ORIGINAL ARTICLE

Expression of a pathogen-response program in peripheral blood cells defines a subgroup of Rheumatoid Arthritis patients

TCTM van der Pouw Kraan^{1,5}, LGM van Baarsen^{1,2,5}, CA Wijbrandts³, AE Voskuyl⁴, F Rustenburg^{1,2}, JM Baggen¹, BAC Dijkmans⁴, PP Tak³ and CL Verweij^{1,2}

¹Department of Molecular Cell Biology and Immunology, VU University Medical Center, Amsterdam, The Netherlands; ²Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands; ³Division of Clinical Immunology and Rheumatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands and ⁴Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands

Rheumatoid arthritis (RA) is a heterogeneous disease with unknown etiology. Here we aimed to distinguish RA subtypes based on peripheral blood (PB) gene expression profiles in comparison with a pathogen-response transcriptional program. PB was obtained from 35 RA patients and 15 healthy individuals. For expression profiling we used DNA microarrays. A combined cluster analysis of RA and control samples together with samples from a viral infection model revealed that the gene expression profile of a subgroup of RA patients (RA^A) was reminiscent to that of poxvirus-infected macaques. Statistical analysis, followed by Gene Ontology analysis of the RA^A patients confirmed that these patients form a distinct group, with activation of several host defense mechanisms that resemble a common host-pathogen response. Analysis of the promoter region of genes that were overexpressed in the RA^A patients, revealed an enrichment of transcription factor binding sites for NFκB and interferon-activated transcription factors. Moreover, this subgroup of RA patients expressed significantly increased titers of anti-cyclic citrullinated peptide antibodies. We conclude that activation of a host-pathogen response defines a subgroup of RA patients characterized by increased autoreactivity against citrullinated proteins.

Genes and Immunity (2008) 9, 16–22; doi:10.1038/sj.gene.6364438; published online 11 October 2007

Keywords: profiling; genomics; interferon; transcription factors; gene ontology

Introduction

Rheumatoid arthritis (RA) is a systemic autoimmune disease characterized by chronic inflammation of the joints. The differential expression of the disease is reflected by the marked differences in the abundance and distribution of T and B cells in the rheumatoid synovium, ranging from a scarce and diffuse distribution to the abundant presence of mononuclear cells, which are organized in germinal center-like structures.^{1,2} The heterogeneous character of RA is further illustrated by the differential responsiveness to treatment^{3–5} and the presence of distinct autoantibody specificities, like rheumatoid factor and anti-citrullinated peptide antibodies (ACPA) in the serum.^{6,7}

The existence of such molecular heterogeneity in RA synovium fits a model proposed by Firestein and

Zvaifler,⁸ who suggested two independent processes, an immune-mediated and a stromal cell-driven form, which might drive destruction of bone and cartilage.

Development of different types of RA may result from specific combinations of environmental factor(s) and a varying polygenic background. Findings from genetically identical twins, where the concordance rate is far less than complete, are indicative for a major role of an environmental factor in the risk of developing RA.^{9,10} From animal models it is clear that an exogenous challenge, like infectious agents may induce arthritis. In man, microbial infections are known to contribute to arthritis.^{11,12} Combined analysis of the Epstein-Barr virus (EBV-), CMV- and Parvo virus B19-data demonstrated that virus infections are far more common in RA or psoriatic arthritis than in reactive arthritis.¹³ Despite this knowledge the environmental factor(s) that contribute to RA remain to be determined.

Upon infection with a pathogen, cells undergo a marked reprogramming of their transcriptome, providing a direct link between infection and the transcriptional profile of the host. Surprisingly, different pathogens induce a remarkable overlap of transcriptional changes in the host, leading to the concept of a common pathogen response.¹⁴ Therefore, analysis of the host transcriptome overcomes the problematic detection of multiple distinct

Correspondence: Dr TCTM van der Pouw Kraan, Department of Molecular and Cellular Biology and Immunology, VU University Medical Center, C262, PO Box 7057, Amsterdam 1007 MB, The Netherlands.

E-mail: t.vanderpouwkraan@vumc.nl

⁵These authors contributed equally to this study.

Received 24 July 2007; revised 5 September 2007; accepted 6 September 2007; published online 11 October 2007

microorganisms, in case no single microorganism but a variety of microorganisms is involved. Large-scale gene expression profiling using DNA microarray technology in blood cells from patients with RA and healthy controls would enable us to detect a molecular portrait representing the contributions and interactions of numerous distinct cells and diverse factors that may be associated with disease. Consequently, if there is a pathogen involved in RA, a systematic analysis of the gene expression profile should reveal similarities with a pathogen-response program.

We previously provided evidence for an activated IFN type I response gene expression program in a subset of patients with RA.¹⁵ This could be indicative for a viral contribution. Here we created a framework for evaluating the relationship between a viral response program and RA in more detail. Therefore, we compared the gene expression profiles of RA patients with the profiles of an *in vivo* virus infection model that was examined on the same microarray platform, allowing a direct comparison.

Results

Identification of a virus response signature in a nonhuman primate model

In order to create a framework, which allows a direct comparison of the RA gene expression profile with that of a physiological viral response program we made use of available *in vivo* data on smallpox virus-infected macaques.¹⁶ This nonhuman primate model for smallpox infection provides a detailed view of the *in vivo* host transcriptional program in peripheral blood (PB) after infection. Because the microarrays used in this study were of the same platform as the ones used in the present study, we could make a direct comparison. First, we extracted microarray data from nine macaques before, and 2 or 3 days post infection from the public access website of the Stanford Microarray Database (SMD).¹⁷ Next, we identified the genes that significantly changed in expression in PB cells of infected macaques by Significance Analysis of Microarrays (SAM).¹⁸ At a false discovery rate of 5%, 911 genes showed a significant change in expression after smallpox infection.

The PB gene expression profile of a subgroup of patients with RA is reminiscent of a viral-response program

We previously analyzed PB gene expression profiles from RA patients, and compared their profiles with healthy controls using microarrays with a complexity of ~26 K unique genes (total 43 K elements).¹⁵

Expression of the viral-response genes was examined in a combined data set including gene expression data from all RA patients and healthy controls, uninfected and infected macaques. Two-way hierarchical clustering clearly separated the profiles of the macaques before and after viral infection (Figure 1), whereas almost all of the healthy controls clustered in the same group as the uninfected animals. Most interestingly, a total of 23 of the 35 RA patients clustered together with the infected macaques.

The subgroup of RA patients that clustered together with the virus-infected primates, was designated the RA^A subgroup. The patients that co-clustered with the

Figure 1 Subgroup of RA patients clusters together with smallpox-virus-infected macaques. Two-way hierarchical clustering was performed on genes that are significantly changed in expression after smallpox infection of macaques.¹⁶ Data from the smallpox study were combined with expression data from RA patients and healthy controls (see Materials and methods).

noninfected primates and healthy individuals were referred to as the RA^B subgroup (Table 1).

A direct comparison between the RA^A and RA^B subgroup by SAM revealed a prominent downregulation of genes in the RA^A group (2003 in total) representing processes involved in nucleotide metabolism, including pre-mRNA processing, mRNA transcription and mRNA splicing (analyzed in the Protein ANalysis THrough Evolutionary Relationships (PANTHER) database, $P = 5 \times 10^{-24}$ data not shown). In the smallpox-infected macaques nucleotide metabolism was also highly significant downregulated ($P = 1.3 \times 10^{-11}$ data not shown).

In our further analysis we compared the different RA subgroups to the healthy controls to obtain a view of the biological aberrations compared to a nondisease state. To identify genes that show an abnormal expression in the RA^A subgroup we performed a statistical analysis by SAM, which revealed that 663 genes were upregulated in RA^A patients, while 1246 genes were downregulated compared to the healthy controls. To interpret the biological significance of the list of differentially expressed genes we performed a Gene Ontology analysis in the PANTHER database (Table 2). RA^A patients showed evidence of 21 upregulated processes, including immunity and defense, interferon-mediated immunity, cytokine/chemokine-mediated immunity, Cell proliferation and differentiation, (anti-) apoptotic processes and blood clotting. Among the six downregulated processes were several processes that indicated reduced transcriptional activity, cell cycling and apoptotic activity (in contrast to the upregulated genes these were mainly genes with pro-apoptotic activity). Within the list of significantly

downregulated genes in the RA^A patient group were several genes suggesting a reduced cytotoxic response mediated by $\gamma\delta$ T cells, including CCL5/RANTES,

Table 1 Patient characteristics

	RA ^A N = 23	RA ^B N = 12
Age (mean)	49	49
Sex: female, male	17 F, 6 M	8 F, 4 M
<i>Disease characteristics</i>		
ESR (mean)	31	29
Rheumatoid factor titer (mean)	129 ^a	95
ACPA titer (mean)	1372 ^a	398 ^b
Disease duration in months (mean)	57	132
Erosions (N)	14	11
DAS28 (mean)	5.4	5.3
<i>Medication</i>		
MTX dose (mean)	14	18
Prednisone $\leq$ 10 mg per day (N)	6	0

Patients were subdivided in RA^A and RA^B, based on two-way co-clustering with virus-infected macaques, as presented in Figure 1.

^aAvailable for 21 of 23 patients.

^bSignificantly different between RA^A and RA^B ($P = 0.03$).

TARP/TCRgammaV9, perforin, NKG2D/KLRK1 and granzyme B, K and M.

The same analysis for the RA^B patients revealed only 375 upregulated genes and 179 downregulated genes compared to healthy controls (Table 3). Accordingly, there were fewer processes that discriminated RA^B patients from controls (increased cell adhesion, cell communication, signal transduction, and macrophage-mediated immunity) (Table 2).

To substantiate these findings we also performed gene ontology analysis on common pathogen-response genes, that is, a set of genes that is commonly induced by viral

Table 3 Overview of statistical analysis by SAM of the different RA subgroups (as defined in Figure 1), compared to healthy individuals

	RA N = 35	RA ^A N = 23	RA ^B N = 12	IFN ^{high} N = 20
No. upregulated genes	259	663	375	484
No. downregulated genes	318	1246	179	229
Total number of genes	577	1909	554	713

Our previous analysis on the IFN^{high} RA group¹⁵ is included for comparison. The number of genes is indicated after a two-class unpaired SAM analysis using a false discovery rate of 5%.

Table 2 Gene ontology of differentially expressed genes compared to controls

	CPR ^a	All patients	RA ^A	RA ^B
<i>Upregulated biological processes</i>				
Immunity and defense	x	1.61E-11	1.32E-13	1.52E-03
Interferon-mediated immunity	x	1.09E-06	3.54E-04	NS
Ligand-mediated signaling	x	3.52E-03	5.38E-04	NS
Cytokine/chemokine-mediated immunity	x	1.63E-03	3.50E-03	NS
Macrophage-mediated immunity	x	6.83E-05	1.13E-02	1.17E-02
Transport		4.82E-04	1.29E-02	4.87E-02
Lipid and fatty acid transport		1.54E-04	4.85E-02	NS
Signal transduction	x	NS	4.44E-06	4.24E-03
Cell proliferation and differentiation	x	NS	8.97E-06	NS
Developmental processes	x	NS	5.35E-05	NS
Blood clotting		NS	1.06E-04	NS
Cell communication	x	NS	2.06E-03	4.32E-02
Intracellular signaling cascade	x	NS	5.97E-03	NS
Inhibition of apoptosis	x	NS	6.35E-03	NS
Cell adhesion		NS	8.08E-03	1.18E-03
Cell structure and motility		NS	8.68E-03	NS
Cell surface receptor mediated signal transduction	x	NS	1.05E-02	NS
Mesoderm development	x	NS	1.69E-02	NS
Cell cycle		NS	1.81E-02	NS
Cell cycle control	x	NS	3.11E-02	NS
Apoptosis	x	NS	3.71E-02	NS
<i>Downregulated biological processes</i>				
Nucleoside, nucleotide and nucleic acid metabolism	x	NS	4.75E-09	NS
mRNA transcription	x	NS	1.16E-04	NS
mRNA transcription regulation	x	NS	3.66E-03	NS
Pre-mRNA processing		NS	5.71E-03	NS
Cell cycle		NS	1.91E-02	NS
Apoptosis	x	NS	4.33E-02	NS

Abbreviations: CPR, common pathogen response; NS, not significant.

Overview of the biological processes represented by the genes that were significantly different expressed in RA patient subgroups (Figure 1), compared to healthy individuals. Processes were identified by Panther Gene Ontology analysis. Processes with $P < 0.05$ (after a Bonferroni correction) were selected as significant.

^aBiological processes that were also represented by the common pathogen response (CPR)¹⁴ are indicated (x).

and bacterial triggers through activation of different TLRs.¹⁴ In comparison with the processes represented by the RA patients, we observed that in the RA^A group, 18 of the identified processes were shared with the common pathogen response, while in the RA^B group only four processes were identified in the common pathogen response (Table 2).

Finally, to gain more insight in the factors that could have induced the increased expression of genes in the RA^A patients we performed an analysis on transcription factor binding sites (TFBS) that are overrepresented in the genes, which were upregulated in the RA^A patients compared to the healthy individuals (by rVISTA¹⁹ at <http://genome.lbl.gov/vista/index.shtml>, data not shown). The most significantly enriched regulatory elements in the 500bp upstream regions of the over-expressed genes were ICSBP/IRF8, IRF-1, ISRE and NFκB binding sites. This indicates that agents that activate the NFκB-pathway and induce IFN type I production, are the most important contributors to the expression signature of RA^A patients. Because these pathways are typical for TLR activation, we next performed this analysis on the common pathogen-response gene set.¹⁴ The search for TFBS overrepresentation in the promoter regions of the common pathogen-response genes revealed an almost identical set of most significant regulatory regions, namely NFκB, NFκB65, IRF-1 and ICSBP/IRF8 binding sites. The combined results suggest that activation of several TLRs, which mediate IFN type I induction and NFκB activation signals, such as TLR3, TLR4, TLR7 and TLR9, could potentially be responsible for the altered transcriptome in RA^A patients.²⁰

The most significantly overrepresented TFBS in genes with a low expression in RA^A patients were the binding sites for E2F1DP1 and E2F1DP2, which are involved in regulation of the cell cycle, DNA synthesis and apoptosis. The reduced activity of genes harboring these TFBS corresponds with the reduced activity of the pathways in RA^A patients identified by gene ontology analysis.

Relation to the IFN-type high and low RA subgroups

We previously performed an RA subclassification based on the expression of IFN-induced genes.¹⁵ To obtain an insight in the differences between IFN- and viral-determined subclassification, we calculated the average expression of 43 correlated IFN-induced genes, which we described previously¹⁵ for the different RA subgroups (Figure 2a). As anticipated the RA^A group showed a significantly higher expression of the IFN-response genes, when compared to both the RA^B group and the controls, corresponding with the Gene Ontology analysis. Although the RA^A group as a whole displayed an enhanced IFN type I response, six of the IFN^{low} patients were also included in the RA^A group, while four of the IFN^{high} patients were not represented in the RA^A group. This indicates that the subclassification of RA patients into RA^A and RA^B is not solely based on the expression of IFN-induced genes, but that the entire pathogen-response program contributes to the subclassification. Statistical comparison of all groups to the healthy individuals indicated that the RA^A group was the most different from healthy controls (Table 3).

Figure 2 (a) IFN-response gene expression is enhanced in RA patients that clustered together with infected macaques (RA^A). The average expression ratios of 43 IFN-response genes, that were identified in RA patients previously,¹⁵ are shown for the two patient groups and controls. (b) ACPA levels are significantly higher in the RA patients that clustered together with infected macaques (RA^A). Data indicate the levels of antibodies to citrullinated proteins measured in the sera of RA patients.

Clinical relevance of the patient stratification

Next, the question arose whether the molecular stratification of RA was associated with clinical differences. To address this possibility, the relationship between the RA subtypes and clinical parameters, summarized in Table 1, was determined. Most strikingly, this analysis revealed that the two groups differed in ACPA levels. The RA^A group showed higher levels of ACPA ($P=0.03$, Figure 2b) than the RA^B group.

Discussion

The combined cluster analysis of RA patients with smallpox virus-infected macaques revealed a striking heterogeneity between RA patients indicated by the co-clustering of a subset of patients (RA^A subgroup) with the virus-infected primates.

The combined analysis of different datasets has already demonstrated its significance by the discovery of a common fibroblast serum response signature that is represented in tumor samples and predicts cancer progression.²¹ Our current analysis further indicates that combining different datasets powers disease subclassification and leads to the identification of an immune defense gene expression profile in PB cells of a subgroup of RA patients.

Infectious agents have long been considered as possible triggers for autoimmune responses.^{22–24} The higher load of EBV DNA identified in PB cells from RA patients²⁵ would support this suggestion. Olsen *et al.* proposed a ‘viral signature’ as a typical feature of early RA, based on the observation that the transcriptome of PB cells from early RA patients partly overlaps with

genes induced by Influenza.²⁶ Although our study was not designed to compare early with longstanding RA, the disease duration was shorter in the RA^A compared to the RA^B group (57 vs 132 months), but did not reach statistical significance ($P=0.099$). Further support for an infectious agent is delivered by the remarkable resemblance of the gene expression signature of the RA^A group with the transcriptional program of the common pathogen response.¹⁴ In addition, the most important TFBS in the upstream regions of the genes expressed by the RA^A patients correspond with the regulatory regions used by genes that comprise the common pathogen response; the binding sites for ICSB/IRF8, IRF1 and NF κ B suggests the involvement of TLR activation in the RA^A group. Interestingly, in genes with a low expression in RA^A patients the most significant TFBS were E2F1DP1 and E2F1DP2. These factors are normally able to inhibit the NF κ B survival signal.²⁷ A reduced activity of these transcription factors would imply that an uninhibited survival signal could be delivered by NF κ B in the RA^A patients.

Another resemblance with a viral infection is the observation that the RA^A patient group showed signs of a reduced cytotoxic response mediated by $\gamma\delta$ T cells, suggested by the reduced expression of CCL5/RANTES, TARP/TCR γ 9, perforin, NKG2D/KLRK1 and granzyme B, K and M. This finding is in line with the previously identified low levels of $\gamma\delta$ T cells in PB of RA patients.²⁸

Furthermore, a lack of expression of granzyme A, RANTES/CCL5 and TCRB, during hepatitis C virus infection in nonhuman primates has shown to be associated with the inability to clear the infection.²⁹ Therefore, RA^A patients may be more prone to viral persistence because of the lack of effective virus clearance. In support of this, a reduced antiviral response, indicated by the decreased T-cell precursor frequencies to EBV gp110 in established RA has been reported.³⁰ Although we were not able to directly compare our data to a chronic viral infection, several characteristics of acute viral infections have also been observed in chronic viral infections. For instance, patients with chronic hepatitis C virus infection show reduced numbers of circulating $\gamma\delta$ T cells as well.³¹ In addition, the type I IFN-response expression profile, is also convincingly present in PB mononuclear cells (PBMC) from patients with a chronic hepatitis C infection.³²

In summary, the RA^A patients show several characteristics of viral infections, (a) the increased type I IFN signature, (b) a reduced $\gamma\delta$ gene expression signature, (c) reduced transcript levels of HLA class II molecules (data not shown) and (d) reduced transcriptional activity.^{31,33,34}

The fact that the RA^B patients were not very different from controls in immunity-related transcript levels, may suggest that local inflammation in the joints is less severe, and therefore not as profoundly reflected in PB cells.

In these patients a more aggressive type of fibroblast-like synoviocytes that are relatively immune-independent could possibly drive the disease, as we suggested earlier.³⁵

Although our current analysis supports the suggestion that an infectious agent is involved in the development of arthritis, the study design does not allow for a firm

conclusion to be drawn. Formally we cannot exclude the possibility that the pathogen-response signature may have been triggered by endogenous danger signals that are released during inflammation, independent of infectious agents, such as heat shock proteins and DNA or RNA from damaged or dying cells.³⁶ Recently, defective degradation of endogenous DNA from apoptotic cells has been identified as a trigger for arthritis and the generation ACPA in an animal model, leading to increased IFN type I and TNF α release by an unknown mechanism.³⁷ We found evidence for diminished apoptosis in the RA^A group, which suggests that a defect in apoptosis may contribute to the autoimmune process, as has been suggested before.³⁸ The increased ACPA levels in the RA^A patients might therefore be related to defects in the apoptotic pathway. Of interest is the recent observation that approximately half of the RA patients show antibody reactivity to citrullinated EBV-derived peptide, which correlates with ACPA levels.³⁹ These findings might suggest that a viral agent could also play a role in the induction of ACPA antibodies.

We are now facing the challenge of unraveling the etiology of the pathogen-response signature in RA.

Materials and methods

Patients and controls

PB was obtained in PAXgene RNA isolation tubes (PreAnalytix, GmbH, Germany) from 35 patients, as described previously.¹⁵ From all 35 patients, 25 used methotrexate, and 10 patients were DMARD naive. The control group consisted of 15 age- and sex-matched healthy individuals, as described.¹⁵ All patients and controls gave their informed consent and the Medical Ethics Committees from AMC and VUmc approved the study protocol.

ACPA measurements

ACPA titers were measured by ELISA (Euro Diagnostica No. RA-96RT, Arnhem, the Netherlands) with a lower detection limit of <1 kAU ml⁻¹.

Sample preparation, labeling and hybridization

Total RNA was isolated from PB using the PAXgene RNA isolation kit according to the manufacturers' instructions including a DNase (Qiagen, Venlo, The Netherlands) step to remove genomic DNA. One microgram of total RNA was linearly amplified using the MessageAmp kit from Ambion. An amplified common reference sample was used for all array hybridizations, consisting of a mixture of mRNAs isolated from 11 different cell lines supplemented with RNA from synovial tissue, synovial fibroblasts and activated PBMCs.

Five micrograms of amplified RNA was labeled with aminoallyl-dUTP during cDNA synthesis by reverse transcriptase, followed by chemical coupling of the aminoallyl group to Cy3 or Cy5 for the experimental and reference samples, respectively. The Cy3- and Cy5-labeled cDNA transcripts were hybridized together on 43 K human cDNA microarrays (Stanford University) at 65 °C for 14–18 h. After washing, arrays were scanned using a G2505B microarray scanner (Agilent Technologies, Amstelveen, the Netherlands). Feature extraction

analysis was performed using GenepixPro version 3.0.6.90 (Axon instruments, Union City, CA, USA).

Data filtering and analysis

Data were stored and pre-analyzed in the SMD¹⁷ at <http://genome-www.stanford.edu/microarray>. Data are expressed as log₂ ratios of fluorescence intensities of the experimental and the common reference sample. Intensity-dependent normalization using local estimation (Loess) was performed separately on each sector of the array. In addition, 'scale' normalization was applied, which makes the data more comparable across different arrays. Spots were included in the analysis when in at least 80% of the microarrays a reliable data point was obtained for that element (defined by a regression correlation coefficient $R > 0.6$ assuring a linear fit between the Cy3 and Cy5 pixel intensities, and a signal intensity of 2.5 times the local background for both Cy3 and Cy5). The use of a common reference sample allows comparison of the relative expression levels across all samples. Therefore, the expression levels (as log₂ ratio's) were median centered, that is, each spotted element was expressed relative to the median expression level of that element across all samples. We corrected for array batch differences by applying single value decomposition.⁴⁰ Genes represented more than once on the microarrays were averaged in SMD from sequences with the same UniGene identifier.

Application of published data from the nonhuman primate model of smallpox

A genome-wide investigation of poxvirus infection in macaques was performed by Rubins *et al.*¹⁶ In this study, PBMCs were obtained from macaques before and after infection with smallpox virus. Gene expression data were obtained using the same microarray platform and data were filtered as described above. The genes in the smallpox dataset were centered separately to remove the bias of a different reference sample. Two-way hierarchical clustering⁴¹ of macaque and human samples was used to subclassify RA patients and visualize the correlation of co-expressed genes in Treeview (available at <http://rana.lbl.gov/EisenSoftware.htm>).

Statistical analysis

Statistical analysis on microarray data was performed using SAM.¹⁸

For an interpretation of the biological processes that are represented by the genes that show a significantly different level of expression in RA patients compared to the controls, we used Gene Ontology analysis using the PANTHER Classification System at <http://PANTHER.appliedbiosystem.com>.⁴² PANTHER uses the binomial statistics tool to compare our gene list to a reference list (NCBI: Homo sapiens genes) to determine the statistically significant overrepresentation of functional groups of genes. A Bonferroni correction was applied to adjust for multiple testing.

Acknowledgements

We would like to thank the following authors for making their microarray data publicly available: RG Jenner (Cambridge, USA) and KH Rubins (Stanford, USA). This

study was supported in part by the EURO-RA Marie Curie Trainings network, the European Community's FP6 funding (Autocure), the Netherlands Organization for Health Research and Development (ZonMw) in assignment of the Netherlands Organization for Scientific Research (NWO, grant number 945-02-029), the Innovation Oriented research Program (IOP) on Genomics and Centre for Medical Systems Biology (a center of excellence approved by the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research). This publication reflects only the authors' views. The European Community is not liable for any use that may be made of the information herein.

Conflicts of interest

We declare that there are no conflicts of interest or financial interests that are relevant to the research.

References

- 1 Takemura S, Braun A, Crowson C, Kurtin PJ, Cofield RH, O'Fallon WM *et al.* Lymphoid neogenesis in rheumatoid synovitis. *J Immunol* 2001; **167**: 1072–1080.
- 2 Timmer TCG, Baltus B, Vondenhoff M, Huizinga TWJ, Tak PP, Verweij CL *et al.* Inflammation and ectopic lymphoid structures in Rheumatoid Arthritis synovial tissues dissected by genomics technology. Identification of the IL-7 signaling pathway in tissues with lymphoid neogenesis. *Arthritis Rheum* 2007; **56**: 2492–2502.
- 3 Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR *et al.* Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-tumor necrosis factor trial in rheumatoid arthritis with Concomitant Therapy Study Group. *N Engl J Med* 2000; **343**: 1594–1602.
- 4 Kremer JM, Westhovens R, Leon M, Di Giorgio E, Alten R, Steinfeld S *et al.* Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. *N Engl J Med* 2003; **349**: 1907–1915.
- 5 Edwards JC, Szczepanski L, Szechinski J, Filipowicz-Sosnowska A, Emery P, Close DR *et al.* Efficacy of B-cell-targeted therapy with rituximab in patients with rheumatoid arthritis. *N Engl J Med* 2004; **350**: 2572–2581.
- 6 Zendman AJ, van Venrooij WJ, Pruijn GJ. Use and significance of anti-CCP autoantibodies in rheumatoid arthritis. *Rheumatology (Oxford)* 2006; **45**: 20–25.
- 7 Hueber W, Kidd BA, Tomooka BH, Lee BJ, Bruce B, Fries JF *et al.* Antigen microarray profiling of autoantibodies in rheumatoid arthritis. *Arthritis Rheum* 2005; **52**: 2645–2655.
- 8 Firestein GS, Zvaifler NJ. How important are T cells in chronic rheumatoid synovitis? II. T cell-independent mechanisms from beginning to end. *Arthritis Rheum* 2002; **46**: 298–308.
- 9 Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A *et al.* Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br J Rheumatol* 1993; **32**: 903–907.
- 10 Deighton CM, Wentzel J, Cavanagh G, Roberts DF, Walker DJ. Contribution of inherited factors to rheumatoid arthritis. *Ann Rheum Dis* 1992; **51**: 182–185.
- 11 Ramirez AS, Rosas A, Hernandez-Berian JA, Orenge JC, Saavedra P, de la FC *et al.* Relationship between rheumatoid arthritis and Mycoplasma pneumoniae: a case-control study. *Rheumatology (Oxford)* 2005; **44**: 912–914.
- 12 Haier J, Nasralla M, Franco AR, Nicolson GL. Detection of mycoplasmal infections in blood of patients with rheumatoid arthritis. *Rheumatology (Oxford)* 1999; **38**: 504–509.

- 13 Mehraein Y, Lennerz C, Ehlhardt S, Remberger K, Ojak A, Zang KD. Latent Epstein-Barr virus (EBV) infection and cytomegalovirus (CMV) infection in synovial tissue of autoimmune chronic arthritis determined by RNA- and DNA-*in situ* hybridization. *Mod Pathol* 2004; **17**: 781–789.
- 14 Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. *Nat Rev Microbiol* 2005; **3**: 281–294.
- 15 van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG, Voskuyl AE, Rustenburg F, Baggen JM *et al*. Rheumatoid Arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. *Ann Rheum Dis* 2007; **66**: 1008–1014.
- 16 Rubins KH, Hensley LE, Jahrling PB, Whitney AR, Geisbert TW, Huggins JW *et al*. The host response to smallpox: analysis of the gene expression program in peripheral blood cells in a nonhuman primate model. *Proc Natl Acad Sci USA* 2004; **101**: 15190–15195.
- 17 Demeter J, Beauheim C, Gollub J, Hernandez-Boussard T, Jin H, Maier D *et al*. The Stanford Microarray Database: implementation of new analysis tools and open source release of software. *Nucleic Acids Res* 2007; **35**: D766–D770.
- 18 Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001; **98**: 5116–5121.
- 19 Loots GG, Ovcharenko I, Pachter L, Dubchak I, Rubin EM. rVista for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Res* 2002; **12**: 832–839.
- 20 Moynagh PN. TLR signalling and activation of IRFs: revisiting old friends from the NF-kappaB pathway. *Trends Immunol* 2005; **26**: 469–476.
- 21 Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K *et al*. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004; **2**: E7.
- 22 Wucherpfennig KW. Mechanisms for the induction of autoimmunity by infectious agents. *J Clin Invest* 2001; **108**: 1097–1104.
- 23 Cacoub P, Poynard T, Ghillani P, Charlotte F, Olivi M, Piette JC *et al*. Extrahepatic manifestations of chronic hepatitis C. MULTIVIRC Group. Multidepartment Virus C. *Arthritis Rheum* 1999; **42**: 2204–2212.
- 24 Stahl HD, Hubner B, Seidl B, Liebert UG, van dH I, Wilbrink B *et al*. Detection of multiple viral DNA species in synovial tissue and fluid of patients with early arthritis. *Ann Rheum Dis* 2000; **59**: 342–346.
- 25 Balandraud N, Roudier J, Roudier C. Epstein-Barr virus and rheumatoid arthritis. *Autoimmun Rev* 2004; **3**: 362–367.
- 26 Olsen N, Sokka T, Seehorn CL, Kraft B, Maas K, Moore J *et al*. A gene expression signature for recent onset rheumatoid arthritis in peripheral blood mononuclear cells. *Ann Rheum Dis* 2004; **63**: 1387–1392.
- 27 Hitchens MR, Robbins PD. The role of the transcription factor DP in apoptosis. *Apoptosis* 2003; **8**: 461–468.
- 28 Lunardi C, Marguerie C, Walport MJ, So AK. T gamma delta cells and their subsets in blood and synovial fluid from patients with rheumatoid arthritis. *Br J Rheumatol* 1992; **31**: 527–530.
- 29 Su AI, Pezacki JP, Wodicka L, Brideau AD, Supekova L, Thimme R *et al*. Genomic analysis of the host response to hepatitis C virus infection. *Proc Natl Acad Sci USA* 2002; **99**: 15669–15674.
- 30 Toussiot E, Wendling D, Tiberghien P, Luka J, Roudier J. Decreased T cell precursor frequencies to Epstein-Barr virus glycoprotein Gp110 in peripheral blood correlate with disease activity and severity in patients with rheumatoid arthritis. *Ann Rheum Dis* 2000; **59**: 533–538.
- 31 Par G, Rukavina D, Podack ER, Horanyi M, Szekeres-Bartho J, Hegedus G *et al*. Decrease in CD3-negative-CD8dim(+) and Vdelta2/Vgamma9 TcR+ peripheral blood lymphocyte counts, low perforin expression and the impairment of natural killer cell activity is associated with chronic hepatitis C virus infection. *J Hepatol* 2002; **37**: 514–522.
- 32 Mihm S, Frese M, Meier V, Wietzke-Braun P, Scharf JG, Bartenschlager R *et al*. Interferon type I gene expression in chronic hepatitis C. *Lab Invest* 2004; **84**: 1148–1159.
- 33 Lyles DS. Cytopathogenesis and inhibition of host gene expression by RNA viruses. *Microbiol Mol Biol Rev* 2000; **64**: 709–724.
- 34 Poccia F, Boullier S, Lecoer H, Cochet M, Poquet Y, Colizzi V *et al*. Peripheral V gamma 9/V delta 2 T cell deletion and anergy to nonpeptidic mycobacterial antigens in asymptomatic HIV-1-infected persons. *J Immunol* 1996; **157**: 449–461.
- 35 Kasperkovitz PV, Timmer TC, Smeets TJ, Verbeet NL, Tak PP, van Baarsen LG *et al*. Fibroblast-like synoviocytes derived from patients with rheumatoid arthritis show the imprint of synovial tissue heterogeneity: evidence of a link between an increased myofibroblast-like phenotype and high-inflammation synovitis. *Arthritis Rheum* 2005; **52**: 430–441.
- 36 Kol A, Bourcier T, Lichtman AH, Libby P. Chlamydial and human heat shock protein 60s activate human vascular endothelium, smooth muscle cells, and macrophages. *J Clin Invest* 1999; **103**: 571–577.
- 37 Kawane K, Ohtani M, Miwa K, Kizawa T, Kanbara Y, Yoshioka Y *et al*. Chronic polyarthritis caused by mammalian DNA that escapes from degradation in macrophages. *Nature* 2006; **443**: 998–1002.
- 38 Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N *et al*. Cutting edge: molecular portrait of human autoimmune disease. *J Immunol* 2002; **169**: 5–9.
- 39 Anzilotti C, Merlini G, Pratesi F, Tommasi C, Chimenti D, Migliorini P. Antibodies to viral citrullinated peptide in rheumatoid arthritis. *J Rheumatol* 2006; **33**: 647–651.
- 40 Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX *et al*. Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 2002; **359**: 1301–1307.
- 41 Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863–14868.
- 42 Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S *et al*. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res* 2005; **33**: D284–D288.