**Supplementary information**

**MATERIALS AND METHODS:**

## Monocyte isolation for transcriptome analyses:

For transcriptome analyses, bone marrow samples from 8 RA and 8 OA patients were collected at the operating room by surgical spoon into sodium citrate solution. They were passed through mesh filter before processing like blood samples from 6 RA and 6 OA as described earlier.([1](#_ENREF_1)) In brief, after erythrocyte lysis (EL buffer, Qiagen, Hilden, Germany), and neutrophil depletion by anti-CD15 microbeads (MACS, Miltenyi, Germany), monocytes were isolated either by labelling with anti-CD14 antibody (BD Biosciences) and sorting on FACS Aria (BD Biosciences) or by labelling with anti-CD14 microbeads and sorting by MACS technology (MACS, Miltenyi, Germany). Only monocytes from RA1, RA2, OA1 and OA2 bone marrow samples and RA1 and OA1 blood samples were isolated with anti-CD14 microbeads. Only 3 matched samples of bone marrow and blood from RA1, OA1 and OA2 were used for transcriptome analysis and other samples of bone marrow and blood from RA and OA patients were not matched. Purity and viability of isolated monocytes were >90% (supplementary table 5).

For flow cytometry analyses, paired samples of bone marrow and blood (RA=11; OA=9) were subjected to lysis of erythrocytes, blockade of FcR (FcR blocking reagent, Miltenyi, Germany), and staining with anti-CD14-APC, anti-CD16-FITC, anti-CD163-PE and anti-HLA-DR-PerCP.

In a similar way, 6 paired samples of RA synovial fluid and blood cells were investigated by cytometry with anti-CD14-PE-Cy7, anti-CD16-APC-Cy7, anti-CD163-PE and anti-HLA-DR-PE-Cy5 (BD Biosciences). The same combination of antibodies was utilised for staining of samples from blood of 12 RA patients and 12 healthy donors in validation study.

For protein analysis by ELISA the matched samples of synovial fluid and serum from RA (n=17) and OA (n=16) patients and sera from healthy donors (n=14) were examined for sCD14, sCD163 (Quantikine, R&D System) and S100P (CircuLex, MBL International) according to the manufacturer’s instructions.

RA patients defined by ACR criteria had long-standing disease (median 14 years, range 4-22 years) with progressive arthritis and radiographically confirmed joint destruction. All RA patients were treated with methotrexate, low dose corticosteroids and/or non-steroidal anti-inflammatory drugs (NSAIDs) but no biologics.

## RNA isolation, Affymetrix GeneChip hybridization and quality controls for gene-expression analyses

RNA preparation, quality controls and array hybridization were performed as previously described.([1](#_ENREF_1), [2](#_ENREF_2)) In brief, sorted monocytes were immediately lysed in RLT-buffer, total RNA extracted (RNeasy Mini Kit, Qiagen, Hilden, Germany), quality assessed (Agilent 2100 Bioanalyzer, Agilent Technologies, Waldbronn, Germany) and quantified (NanoDrop ND-1000 spectrophotometer, NanoDrop Technologies, Wilmington, DE, USA). After reverse transcription of 300 ng total RNA (Ambion, Austin, TX, USA) and generation of cRNA probes (Enzo RNA Transcript Labeling kit; Affymetrix), 50 μg/mL of biotinylated and fragmented cRNA was hybridized to HG-U133 Plus 2.0 arrays. After washing and fluorescence labelling, arrays were scanned (Affymetrix GeneChip Scanner 3000).

## Functional analyses of microarray data

MeV (MultiExperiment Viewer, version 4.0, MA, USA) was applied for hierarchical clustering of signal and correlation coefficient matrices and Qlucore (Lund, Sweeden) for principal component analyses (PCA).([3](#_ENREF_3)) GeneCodis was applied for identification and grouping of genes into different Gene Ontology (GO) molecular functions.([4](#_ENREF_4)) Ingenuity pathway analysis (IPA) was applied to assign the RA-related profiles from bone marrow and blood to distinct molecular networks (IPA, Qiagen Redwood City, USA).

**Principal component analyses and Gaussian Modelling**

To calculated distance between 4 groups (RA-BM, OA-BM, RA-PB and OA-PB) we applied principal component analyses (PCA) projection and Gaussian modelling. In short all differentially expressed genes identified in comparisons between RA and OA bone marrow and RA and OA blood samples were used for PCA analyses. Signals from differentially expressed genes were log-transformed and standardised by gene across all samples (mean=0, StD=1) and PCA was calculated by the R-function prcomp. The first 3 principal components (PCA1, PCA2 and PCA3) were used to estimate mean and variance between the 4 groups. The distances were calculated by Mahalanobis metrics (included in supplementary figure 1). The smallest distance of 2.8 was detected between RA-BM and OA-BM, indicating that they rather form one group of samples. The greatest distance was obvious between OA-PB and RA-BM groups (151.6) and it was followed by the distance between OA-PB and OA-BM (139.3). Thus, the PCA analysis indicated clustering of samples into 3 groups composed of 1) RA-BM and OA-BM samples, 2) OA-PB samples and 3) RA-PB samples. In addition, PCA suggested that RA-PB is position between BM samples from RA and OA (distance from RA-BM and OA-BM is 21.2 and 23.5, respectively) and OA-PB (with slightly smaller distance of 16.8).

## Functional interpretation of RA bone marrow and blood transcriptomes with reference signatures

Own and public data from Gene expression omnibus (GEO) were used for mapping of differentially expressed genes to reference conditions. In total, 68 Affymetrix HG-U133 Plus 2.0 microarrays from 4 different studies were utilised including i) myeloid progenitors from bone marrow (GSE42519), ii) monocytes activated in the blood with defined inflammatory stimuli (GSE38351), iii) monocyte subsets from the blood, where CD16- portrayed the classical subset and CD16+ the non-classical, more differentiated subset (GSE18565), and iv) all blood leukocytes from healthy donors before and after treatment with G-CSF (granulocyte-colony stimulatory factor) (GSE7400).([1](#_ENREF_1), [5-8](#_ENREF_5))

In detail, the samples from bone marrow included 34 microarrays from 11 different cell types: hematopoietic stem cells (HSC; n=4), multipotent progenitors (MPP; n=2), common myeloid progenitors (CMP; n=3), megakaryocyte-erythrocyte progenitors (MEP; n=2), granulocyte-monocyte progenitors (GMP; n=5), early promyelocytes (early-PM; n=3), late promyelocytes (late-PM; n=3), myelocytes (MY; n=2), metamyelocytes (MM; n=3), band cells (BC; n=4) and polymorphonuclear (PMN; n=3). The samples from blood monocytes included 34 microarrays from 3 different studies: (1) a study related to monocyte activation *ex vivo* by various stimuli and including 6 groups of microarrays: control monocytes (Ctr\_0, n=3), control monocytes incubated for 90 minutes (Ctr\_90, n=3), monocytes stimulated for 90 minutes with TNF (n=3), or LPS (n=3), or IFNγ (n=3), or IFNα (n=3); (2) a study that analysed 2 different monocyte subsets in blood: CD16- (n=3) and CD16+ (n=3), and (3) whole blood leukocytes from healthy stem cell donors before and after administration of G-CSF for 5 days: pre G-CSF (n=5) and post G-CSF (n=5). An overview is included in supplementary figure 14.

In addition, 21 reference transcriptomes were included in the analyses to demonstrate that potential contamination of monocytes with neutrophils and T-cells does not correspond to the gene-patterns identified by co-expression analyses in figures 2 and 3. These transcriptomes portrayed neutrophils (n=3), CD4+ T-cells (n=3), CD8+ T-cells (n=3), CD19+ B-cells (n=3), CD56 NK-cells (n=3), CD14+ monocytes (n=3) and are obtained from the GEO repository (GSE58173). To demonstrate the influence of various methods used for monocyte isolation we included 3 additional transcriptomes that presented monocytes isolated by negative selection, obtained from GSE49910. All these samples are presented in supplementary figure 3 and are coloured in grey.

To harmonize data from different studies, all transcriptomes were quantile normalised and subsequently applied for co-expression analysis. Pearson correlation coefficients were calculated between signals of all differentially expressed probe sets in the 68 different reference transcriptomes. Hierarchical clustering of this gene-to-gene correlation matrix was performed by applying Euclidean distance and average linkage as an agglomeration rule.

**Calculation of scores that quantify changes of gene expression in reference transcriptomes and determine the impact of these changes on RA and OA profiles in blood and BM**

To select probe sets differentially expressed in the reference transcriptomes we utilised the same criteria as for characterisation of RA and OA bone marrow and blood profiles in this study.([1](#_ENREF_1), [5](#_ENREF_5), [9](#_ENREF_9)) In brief, probe sets differentially expressed in at least 60% of MAS5.0 pair-wise comparisons between the reference transcriptomes were selected.

To determine profiles related to G-CSF stimulation, CD16+ differentiation as well as early stage, metamyelocyte (MM) stage and late stage of myeloid differentiation in bone marrow, the following pair-wise comparisons were performed: i) Post G-CSF vs Pre G-CSF (GSE7400), ii) CD16+ vs CD16- (GSE18565), iii) GMP vs HSC (GSE42519), iv) MM vs HSC (GSE42519) and v) PMN vs HSC (GSE42519) (supplementary figures 9-13).

To quantify differential expression of probe sets in these comparisons, we calculated a score for each probe set by multiplication of the frequency of change call (increased or decreased up to 100%, as indicated by MAS5.0 pair-wise comparisons) with the average signal log ration (SLR) (obtained from all pair-wise comparisons). For example, the G-CSF induced profile was characterised with 7353 probe sets and the overall distribution of their scores was presented in supplementary figure 9. Distribution of scores for probe sets that are common for G-CSF response and RA-PB, OA-PB, RA-BM or OA-BM profiles characterise the relevance of overlap with G-CSF induced profiles. In detail, 107, 78, 55 and 29 probe sets were common for G-CSF and RA-PB, OA-PB, RA-BM and OA-BM profiles, respectively. Distribution of scores for these common probe sets indicated a strong G-CSF response in RA-PB when compared to OA-PB (p=6.067x10-12). Interestingly, 76 out of 107 probe sets were identified by co-expression analysis in figure 3C. This demonstrates the power of co-expression analysis, which identified leading candidate genes of reference transcriptome patterns and emphasises this analysis as a very comprehensive and detailed approach for functional interpretation of differentially expressed genes in RA and OA.

The same approach was used for calculating scores for 5964 probe sets identified in pair-wise comparisons between CD16+ and CD16- monocytes in blood (supplementary figure 10). In the same way, scores were calculated for 3 signatures of myelopoiesis obtained in comparisons between GMP vs HSC, MM vs HSC, and PMN vs HSC, which were determined with 3129, 9976 and 13449 probe-sets, respectively (supplementary figures 11-13).

**References:**

1. Smiljanovic B, Grun JR, Steinbrich-Zollner M, Stuhlmuller B, Haupl T, Burmester GR, et al. Defining TNF-alpha- and LPS-induced gene signatures in monocytes to unravel the complexity of peripheral blood transcriptomes in health and disease. J Mol Med (Berl). 2010 Oct;88(10):1065-79. PubMed PMID: 20640394. Epub 2010/07/20. eng.

2. Biesen R, Demir C, Barkhudarova F, Grun JR, Steinbrich-Zollner M, Backhaus M, et al. Sialic acid-binding Ig-like lectin 1 expression in inflammatory and resident monocytes is a potential biomarker for monitoring disease activity and success of therapy in systemic lupus erythematosus. Arthritis and rheumatism. 2008 Apr;58(4):1136-45. PubMed PMID: 18383365. Epub 2008/04/03. eng.

3. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003 Feb;34(2):374-8. PubMed PMID: 12613259. Epub 2003/03/05. eng.

4. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A. GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W478-83. PubMed PMID: 22573175. Pubmed Central PMCID: 3394297.

5. Smiljanovic B, Grun JR, Biesen R, Schulte-Wrede U, Baumgrass R, Stuhlmuller B, et al. The multifaceted balance of TNF-alpha and type I/II interferon responses in SLE and RA: how monocytes manage the impact of cytokines. J Mol Med (Berl). 2012 Nov;90(11):1295-309. PubMed PMID: 22610275.

6. Frankenberger M, Hofer TP, Marei A, Dayyani F, Schewe S, Strasser C, et al. Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint. European journal of immunology. 2012 Apr;42(4):957-74. PubMed PMID: 22531920.

7. Rapin N, Bagger FO, Jendholm J, Mora-Jensen H, Krogh A, Kohlmann A, et al. Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. Blood. 2014 Feb 6;123(6):894-904. PubMed PMID: 24363398.

8. Buzzeo MP, Yang J, Casella G, Reddy V. Hematopoietic stem cell mobilization with G-CSF induces innate inflammation yet suppresses adaptive immune gene expression as revealed by microarray analysis. Experimental hematology. 2007 Sep;35(9):1456-65. PubMed PMID: 17761290.

9. Menssen A, Edinger G, Grun JR, Haase U, Baumgrass R, Grutzkau A, et al. SiPaGene: A new repository for instant online retrieval, sharing and meta-analyses of GeneChip expression data. BMC genomics. 2009;10:98. PubMed PMID: 19265543. Epub 2009/03/07. eng.