of SMART amplified probes in agreement with our own data, at least when used in combination with Atlas array, was described. Our present analysis of variability due to different methods of RNA amplification confirms and extends these findings by indicating a high homology between the standard linear procedure and the exponential SMART-PCR in combination with two different array systems. For the Atlas array, SMART-PCR showed a higher sensitivity yielding three more



**Figure 4** Average linkage analysis of 1185 transcripts from all 10 samples as assayed by microarray. The left column (A) displays all 10 assayed samples with 1185 genes placed in rows. Each gene is represented by a single row of coloured boxes with correlation between level of expression and intensity of colour ranging from black (no expression) to bright red (high expression). The yellow lines in column A frame one identified homogeneous cluster of genes highly expressed in all three SpA samples. Column B is a magnified segment of the left column showing 16 genes within that particular cluster. Nine of the 16 genes, underlined in red, were also identified by *t* test in either SpA v OA or SpA v HC, and five of nine genes within that cluster were found to be among the genes expressed significantly in both the *t* test comparison groups.

**Table 4** Genes found to be significantly expressed as assessed by Student's *t* test and Bonferroni adjustment in SpA synovium when compared with OA synovium and with normal PBMC

Gene	Mean SpA	CI	Mean OA	CI	t Test (p<0.0017)	Mean HC	CI	t Test (p<0.00063)
1) <b>IL3</b>	<b>26518</b>	<b>19</b>	<b>2255</b>	<b>10</b>	<b>0.00001</b>	1580	56	<b>0.000002</b>
2) <b>CRH-R1</b>	<b>25797</b>	<b>26</b>	<b>578</b>	<b>82</b>	<b>0.0001</b>	2148	55	<b>0.00001</b>
3) <b>IGFBP1</b>	<b>27304</b>	<b>14</b>	<b>2586</b>	<b>86</b>	<b>0.0001</b>	3218	76	<b>0.00001</b>
4) <b>SBP56</b>	<b>32663</b>	<b>53</b>	<b>775</b>	<b>95</b>	<b>0.0004</b>	1748	85	<b>0.0001</b>
5) <b>HSP70.1</b>	<b>24038</b>	<b>48</b>	<b>1255</b>	<b>77</b>	<b>0.0006</b>	2523	67	<b>0.0002</b>
6) <b>GSR</b>	<b>23666</b>	<b>55</b>	<b>2298</b>	<b>91</b>	<b>0.002</b>	2599	53	<b>0.0002</b>
7) NMBR	27557	7	5428	84	0.0003	–	–	–
8) CALNLC	22126	41	845	70	0.0003	–	–	–
9) IL2	32674	47	4148	57	0.0005	–	–	–
10) NFκB	32733	68	1514	92	0.001	–	–	–
11) PURA	25590	47	3193	104	0.001	–	–	–
12) Pbx1	28048	9	–	–	–	1253	54	0.0000002
13) RENBP	25600	11	–	–	–	2966	31	0.0000004
14) IL7	20855	29	–	–	–	1014	40	0.00001
15) ZNF92	18870	21	–	–	–	1991	44	0.00001
16) MMP-17	19850	31	–	–	–	1432	50	0.00002
17) SAS	28320	42	–	–	–	2457	47	0.00004
18) BTEB2	28427	44	–	–	–	1600	62	0.00004
19) IFNβ	23095	40	–	–	–	2291	49	0.0001
20) IL10	22129	45	–	–	–	2955	38	0.0001
21) HSP40	20634	46	–	–	–	4721	54	0.0005
22) XPB	26712	73	–	–	–	1809	76	0.0005
23) MCSF	25502	68	–	–	–	2257	87	0.0005

Six of 11 (SpA v OA) and 18 (SpA v HC) genes (numbers 1–6, printed in bold) were found to be expressed in both comparisons, indicating a reasonably low heterogeneity of synovium among the various samples (genes 7–11 are identified only in the SpA v OA comparison, whereas genes from 12 to 23 are expressed only in the SpA v HC comparison). Mean values represent signal intensities as measured by phosphor imaging and AtlasImage software; CI, confidence interval at 95%; according to Bonferroni adjustment p value cut off points for the two comparisons refer to the number of genes fulfilling the criteria described in "Methods". Full name of transcripts and GenBank accession numbers: IL2, interleukin 2 (A14844); IL3, interleukin 3 (M14743); IL7, interleukin 7 (J04156); IL10, interleukin 10 (M57627); CRH-R1, corticotrophin releasing hormone receptor 1 (X72304); IGFBP1, insulin-like growth factor binding protein 1 (M31145); SBP56, selenium binding protein (U29091); HSP70.1, 70 kDa heat shock protein 1 (M11717); GSR, glutathione reductase (X15722); NMBR, neuromedin B receptor (M73482); CALNLC, nucleobindin (M96824); NFκB, nuclear factor κB DNA binding subunit (M58603); PURA, purine-rich single stranded DNA binding protein α (M96684); Pbx1, pre-B-cell leukaemia transcription factor-1 (M86546); RENBP, renin binding protein (D10232); ZNF92, zinc finger protein 91 (L11672); MMP-17, matrix metalloproteinase 17 (X89576); SAS, transmembrane 4 superfamily protein (U01160); BTEB2, basic transcription element binding protein 2 (D14520); IFNβ, interferon β (M28622); HSP40, heat shock protein 40 (D49547); XPB, xeroderma pigmentosum group B complementing protein (M31899); MCSF, macrophage-specific colony stimulating factor (M37435).

cytokines/chemokines than the linear method, which most likely is due to a higher specific activity of SMART amplified and <sup>32</sup>P labelled cDNA probes used for hybridisation. Another reason might be the use of a random prime labelling system for SMART amplified cDNA templates, further contributing to a higher sensitivity.

When the two different aliquots prepared by the SMART-PCR assay were compared (150 ng and 300 ng), there was complete concordance for all the cytokines and chemokines identified by the Atlas 1.2 array, indicating a low variability of the RNA amplification and this particular microarray assay. For the GF211 array, one cytokine (IL1β) was detected

**Table 5** Confirmation of results by permutation tests: statistical analysis of all 20 possible combinations (shuffling) of six synovial tissue samples studied by microarray

20 Possible combinations of samples		Number of significantly expressed genes			
		(A) Original stringent thresholds		(B) Less stringent thresholds	
Group 1	Group 2	Group 1 v group 2	Group 2 v group 1	Group 1 v group 2	Group 2 v group 1
<b>SpA1-SpA2-SpA3</b>	<b>OA1-OA2-OA3</b>	11	0	61 (+50)	2
SpA1-SpA2-OA1	SpA3-OA2-OA3	0	0		
SpA1-SpA2-OA2	OA1-SpA3-OA3	0	0		
SpA1-SpA2-OA3	OA1-OA2-SpA3	0	0	1	2
SpA1-OA1-SpA3	SpA2-OA2-OA3	0	0		
SpA1-OA2-SpA3	OA1-SpA2-OA3	0	0		
SpA1-OA3-SpA3	OA1-OA2-SpA2	0	0		
<b>OA1-SpA2-SpA3</b>	<b>SpA1-OA2-OA3</b>	1	0	2 (+1)	
OA2-SpA2-SpA3	OA1-SpA1-OA3	0	0		
OA3-SpA2-SpA3	OA1-OA2-SpA1	0	0		

Gene expression data of samples were rearranged and randomly assigned to two groups. Data were analysed by *t* test and Bonferroni adjustment. SpA1-3 v OA1-3 disclosed 11 significant genes, whereas only one of the 19 remaining combinations (SpA2, SpA3, OA1 v OA2, OA3, SpA1) disclosed one single significantly expressed gene. The same analysis was performed using less stringent cut off point criteria: difference of mean higher than onefold background signal, ratio higher than 2, analysis of variance, Student's *t* test without Bonferroni adjustment—that is, a p value <0.05 was considered to be significant. Accordingly, 50 additional (61 altogether) genes were found to be expressed in SpA compared with OA synovial tissue. However, the number of false positive genes was increased as well with six additional (seven altogether) genes in the various combinations as displayed in the table.



in the 5 µg Atlas array sample as well as in the 150 ng but not in the 300 ng sample, whereas MCP-4 was exclusively detected in the 5 µg sample. Nevertheless, four of eight cytokine/chemokine genes were detected in all of the six experiments performed across the two different array systems and six of eight cytokine/chemokine genes were detected in at least five of six experiments, indicating a low variability not only between the different RNA amplification techniques but also between the two different microarray systems. However, when the seven LPS induced non-cytokine/chemokine genes were analysed, reproducibility between exponential and linear amplification as well as between the two different array systems was low. This is probably because these genes are expressed on a threefold lower level than the cytokine/chemokine genes—a finding which supports the application of stringent thresholds in microarray data mining in order to avoid detection of false positive events. Assessing the reproducibility on synovial tissue samples of the microarray itself, we performed duplicate experiments on samples from two patients with SpA. The high correlation coefficients between the two duplicate datasets and little variation in signal intensity ratios of expressed genes calculated from duplicate data suggest a high experimental reproducibility. These data together with the aforementioned results indicate that the low technical variability of microarray applied to homogeneous cell populations<sup>5, 17</sup> is also applicable to synovial tissue samples.

A possible major problem of microarray experiments on synovial tissue biopsies is the sample heterogeneity, which might be another cause of experimental variability. However, no concise explorations have been published on this point. The results of our study indicate low variability due to tissue and/or patient heterogeneity when comparing the gene expression profiles of SpA synovium with OA synovium and with PBMC. Firstly, as expected, hierarchical cluster analysis could separate clearly gene expression profiles from synovium and PBMC. However, there was also a high degree of similarity between gene expression data from the three SpA synovial tissue samples, which were unambiguously separated from OA synovium. In fact, the similarity between SpA synovial tissue samples was even higher than between the PBMC samples. Secondly, the histopathological analysis showed clear differences between two SpA samples but similarity between one SpA and one OA sample, indicating that the gene expression clusters are more dependent on the disease than on the cellular composition of the tissues. Thirdly, the relative homogeneity of the SpA data is further indicated by the significant expression of 11 and 18 genes compared with heterogeneous tissue such as OA synovium and a rather homogeneous cell population such as PBMC, respectively. Six genes were shared between both analyses and we found just one expressed gene in OA synovium compared with PBMC, which emphasises that these results are not merely biased by heterogeneity of the tissues. It also demonstrates that RT-PCR and array technology only mirror actively up regulated genes, whereas immunohistochemistry and proteomics potentially can reflect the entire proteome. Finally, the relatively low number of samples studied here (three v three v four) probably leads to an underestimation of the total number of differentially expressed genes, though they indicate a high homology between patients within one disease group in the expression of various genes. Further studies will have to determine whether this is mainly biased by the particular phenotype of patients studied here (early disease, no DMARD treatment, HLA-B27 positivity, peripheral joint inflammation).

Apart from reproducibility of the technique and sample heterogeneity, the scientific validity of microarray data is strongly dependent on the statistical methods used for

analysis. Applied methods should be stringent in order to select a relatively small number of differentially expressed genes out of thousands but, more importantly, should also avoid false positive results.

This study did not compare different statistical methods, but attempted to evaluate the validity of a widely used approach: difference and ratio of signal intensities higher than predetermined cut off points, test for homogeneity of variance of data followed by *t* test with Bonferroni adjustment.<sup>17–21</sup>

The validity is also confirmed by recalculation of *p* values for all possible combinations of signal intensity data from the 10 samples. Moreover, permutation of class labels—that is, a complete reanalysis of datasets between SpA and OA groups bears evidence of a low rate of false positive results, thus supporting the specificity of the SpA v OA synovium comparison. Using the permutation analysis, we additionally demonstrate that the approach can be considered as feasible for microarray data analysis because less stringent threshold criteria inevitably lead to a higher number of false positive results.

In addition, commonly used average linkage analysis identified one homogeneous cluster of genes expressed in all three SpA samples. Nine of 16 genes present in this cluster were also identified by *t* test when SpA was compared with OA and SpA with HC (fig 4). Five of the nine transcripts were also found among the group of six shared *t* test genes expressed in both the statistical comparisons—a finding which helps to confirm our *t* test genes by another analytical tool.

In summary, our results indicate that the approach reported here can narrow down the spectrum of potentially expressed genes of more than 1000 to fewer than 20. The microarray technology is emerging as a useful screening tool for identification of molecular mediators in complex diseases. We have technically evaluated this tool for analysis of synovial tissue samples obtained by needle arthroscopy. It should clearly be emphasised that it was not the aim of our study to identify and confirm pathogenically relevant genes involved in SpA pathogenesis but to explore and validate the microarray technology for studying synovial tissue samples from patients with arthritis.

Our data suggest that microarray technology can be used reliably to detect molecular differences between SpA and OA synovium. Moreover, future studies applying molecular profiling, combined with confirmatory techniques such as RT-PCR or immunohistochemistry, to a larger number of patients with distinct phenotypical features might help to identify distinct gene expression patterns (“signatures”) and thus diagnostic and prognostic subgroups of SpA.<sup>2, 3</sup> In this context, predicting the response to particular forms of newly emerging, expensive treatments could indeed be useful. In conclusion, microarray on synovial tissue biopsy samples might be a valuable tool for enhancing the search for clues to SpA pathogenesis.

## ACKNOWLEDGEMENTS

The authors thank Mrs Jenny Vermeersch for excellent technical assistance.

This work was supported by the Nora Eccles Treadwell Foundation. Dr M Rihl and Dr JG Kuipers are supported by the Deutsche Forschungsgemeinschaft DFG (RI 1119/1-1 and KU 1182/1-3). Dr M Rihl is also supported by the Rheumatology Competence Network, Berlin. Dr D Baeten is supported by the Fund for Scientific Research-Flanders (FWO-Vlaanderen).

## Authors' affiliations

M Rihl, J G Kuipers, H Zeidler, Hannover Medical School, Department of Rheumatology, Hannover, Germany

**D Baeten, F De Keyser, E M Veys**, Ghent University Hospital, Department of Rheumatology, Ghent, Belgium  
**M Rihl, N Seta, J Gu, D T Y Yu**, University of California Los Angeles, UCLA School of Medicine, Division of Rheumatology, Los Angeles, CA, USA

## REFERENCES

- 1 **Alizadeh AA**, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**:503–11.
- 2 **Rosenwald A**, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, *et al*. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;**346**:1937–47.
- 3 **Van De Vijver MJ**, He YD, Van't Veer LJ, Dai H, Hart AAM, Voskuil DW, *et al*. A gene expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**:1999–2009.
- 4 **Gu J**, Märker-Hermann E, Baeten D, Tsai WC, Gladman D, Xiong M, *et al*. A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthritis patients. *Rheumatology (Oxford)* 2002;**41**:759–66.
- 5 **Gu J**, Rihl M, Märker-Hermann E, Baeten D, Kuipers J, Song YW, *et al*. Clues to pathogenesis of spondyloarthritis derived from synovial-fluid-mononuclear-cell gene expression profiles. *J Rheumatol* 2002;**29**:2159–64.
- 6 **Baeten D**, Demetter P, Cuvelier C, Van Den Bosch F, Kruithof E, Van Damme N, *et al*. Comparative study of the synovial histology in rheumatoid arthritis, spondyloarthritis, and osteoarthritis: influence of disease duration and activity. *Ann Rheum Dis* 2000;**59**:945–53.
- 7 **Heller RA**, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, *et al*. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci USA* 1997;**94**:2150–5.
- 8 **Zanders ED**, Goulden MG, Kennedy TC, Kempell KE. Analysis of immune system gene expression in small rheumatoid arthritis biopsies using a combination of subtractive hybridization and high-density cDNA arrays. *J Immunol Methods* 2000;**233**:131–40.
- 9 **Firestein GS**, Pisetsky DS. DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology. *Arthritis Rheum* 2002;**46**:859–61.
- 10 **Guha M**, Mackman N. LPS induction of gene expression in human monocytes. *Cell Signal* 2001;**13**:85–94.
- 11 **Baeten D**, Van den Bosch F, Elewaut D, Stuer A, Veys EM, De Keyser F. Needle arthroscopy of the knee with synovial biopsy sampling: technical experience in 150 patients. *Clin Rheumatol* 1999;**18**:434–41.
- 12 **Dougados M**, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, *et al*. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;**34**:1218–30.
- 13 **Chenchik A**, Zhu YY, Diatchenko L, Li R, Hill J, Siebert PD. Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. In: Siebert PD, Larrick JW, eds. *Gene cloning and analysis by RT-PCR*. Biotechniques Books. Natick, MA, USA: Eaton Publishing, 1998:305–19.
- 14 **Eisen MB**, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;**95**:14863–8.
- 15 **Zhumabayeva B**, Diatchenko L, Chenchik A, Siebert PD. Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques* 2001;**30**:158–63.
- 16 **Puskas LG**, Zvara A, Hackler L Jr, Van Hummelen P. RNA amplification results in reproducible microarray data with slight ratio bias. *Biotechniques* 2002;**32**:1330–4, 1336, 1338, 1340.
- 17 **Neumann E**, Kullmann F, Judex M, Justen HP, Wessinghage D, Gay S, *et al*. Identification of differentially expressed genes in rheumatoid arthritis by a combination of complementary DNA array and RNA arbitrarily primed-polymerase chain reaction. *Arthritis Rheum* 2002;**46**:52–63.
- 18 **Baggerly KA**, Coombes KR, Hess KR, Stivers DN, Abruzzo LV, Zhang W. Identifying differentially expressed genes in cDNA microarray experiments. *J Comput Biol* 2001;**8**:639–59.
- 19 **Haslett JN**, Sanoudou D, Kho AT, Bennett RR, Greenberg SA, Kohane IS, *et al*. Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. *Proc Natl Acad Sci USA* 2002;**99**:15000–5.
- 20 **Zou TT**, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y, *et al*. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. *Oncogene* 2002;**21**:4855–62.
- 21 **Pan W**. A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* 2002;**18**:546–54.