

Supplementary File 1:

Material and Methods:

Monocyte purification and transcriptomic profiling

Monocytes from heparinized whole blood were isolated by positive magnetic selection using CD14 MicroBeads (MACS WholeBlood Column Kit, Miltenyi Biotec). The purity of CD14⁺ cells was at least 95%. Cells were lysed in RLT buffer (Qiagen) and stored at -80°C before RNA extraction.

In cohort 1, mRNA profiling was performed by GeneChip microarrays (Affymetrix HG-U133 Plus 2.0, Thermo Fisher Scientific). RNA isolation, gene-chip hybridization and quality checks were performed as previously described (Smiljanovic et al., 2010).

In cohort 2, monocyte transcriptomes were generated by RNA-sequencing. Sequencing libraries were prepared from high quality RNA (RIN>8) using the Illumina TruSeq® stranded total RNA protocol. Libraries were then sequenced paired-ended using an Illumina HiSeq4000 sequencer to generate an average of 65 million of uniquely mapped and properly paired reads per sample.

Statistical analysis

For GeneChip microarray analysis, CEL-files were normalized with Robust Multichip Averaging (RMA) algorithm using RMAExpress with background correction, quantile normalization, and Median Polish as summarization. This quantile normalization was done for each comparison separately. To determine significant differences in expression between two groups high performance chip data analysis (HPCDA) was performed according to Menssen et al. (Menssen et al., 2009). In group-wise comparisons each transcriptome of one group was compared to each transcriptome from the other group (i. e. 250 comparisons for 25 axSpA vs. 10 HC transcriptomes). Genes with a HPCDA score of > 100 and all values described in Menssen et al. (Menssen et al., 2009) were considered significant. This empirically developed HPCDA

score includes all values calculated with HPCDA, as Fold Change, signal differences between two groups, four different t tests, presence or absence of each gene, Increased or Decreased calls for every chip to chip comparison, high and low confidence intervals and HPCDA-Volcano plot x and y values (compare Table 1 at the end and further explanations in DOI: 10.13140/2.1.4942.9289 or DOI: 10.13140/RG.2.1.1426.8321). This analysis promotes detection of small changes between samples which helps in identifying context-specific pattern.

For RNA-sequencing, RNA-seq reads were aligned with the splice-aware STAR aligner on the human genome (GRCh38) and transcriptome (gencode v28) after adapter and low-quality bases trimming. Uniquely mapping fragments were assigned to exons and genes with featureCounts. Counts data were normalized using the TMM (trimmed means of M-values) method. Analysis of differential gene expression was carried out with edgeR, using a linear model including the disease status, the gender and the sequencing experiment. Scores reaching a p-value < 0.05 were considered significant.

Hierarchical clustering of transcriptomes using differentially expressed probe-sets was performed with Genes@Work software which applies Pearson correlation with z normalized gene vectors as similarity/distance measure and center of mass as cluster type (Lepre et al., 2004). Principal component analysis (PCA) was performed using Qlucore (Lund, Sweden).

Functional allocation and co-expression analysis of differentially expressed genes

For functional allocation, gene set enrichment analysis (GSEA) using the pre-ranked list of differentially expressed probe-sets (GeneChip) or genes (RNA-seq) was performed using the GSEA software (Broad Institute) and the annotated gene sets of

the Molecular Signature Database (MSigDB) v7.2 (Subramanian et al., 2005; Mootha et al., 2003).

In addition, differentially expressed probe-sets identified in comparisons between 1) axSpA versus HC, 2) BASDAI^{high} vs. BASDAI^{low} and 3) nr-axSpA versus AS were analyzed for co-expression in 70 reference transcriptomes (all Affymetrix HG-U133 Plus 2.0 transcriptomes) generated by our own or retrieved from Gene expression omnibus (GEO) data repository.

The reference transcriptomes included (1) 34 transcriptomes of 11 different cell types of early and late myelopoiesis from bone marrow (BM; GSE42519) (Rapin et al., 2014), (2) 15 transcriptomes from blood cells comprising BDCA1⁺ (n=3) and BDCA3⁺ DCs (n=3), CD15⁺ PMN (n= 3) and blood monocyte subsets including classical CD14⁺CD16⁻ monocytes (Mo-CD16⁻; n=3), and non-classical monocytes CD14⁻CD16⁺ (Mo-CD16⁺; n=3) (Ziegler-Heitbrock et al., 2010) (GSE18565), (3) six transcriptomes of blood leukocytes from healthy donors before and after treatment with G-CSF (GSE7400) (Buzzeo et al., 2007) and (4) 12 transcriptomes of monocytes isolated from peripheral blood and incubated for 90 minutes without stimulus or stimulated with TNF α , LPS, IFN γ or IFN α (GSE38351) (Smiljanovic et al., 2012).

For the coexpression analysis, 70 reference transcriptomes were utilized and Pearson correlation coefficients were calculated between signals of all differentially expressed probe-sets. Hierarchical clustering of this gene-to-gene correlation matrix was performed by applying Euclidean distance and average linkage as an agglomeration rule as previously described (Smiljanovic et al., 2020). The order of probe-sets determined by these calculations was used to determine co-expressed probe-sets among the 70 reference transcriptomes. To harmonize data from different studies, all transcriptomes were quantile normalized and subsequently applied for co-expression analysis.

Intensity of expression is color-coded – red indicated increased signal expression (max = 2) or positive correlation (max=1) and blue indicates decreased signal expression. Cluster of coexpression were visually depicted and boxed in graphical depictions.

