

**RUNNING HEAD:** Genomic alignments of mouse and human arthritogenic synovial fibroblasts.

**TITLE:** Genomic responses of mouse synovial fibroblasts during TNF-driven arthritogenesis greatly mimic those of human rheumatoid arthritis.

**AUTHORS:** Evangelos Ntougkos PhD<sup>1,2</sup>, Panagiotis Chouvardas MSc<sup>1,2,7</sup>, Fani Roumelioti MSc<sup>1,3</sup>, Caroline Ospelt PhD<sup>4</sup>, Mojka Frank-Bertoncelj PhD<sup>4</sup>, Andrew Filer PhD<sup>5</sup>, Christopher D. Buckley PhD<sup>5</sup>, Steffen Gay PhD<sup>4</sup>, Christoforos Nikolaou PhD<sup>1,6</sup>, George Kollias PhD<sup>1,7</sup>.

**Financial support:** This work has been funded by the IMI project BTCure (Grant Agreement no. 115142-2) to GK and SG, including a Research grant from Janssen Biologics B.V.; the Greek GSRT project INNATE FIBROBLAST (ERC06, co-financed by the ESF and NSRF 2007–2013) to GK; the FP7 Advanced ERC grant MCs-inTEST (Grant Agreement no. 340217) to GK; and Research Projects for Excellence IKY/SIEMENS (Grant Agreement no. 3288) to GK.

---

<sup>1</sup> BSRC Alexander Fleming, Vari, Greece.

<sup>2</sup> Authors contributed equally to this work.

<sup>3</sup> Department of Pathophysiology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece.

<sup>4</sup> Centre of Experimental Rheumatology, University Hospital of Zurich, Zurich, Switzerland.

<sup>5</sup> University of Birmingham, Birmingham, United Kingdom.

<sup>6</sup> Computational Genomics Group, Department of Biology, University of Crete, Heraklion, Greece.

<sup>7</sup> Department of Physiology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/art.40128

© 2017 American College of Rheumatology

Received: Dec 21, 2016; Revised: Mar 15, 2017; Accepted: Apr 11, 2017

**Correspondence to:** George Kollias, BSRC Alexander Fleming, Vari, 16672, Greece,

g.kollias@fleming.gr, tel.: +302109656507.

Accepted Article

## ABSTRACT

**Objective.** Aberrant activation of synovial fibroblasts (SFs) is a key determinant in the pathogenesis of rheumatoid arthritis (RA). We aimed to produce a map of gene expression and epigenetic changes occurring in this cell type during disease progression in the human TNF-transgenic model of arthritis, and identify commonalities with human SFs.

**Methods.** We used deep sequencing to probe the transcriptome, the methylome and the chromatin landscape of cultured mouse arthritogenic SFs at three stages of disease, as well as SFs stimulated with human TNF. We performed bioinformatics analyses at the gene, pathway and network levels, compared mouse and human data, and validated selected genes in both species.

**Results.** We report that SF arthritogenicity is reflected on distinct dynamic patterns of transcriptional deregulation, enriched in pathways of the innate immune response and mesenchymal differentiation. A functionally-representative subset of these changes is associated with methylation, mostly in gene bodies. The arthritogenic state involves highly active promoters, marked by H3K4 trimethylation. There is significant overlap between mouse and human data, at the level of deregulated genes and to an even higher extent at the level of pathways.

**Conclusion.** This work presents the first systematic examination of the pathogenic changes that occur in mouse synovial fibroblasts in progressive TNF-driven arthritogenesis. Significant correlations with respective human RA SF data further validate the human TNF-transgenic mouse as a reliable model of the human disease. The resource of data generated here may serve as a framework for the discovery of novel pathogenic mechanisms and disease biomarkers.

Rheumatoid arthritis (RA) is a chronic, inflammatory, disease, characterised by extensive damage to diarthroidal joints. Our understanding of RA has largely relied on the use of mouse models, such as collagen-induced arthritis, collagen antibody-induced arthritis and the TNF-transgenic models; work on the latter established the proof of principle for the pivotal role of TNF in disease pathogenesis and its therapeutic targeting by anti-TNF biologics (1). The same models established the synovial fibroblast (SF) as a key cellular player and target of TNF in RA (2). Whilst under physiological conditions SFs provide nourishment and protection to the joint, upon chronic innate immune activation, they undergo a complex process described as arthritogenic “transformation”: they proliferate aberrantly and acquire resistance to death, thus becoming hyperplastic; and they express pro-inflammatory cytokines, matrix degrading enzymes and cell adhesion molecules, thus contributing to an inflammatory and tissue-damaging microenvironment (3–5). Moreover, SFs can migrate to distant sites via the vasculature, maintaining and transferring their newly-acquired destructive properties, both in human and in mouse models (6,7).

Epigenetic changes underlie the activated and aggressive phenotype of RA SFs (8). Most of the relevant studies have examined the contribution of the methylome, reporting global genomic hypo-methylation in the arthritogenic human synovial fibroblast (9–12). In addition, micro RNA deregulations are known to accompany disease both in human and mouse (13).

In order to assess the gene expression and epigenetic changes that occur in synovial fibroblasts during the arthritogenic process, we have initiated a multi-level omics analysis of human TNF-transgenic (Tg197) SFs using next generation sequencing (NGS) approaches. Here, we report the molecular profile of the activated synovial

fibroblast at the level of the transcriptome, the methylome and active chromatin as evidenced by H3K4 trimethylation. This profile entails distinct deregulations associated with functional pathways, notably related to the innate immune response and mesenchymal differentiation; methylation changes correlated with expression in specific genes; and an activated chromatin profile. In the context of this key pathogenic cell type, at the level of genes and even more at the level of pathways, the human TNF-transgenic model aligned significantly with human data.

## **MATERIALS AND METHODS**

### **Mouse and human synovial fibroblasts**

Mouse synovial fibroblasts were isolated from CBAXC57BL/6 human TNF-transgenic (TghuTNF; Tg197) mice (1) and wild-type littermates housed under specific pathogen-free conditions at three disease stages: early (3 weeks), established (8 weeks) and late (11 weeks), as previously described (14). Three biological replicates were isolated per condition; for each, a pool of three mixed-gender mice was used. Purity of all isolations was assessed by FACS, with acceptance criteria being: more than 85% positive for CD90.2 and less than 2.5% positive for CD45.

Human samples used for RNA sequencing were isolated from the knees of three RA patients, as well as of three healthy individuals (as defined below), as detailed elsewhere (15). Human RA synovial fibroblasts used for validations were isolated from synovial tissues obtained from patients undergoing joint replacement surgery at the Schulthess Clinic, Zurich, Switzerland. All patients fulfilled the 2010 ACR/EULAR criteria for the classification of RA. Samples described as healthy synovial

Genomic alignments of mouse and human arthritogenic synovial fibroblasts.

fibroblasts were grown from synovial biopsies from arthralgia patients without clinical or histological signs of inflammation or cartilage destruction, at the Queen Elizabeth Hospital, Birmingham, UK. Details regarding these human samples can be found in Supplementary Table 1. Human studies were approved by the local ethics committees of the University Hospital Zurich, Switzerland and the University of Birmingham, UK. Informed consents were obtained from all patients.

### **Extraction of RNA and DNA**

RNA was extracted from mouse SFs using the Absolutely RNA Miniprep Kit (Agilent Technologies) and from human SFs using the miRNeasy Mini kit (Qiagen). DNA was extracted from mouse SFs using the PureLink Genomic DNA Mini Kit (Invitrogen).

### **Enrichment for MethylCap and H3K4me3**

In order to enrich for methylated DNA, the MethylMiner DNA Enrichment Kit from Invitrogen was used as per the manufacturer's instructions. DNA was firstly fragmented by sonication to an average size of 250 bp using a Covaris S220 focused-ultrasonicator. Then enrichment was performed with a double elution (high and low salt) in order to capture both sparsely- and densely-methylated DNA; fractions were combined in the end. To enrich for H3K4me3, DNA was processed using the iDeal ChIP-seq Kit (Diagenode) in combination with a ChIP-grade H3K4me3 antibody (Diagenode), as per the manufacturer's instructions. Chromatin was firstly sheared by sonication using a Bioruptor (Diagenode) and then subjected to the H3K4me3 enrichment protocol with a minor modification: after performing an RNase I digestion for 30 min at 37°C, the de-crosslinked DNA was eluted using the QIAquick

Genomic alignments of mouse and human arthritogenic synovial fibroblasts.

PCR Purification Kit (Qiagen). Enrichment (in comparison to a non-enriched input control) was confirmed by quantitative PCR.

### **Preparation of libraries and next generation sequencing**

All library preparations, next generation sequencing and quality control steps were performed by the McGill University and Genome Quebec Innovation Centre. More specifically, for RNA-seq, TruSeq RNA libraries were prepared and samples were run in an Illumina HiSeq2000 platform using a 100 bp paired-end setup. For MethylCap-seq, TruSeq genomic DNA libraries were prepared and samples were run in an Illumina HiSeq platform using a 100 bp paired-end setup. For ChIP-seq, TruSeq genomic DNA libraries were prepared and samples were run in an Illumina HiSeq platform using a 100 bp paired-end setup. For human samples, preparation of libraries and next generation sequencing was performed as detailed elsewhere (15)

### **Bioinformatic analyses**

**General:** Quality of sequencing was accessed using FastQC software (16), which also calculates the average quality of reads. The number of mapped reads was calculated with the use of the samtools tool (-flagstat) (17). Conversion of alignment files (bam) to bed format was done with bedtools (bamtobed) (18). Bam files were also converted to bedgraph with bedtools (genomecov -bga) and bedgraph files were converted to bigwig using the UCSC bedGraphToBigWig utility. Volcano plots were created in R with an in-house developed script, which is based on the ggplot package. Heat maps were also generated in R with the *heatmap.2* function of the gplots package. Functional analyses were performed with the online tool Enrichr

(19). For all statistical comparisons the cut-off for significance was set to 0.05. With the exception of KEGG pathway enrichment and human RNA-seq analyses, in all other comparisons  $p$  values were adjusted for multiple comparisons. Human-mouse overlap was tested using an online tool based on a normal approximation to the exact hypergeometric probability (20).

**RNA-seq:** Mouse RNA-seq samples were analysed with the tophat-cufflinks pipeline (21–23). Reads were mapped to the mm9 genome with tophat2. Transcripts were assembled using cufflinks. Final transcriptome assembly was performed with cuffmerge and differential expression was identified with cuffdiff. Venn diagrams were created with InteractiVenn (24) or the online tool Venny (25). Clustering of genes was performed with  $k$  means based on both the fold change and the  $q$  value of all time points. The number of clusters was selected with the use of the elbow rule. The average cluster profile was plotted in R using the fold changes and the  $q$  values of the clusters' centroids. RNEA was performed as previously described (26). Networks were visualised in Cytoscape (27) and the characteristics of nodes (colour and size) were based on betweenness centrality and outdegree. The compilation of differentially-expressed gene lists from human RNA sequencing data has been described elsewhere (15).

**MethylCap-seq/H3K4me3-seq:** MethylCap- and H3K4me3-seq reads were mapped using the BWA tool (28). Visualisations of general profiles were created with seqplots (29). Methylation plots were generated with the use of a pseudo-length of 10 kb, to which the genes were extended or shrunk using linear approximation (“anchored features”). Differential analysis was performed with diffReps (30) using the default



Genomic alignments of mouse and human arthritogenic synovial fibroblasts.

normalisation procedure and windows of 500 bases for MethylCap-seq and 1000 bases for H3K4me3-seq.

### **Quantitative RT-PCR**

For validation of deregulated genes identified by RNA-seq in the mouse, an independent set of mouse SF RNA was used, comprising triplicate biological samples for each condition. All primers were QuantiTect Primer Assays (Qiagen) and were used in conjunction with QuantiFast SYBR Green RT-PCR Kit (Qiagen), according to the manufacturer's instructions. Thermal cycling was performed in a Bio-rad CFX96

Touch Real-Time PCR Detection System. Expression was quantified relative to *B2m*.

For validations in human SFs, reverse-transcribed total RNA was measured in a 7500 Real-Time PCR System (Applied Biosystems) using self-designed primers (Suppl. Table 2) in combination with SYBR green dye. No template control samples, dissociation curves and samples containing the non-transcribed RNA were run as controls. For normalisation, expression of *HPRT1* was used (target – *HPRT1* expression =  $\Delta C_T$ ).

## RESULTS

### **Arthritogenic synovial fibroblasts from the Tg197 mouse exhibit distinct patterns of transcriptional deregulation during disease progression.**

In order to explore cell-specific transcriptional changes accompanying disease progression in the Tg197 mouse model of rheumatoid arthritis, synovial fibroblasts were isolated from diseased and healthy animals at three stages of disease: 3 weeks, representing an early stage of not overtly manifested pathology; 8 weeks, which is a stage of established disease; and 11 weeks, a late stage of severe pathology (Suppl.

Figs. 1 and 2A). RNA samples were processed for NGS, confirming the quality of sequencing as shown in Supplementary Figures 2B and C. As can be seen in Figure 1A, more than one thousand significant differentially-expressed genes (DEGs) were identified. The early stage of disease is characterised by the least, equally represented by up- and down-regulation, whereas the most were identified in the established stage, where up-regulation predominates; this is also true for the late stage of disease. A considerable number of DEGs are persistent throughout disease (177 up-regulated; 114 down-regulated), while there are also stage-specific ones (Fig. 1B). Interestingly, the vast majority of early DEGs are met throughout the disease, indicating genes that play a central role in disease development.

In an effort to explore further the patterns of deregulation identified, we classified DEGs in six clusters, according to their occurrence in disease progression (Fig. 1C). Three clusters were associated with up-regulated genes and three with down-regulated genes. These clusters were then subjected to functional enrichment analysis, so as to examine whether they are associated with distinct functional

categories. Regarding up-regulation clusters, immune response was identified as the most prominent functional category, being significant throughout disease progression. Stage-specific up-regulation encompasses functions of cell signalling, cell adhesion, cytokine response, cell motility and the extracellular matrix. Notably, genes down-regulated in all disease stages are functionally related to differentiation. This becomes more pronounced in the cluster of genes down-regulated at the late stage, where we observed a deficit in the culmination of differentiation processes (termed morphogenesis). Additional functions of down-regulation across disease are related to the extracellular matrix and negative regulation of proliferation (hence an overall positive proliferative potential). Stage-specific down-regulation clusters include functions of cell division and regulation of WNT signalling. In the context of all these clusters, late is the most distant hierarchically, indicating the farthest, functionally evolved, stage of disease.

In a complementary functional approach, the deregulation of each disease stage was subjected to functional enrichment and clustering was performed according to the top 25 terms of the early stage (Fig. 1D). Functions enriched in the up-regulated genes mostly correspond to the immune response and their enrichment becomes more pronounced with disease progression. The majority of the top immune response-related ontology terms enriched in the three stages are related to the innate immune response and TLR signalling. Functions enriched in the down-regulated genes are mostly linked to differentiation and the extracellular matrix; the tendency across disease applies here too, albeit to a lesser extent. As in the first functional analysis, the late stage of disease is the most distant one.

In addition, we sought to identify transcriptional regulators that underlie the identified deregulation profiles. To this end, an in-house developed tool, RNEA, was used, which performs enrichment analysis using transcriptomic data in combination with transcriptional network assignment algorithms (26). Having observed that despite the high complexity of networks derived for each disease stage the top significant central players were consistent (Suppl. Fig. 3), we focused on the common transcriptional network that applies throughout disease. As can be seen in Figure 1E, this transcriptional network was found to have three central regulators: NFKB1, PPARG and CEBPA. These transcription factors are highlighted to have both a central and strong impact, placing the NF- $\kappa$ B and mesenchymal differentiation pathways in the core of our transcriptome profiles and thus Tg197 pathogenesis.

#### **Global DNA methylation changes associate with the activated Tg197 SF profile.**

We next examined the methylome of the arthritogenic synovial fibroblast. MethylCap sequencing was performed using the same biological sample set; quality control metrics can be seen in Supplementary Figures 2B and C. Approaching differential methylation using a promoter/gene body dichotomy in our analyses, we observed that in the early and late stage promoters, differentially-methylated regions (DMRs) show similar levels of enrichment (hyper-) and depletion (hypo-methylation; Fig. 2A). In the established stage, there is a tendency for more hyper-methylation. As regards gene body methylation, there is pronounced hypo-methylation associated with the established stage of disease. Looking into the distribution of these DNA methylation changes, as disease progresses the signal tends to spread away from the transcription start site (Fig 2B). There is a higher

enrichment in gene body methylation, most of which being exonic, while most overall differential methylation is found in CpG islands.

As to the relationship between DNA methylation and expression, we firstly noted that if we stratify our expression data in quantiles then methylation appears to be an indicator of the level of expression, positively downstream of the transcription start site, mostly negatively upstream of it (Fig. 2C). Trying to correlate the deregulations identified at the RNA level with changes in DNA methylation, we observed that approximately one quarter of our DEGs at any disease stage are accompanied by a change in methylation (Fig. 2D). Out of these, half follow the “canonical” correlation between methylation and expression (inverse for the promoter, positive for the gene body), the vast majority of which corresponding to gene body DMRs. Approximately one fifth of the latter, i.e. of genes whose deregulation at the RNA level is positively correlated with gene body methylation, are associated with exonic DMRs.

We then narrowed down our analysis to the genes that are both differentially-expressed and methylated in the expected manner and performed functional enrichment. This approach returned functions most significantly related, at all stages, to the extracellular matrix and, at the established and late stages, to cytokine signalling, proliferation, adhesion, migration, differentiation, and cell death. These enrichments constitute a representative subset of the above-described functional profile based on the RNA-seq data.

These results suggest that DNA methylation accompanies a subset of functionally-important, disease-associated deregulation in the Tg197 mouse.

**Activation of the synovial fibroblast is marked by open chromatin.**

To assess transcriptionally-active genes, we probed for H3K4 trimethylation, a histone mark associated with active promoters. Quality control metrics of the sequencing can be found in Supplementary Figure 2. Most differentially-marked regions were identified in the early time-point, where H3K4me3 depletion predominates, whereas in the other two time-points there are approximately equal numbers of enriched and depleted regions (Fig. 3A). Inspecting the distribution of changes, we observed that in all three time-points there is enrichment in the promoter, as expected for this histone mark, without any clear differences in distance to the transcription start site (TSS; Fig. 3B). Focusing therefore around the TSS, we noted that H3K4me3 signal mirrors the level of expression, both of the Tg197 and the wild-type (Fig. 3C). Interestingly, the activated state of the arthritic SF is evident in this marker of open chromatin both at the early and the established stage, as Tg197 H3K4 trimethylation levels are higher overall. Significant H3K4 trimethylation in the synovial fibroblast was hence found to be a good predictor of differential expression: the correlation between H3K4me3 and RNA expression is high for all disease stages ( $r_s = 0.69$  for the early, 0.61 for the established and 0.72 for the late time-point; Fig. 3D). Due to this high correlation, the output of functional enrichment analysis of differentially H3K4-marked and expressed genes mirrors that of the general profile (data not shown).

Based on the above, pathogenic activation of the synovial fibroblast is engraved on its chromatin, with H3K4 trimethylation being a good marker for this activation.

### **Assessment of the impact of exogenously-administered TNF on the identified signatures.**

TNF plays a pivotal role in both Tg197 and human disease, where it is a major pathogenic cytokine in the inflammatory milieu of the arthritic joint (1). We aimed to assess the *ex vivo* impact of TNF at the level of the transcriptome, methylome and active chromatin, so as to build on our understanding of the interplay of TNF and synovial fibroblast biology, also given that TNF stimulation is widely used in human RA SF studies. To this end, in parallel to our above efforts, wild-type SFs were treated with human TNF for 24 hours and analysed as the rest of the samples (see Suppl. Fig. 1 for sequencing metrics). Figure 4A summarises our findings of genes up- or down-regulated at the level of RNA, hyper- or hypo-methylated, and enriched or depleted in H3K4me3, following stimulation with human TNF.

At the level of differentially-expressed genes or differentially-marked regions, a considerable number of deregulations forming part of the Tg197 signature are shared with the TNF-induced profiles in all three time-points: approximately 85% for RNA and DNA methylation and 70% for H3K4 trimethylation (Fig. 4B). At the level of pathways, the overlap becomes more pronounced, where we observed that sixteen out of twenty KEGG pathways enriched following TNF stimulation were also found in the Tg197 signatures. Nonetheless, a high number of significant DEGs (100 for the early, 718 for the established and 554 for the late stage) and enriched pathways (1, 18 and 18, respectively), were only identified in the Tg197 profiles.

It follows from the above that, *ex vivo* as *in vivo*, TNF sensed by the synovial fibroblast contributes to the creation of a pro-inflammatory and pathogenic environment that is reflected, at the level of both genes and pathways, on pathogenic changes to the transcriptome, the methylome and chromatin.

### **Validations and comparisons with human data.**

In a technical validation of our sequencing results, a panel of 20 genes selected among the top deregulated ones identified by RNA-seq were validated by quantitative RT-PCR (qRT-PCR), using an independent sample set. As can be seen in Figure 5A, there was a very high correlation between RNA-seq and qRT-PCR data, establishing experimental reliability ( $r_s = 0.96$  for the early, 0.90 for the established and 0.90 for the late time-point).

To assess whether our findings in the mouse can be verified in the context of human disease, we made use of an RNA-seq data set comparing human RA SFs to healthy SFs. The broad picture of the identified human deregulations can be seen in Figure 5B. At the level of genes, approximately a tenth of mouse SF deregulations were also identified in the same direction in the human, a very significant overlap ( $p < 0.001$ ), despite differences in isolation and culturing conditions, inter-species variability and the high inter-patient variability in humans (Fig. 5C). At the level of pathways, this significant overlap becomes higher, with approximately one third of the pathways enriched in mouse SFs being also enriched in human SFs ( $p < 0.001$ ); key pathways identified above, such as innate immune and differentiation pathways, were found to be shared.

Eight genes that were validated in mouse SFs and represent key pathways, such as immune response and differentiation, shared by mouse and human, were also examined by qRT-PCR in human SFs (Fig. 5D). Five out of them were significantly deregulated in these samples, namely *MMP3* (up-regulated), and *LECT1*, *WNT16*, *PPARG* and *SP7* (down-regulated). *NOD2* (an up-regulated DEG) was also found to



differ significantly upon TNF stimulation, as its baseline levels are very low in the human (data not shown).

For a comparison between mouse and human at the level of the methylome, we exploited two studies reporting human SF data (9,31). We compared the intersection of changes reported by these studies with our mouse methylation data (Fig. 5E). As can be seen, approximately half of the genes identified in the human are also found in the mouse, a highly significant overlap ( $p < 0.001$ ).

Our findings contribute towards a systems-level validation of the human TNF-transgenic mouse as a model of the human disease, highlighting cell type-specific commonalities in key genes and pathways.

## DISCUSSION

In light of recent controversies on whether animal models accurately reflect human disease (32,33), we re-approached one of the key mouse models of rheumatoid arthritis, the human TNF-transgenic mouse, so as to achieve a high-resolution profile of the activated synovial fibroblast and compare it to human disease. In this disease-driving cell type, arthritogenicity is mirrored on distinct expression profiles characterised by early-onset and persistent changes, indicating that, even before the joint has been extensively damaged, pathogenic molecular processes are already at work.

Functionally, continuous up-regulation of gene expression was connected with the innate immune response, highlighting the role of the synovial fibroblast as an innate immune sensor (34–36). Down-regulation was linked to functions related to cell

differentiation, with two out of the three identified central transcriptional regulators, namely CEBPA and PPARG, being involved in mesenchymal differentiation (37,38); the latter was also validated here in human RA SFs by quantitative RT-PCR.

Given the unclear relationship between mesenchymal stem cells and RA SFs and the importance of differentiation processes in disease pathogenesis (39,40), the aspect of synovial fibroblast ontogeny in physiology and pathology merits further exploration and presents a good case of how omics data can be the starting point for more focused studies.

To date, the most studied epigenetic change in RA has been DNA methylation. The examination of our data from the perspective of the gene body, which represents the majority of genic space, reveals a hypo-methylated profile, in agreement with previous publications in human RA SFs (12). Genome-wide correlations between DNA methylation and gene expression are hampered by the complexity of the former as an epigenetic phenomenon. Nevertheless, we have been able to link differential methylation, most abundantly at the gene body, with approximately a quarter of our identified expression changes. In concert with the above-described functional highlights, among the functions enriched in differentially-methylated and expressed genes, we also find highly significant examples of mesenchymal differentiation processes and the innate immune response. *Sox10*, through methylation-dependent transcription is implicated to suppress stemness by binding to beta catenin, and is also expressed by a novel rat synovial stem cell population capable to give rise to mesenchymal lineages (41,42). Moreover, this is the first report of methylation-modulated expression of TLR2, a known innate immune sensing receptor in SFs, as observed in other inflammatory conditions (34,43). It is

also the first time that increased expression of podoplanin (*Pdpn*), which is well-documented in human arthritic synovium, is linked with changes in promoter methylation, similar to other cells of mesenchymal origin (44–46).

Furthermore, this work represents the first attempt to explore H3K4 trimethylation in SFs. Their arthritogenicity was shown to be defined by activated chromatin, making the promoters of pathogenic genes highly active and ready to drive disease-promoting transcription. Interestingly, in another innate immune cell type, the macrophage, it is known that TLR signalling induces H3K4 trimethylation so as to facilitate rapid transcription of target genes (47). This histone mark has also been reported to play a role in mesenchymal differentiation (48). Even if genome-wide experiments cannot resolve the causal link between histone modifications and gene expression changes, time-dependent approaches as the one presented here could provide the basis for more sophisticated models addressing this issue.

Bioinformatic probing of the present Tg197 mouse RNA-seq data by functional enrichment analysis against the human-specific dbGaP database (49) revealed that the most significant term returned in all three disease stages is “rheumatoid arthritis”, a finding that, when combined with the overall high similarity between mouse and human data presented here, reinforces the value of this model. Despite its TNF dependency and the observed overlap between the Tg197 profiles and those identified upon human TNF treatment, the latter failed to include many gene deregulations, especially of the established and late stages of disease. More importantly, the mouse pathways that collectively account for the resemblance to human disease are significantly enriched only in the Tg197 profiles highlighting the importance of analyses deriving from *in vivo* contexts.

To our knowledge this is the first cell type-specific multi-level characterisation of a mouse model of RA in alignment with human disease. With our current findings, we posit that, rather than comparing mixed and often secondarily-activated material from animal models and humans, careful validation of primarily disease-associated cell types, such as SFs in RA, should be ensured before concluding that an animal model mimics or not the human disease and to what extent. The present compendium of data can serve as the foundation for a range of future studies: from the basic mechanistic investigation of disease pathogenesis to the translational examination of novel biomarkers and the preclinical testing of therapeutics through the comparative alignment of readouts against reference profiles presented in this work.

### **Acknowledgements**

We would like to thank the InfrafrontierGR infrastructure (co-funded by the European Regional Development Fund and Greek NSRF 2007-2013) for mouse-hosting facilities. Also, we are grateful to Mr Spyros Lalos for histology, Dr Vaggelis Harokopos at the Genomics Facility, and Dr Piyi Papadaki for critical reading of this manuscript.

### **Data availability**

Data have been deposited in the GEO database under the following accession numbers: GSE95251, GSE95252, GSE95254 and GSE95256.

## REFERENCES

1. Keffer J, Probert L, Cazaris H, Georgopoulos S, Kaslaris E, Cazaris H, et al. Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis. *EMBO J* 1991;10:4025–31.
2. Armaka M, Apostolaki M, Jacques P, Kontoyiannis DL, Elewaut D, Kollias G. Mesenchymal cell targeting by TNF as a common pathogenic principle in chronic inflammatory joint and intestinal diseases. *J Exp Med* 2008;205:331–7.
3. Pap T, Müller-Ladner U, Gay RE, Gay S. Fibroblast biology. Role of synovial fibroblasts in the pathogenesis of rheumatoid arthritis. *Arthritis Res* 2000;2:361–7.
4. Ospelt C, Neidhart M, Gay RE, Gay S. Synovial activation in rheumatoid arthritis. *Front Biosci* 2004;9:2323–34.
5. Bartok B, Firestein GS. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol Rev* 2010;233:233–55.
6. Aidinis V, Plows D, Haralambous S, Armaka M, Papadopoulos P, Kanaki MZ, et al. Functional analysis of an arthritogenic synovial fibroblast. *Arthritis Res Ther* 2003;5:R140-57.
7. Lefèvre S, Knedla A, Tennie C, Kampmann A, Wunrau C, Dinser R, et al. Synovial fibroblasts spread rheumatoid arthritis to unaffected joints. *Nat Med* 2009;15:1414–20.
8. Klein K, Gay S. Epigenetic modifications in rheumatoid arthritis, a review. *Curr Opin Pharmacol* 2013;13:420–5.
9. la Rica L de, Urquiza JM, Gómez-Cabrero D, Islam ABMMK, López-Bigas N, Tegnér J, et al. Identification of novel markers in rheumatoid arthritis through integrated analysis of DNA methylation and microRNA expression. *J Autoimmun* 2013;41:6–16.

10. Nakano K, Whitaker JW, Boyle DL, Wang W, Firestein GS. DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis* 2013;72:110–7.
11. Whitaker JW, Shoemaker R, Boyle DL, Hillman J, Anderson D, Wang W, et al. An imprinted rheumatoid arthritis methylome signature reflects pathogenic phenotype. *Genome Med* 2013;5:40.
12. Karouzakis E, Gay RE, Michel B a, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum* 2009;60:3613–22.
13. Pandis I, Ospelt C, Karagianni N, Denis MC, Reczko M, Camps C, et al. Identification of microRNA-221/222 and microRNA-323-3p association with rheumatoid arthritis via predictions using the human tumour necrosis factor transgenic mouse model. *Ann Rheum Dis* 2012;71:1716–23.
14. Armaka M, Gkretsi V, Kontoyiannis D, Kollias G. A standardized protocol for the isolation and culture of normal and arthritogenic murine synovial fibroblasts. *Protoc Exch* 2009.
15. Frank-Bertoncelj M, Trenkmann M, Klein K, Karouzakis E, Rehrauer H, Bratus A, et al. Epigenetically-driven anatomical diversity of synovial fibroblasts guides joint-specific fibroblast functions. *Nat Commun* 2017;8:14852.
16. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
18. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 2014;47:11.12.1-11.12.34.
19. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive

and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;14:128.

20. Lund J. Statistical significance of the overlap between two groups of genes.

Available at: [http://nemates.org/MA/progs/overlap\\_stats.html](http://nemates.org/MA/progs/overlap_stats.html).

21. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.

22. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.

23. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11.

24. Heberle H, Meirelles GV, Silva FR da, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 2015;16:169.

25. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. Available at: <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

26. Chouvardas P, Kollias G, Nikolaou C. Inferring active regulatory networks from gene expression data using a combination of prior knowledge and enrichment analysis. *BMC Bioinformatics* 2016;17 Suppl 5:181.

27. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.

28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler

transform. *Bioinformatics* 2009;25:1754–60.

29. Stempor P. seqplots: An interactive tool for visualizing NGS signals and sequence motif densities along genomic features using average plots and heatmaps. 2014.

Available at: <http://przemol.github.io/seqplots/>.

30. Shen L, Shao N-Y, Liu X, Maze I, Feng J, Nestler EJ. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS One* 2013;8:e65598.

31. Nakano K, Whitaker JW, Boyle DL, Wang W, Firestein GS. DNA methylome signature in rheumatoid arthritis. *Ann Rheum Dis* 2013;72:110–7.

32. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker H V, Xu W, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* 2013;110:3507–12.

33. Takao K, Miyakawa T. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* 2015;112:1167–72.

34. Pierer M, Rethage J, Seibl R, Lauener R, Brentano F, Wagner U, et al. Chemokine secretion of rheumatoid arthritis synovial fibroblasts stimulated by Toll-like receptor 2 ligands. *J Immunol* 2004;172:1256–1265.

35. Sacre S, Lo A, Gregory B, Stephens M, Chamberlain G, Stott P, et al. Oligodeoxynucleotide inhibition of Toll-like receptors 3, 7, 8, and 9 suppresses cytokine production in a human rheumatoid arthritis model. *Eur J Immunol* 2016;46:772–781.

36. Brentano F, Schorr O, Gay RE, Gay S, Kyburz D. RNA released from necrotic synovial fluid cells activates rheumatoid arthritis synovial fibroblasts via Toll-like receptor 3. *Arthritis Rheum* 2005;52:2656–2665.



37. Chen W, Zhu G, Hao L, Wu M, Ci H, Li Y-P. C/EBP $\alpha$  regulates osteoclast lineage commitment. *Proc Natl Acad Sci U S A* 2013;110:7294–9.
38. Rosen ED. The transcriptional basis of adipocyte development. *Prostaglandins Leukot Essent Fatty Acids* 2005;73:31–4.
39. El-Jawhari JJ, El-Sherbiny YM, Jones EA, McGonagle D. Mesenchymal stem cells, autoimmunity and rheumatoid arthritis. *QJM* 2014;107:505–14.
40. Bari C de. Are mesenchymal stem cells in rheumatoid arthritis the good or bad guys? *Arthritis Res Ther* 2015;17:1–9.
41. Lee BL-P, Tang Z, Wang A, Huang F, Yan Z, Wang D, et al. Synovial stem cells and their responses to the porosity of microfibrous scaffold. *Acta Biomater* 2013;9:7264–75.
42. Tong X, Li L, Li X, Heng L, Zhong L, Su X, et al. SOX10, a novel HMG-box-containing tumor suppressor, inhibits growth and metastasis of digestive cancers by suppressing the Wnt/ $\beta$ -catenin pathway. *Oncotarget* 2014;5:10571–83.
43. Benakanakere M, Abdolhosseini M, Hosur K, Finoti LS, Kinane DF. TLR2 Promoter Hypermethylation Creates Innate Immune Dysbiosis. *J Dent Res* 2014;94:183–191.
44. Hantusch B, Kalt R, Krieger S, Puri C, Kerjaschki D. Sp1/Sp3 and DNA-methylation contribute to basal transcriptional activation of human podoplanin in MG63 versus Saos-2 osteoblastic cells. *BMC Mol Biol* 2007;8:20.
45. Ekwall A-KH, Eisler T, Anderberg C, Jin C, Karlsson N, Brisslert M, et al. The tumour-associated glycoprotein podoplanin is expressed in fibroblast-like synoviocytes of the hyperplastic synovial lining layer in rheumatoid arthritis. *Arthritis Res Ther* 2011;13:R40.
46. Croft AP, Naylor AJ, Marshall JL, Hardie DL, Zimmermann B, Turner J, et al.

Genomic alignments of mouse and human arthritogenic synovial fibroblasts.

Rheumatoid synovial fibroblasts differentiate into distinct subsets in the presence of cytokines and cartilage. *Arthritis Res Ther* 2016;18:270.

47. Kayama H, Ramirez-Carrozzi VR, Yamamoto M, Mizutani T, Kuwata H, Iba H, et al.

Class-specific regulation of pro-inflammatory genes by MyD88 pathways and IkappaBzeta. *J Biol Chem* 2008;283:12468–77.

48. Matsumura Y, Nakaki R, Inagaki T, Yoshida A, Kano Y, Kimura H, et al.

H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol Cell* 2015;60:584–596.

49. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet* 2007;39:1181–6.

## FIGURE LEGENDS

**Figure 1** Arthritogenic synovial fibroblasts are deregulated at the level of the transcriptome across disease stages in the Tg197 mouse model of RA. RNA-seq profile of Tg197 SFs as compared to wild-type (WT) controls at three different disease stages (early: 3 weeks; established: 8 weeks; and late: 11 weeks). FC = fold change; PV = corrected *p* value. (A) Volcano plots depicting differentially-expressed genes (DEGs) at the three stages of disease. (B) Venn diagrams showing overlap of significant DEGs among disease stages. (C) Heat map showing clustering of significant DEGs across disease stages. Genes were divided in six clusters, each showing a distinct pattern of deregulation as shown in the right hand-side graph. Functional analysis was performed for these clusters; shown (left) are descriptors summarising the top 10 gene ontology (biological process; GO-BP) terms (ranked by *q* value). (D) Heat map showing clustering of the top 25 (as ranked by *q* value) significant GO-BP terms associated with each disease stage. (E) The common transcriptional network of the three disease stages, as identified by RNEA. Colour bar indicates node centrality; circle size indicates number of edges from the node, i.e. number of targets.

**Figure 2** Global DNA methylation changes constitute part of the arthritogenic profile of the Tg197 synovial fibroblast. MethylCap-seq profile of Tg197 SFs as compared to WT controls at the three different disease stages. FC = fold change; PV = corrected *p* value. (A) Volcano plots depicting differentially-methylated regions (DMRs) at the three stages of disease in either promoters or gene bodies. (B)

Distribution of significant DMRs in promoters and gene regions, as well as in relation to CpG islands, across disease stages. *y* axes represent relative enrichment; CGI: CpG island. (C) Expression data of the three disease stages displayed in four quantiles against DNA methylation signal around the transcription start site (TSS) up to the transcription stop site (TES). (D) Pie charts showing percentages of significant DEGs that are also significantly differentially-methylated (DM) in each disease stage. Expected DM = expression correlated positively with gene body or inversely with promoter methylation.

**Figure 3** Activated synovial fibroblasts are characterised by the presence of open chromatin. H3K4me3-seq profile of Tg197 SFs as compared to WT controls at the three different disease stages. FC = fold change; PV = corrected *p* value. (A) Volcano plots depicting differentially H3K4-trimethylated regions in promoters at the three stages of disease. (B) Distribution of significant differentially H3K4-trimethylated regions in promoters, across disease stages. (C) Expression data displayed in four quantiles against H3K4me3 signal around the transcription start site (TSS). (D) Scatter plots showing correlation between significant H3K4 trimethylation (*y* axis) and expression (*x* axis).

**Figure 4** TNF has a pronounced contribution in the identified signatures of the Tg197 synovial fibroblast. RNA-seq, MethylCap-seq and H3K4me3-seq were performed in WT SFs treated with 10 ng/ml human TNF for 24 h and compared to the untreated controls. FC = fold change; PV = corrected *p* value. (A) Volcano plots depicting changes induced by TNF at the levels of transcriptome, methylome and

active chromatin. (B) Venn diagrams showing overlap of significant deregulations either at the level of genes (RNA) / regions (DNA methylation and H3K4 trimethylation) or at the level of KEGG pathways (RNA), among the three different stages of Tg197 disease and treatment with human TNF.

**Figure 5** Validation of selected deregulations and comparisons between mouse and human data. FC = fold change; PV = corrected *p* value. (A) Scatter plots showing correlation between mouse RNA-seq and qRT-PCR data. (B) Volcano plot depicting significantly deregulated genes in human RA SFs in comparison to healthy SFs as identified by RNA-seq. (C) Venn diagrams showing overlap between significantly up- or down-regulated genes or their respective KEGG pathways in mouse and human RNA signatures. Mouse data represent the union of deregulated genes across Tg197 disease stages. (D) qRT-PCR validations in a panel of genes in human SFs. *y* axis shows  $\Delta C_t(\text{gene of interest} - HPRT1)$ . \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . (E) Venn diagrams showing overlap in differentially methylated genes between mouse SFs (our data) and the intersection of human SF deregulations reported by two studies (9,10).

**Supplementary Figure 1** Pathology of Tg197 disease stages. Representative photomicrographs of ankle joint sections of wild-type (at 8 weeks; A,B) and arthritic mice at three stages of disease: 3 weeks (early; C,D); 8 weeks (established; E,F); and 11 weeks (late; G,H). Haematoxylin-eosin (left) and toluidine blue (right) staining. Error bars = 200  $\mu\text{m}$ . Note that in the early stage there is near absence of pathology; in the established stage we observe the histological hallmarks of disease (pannus

formation with synovial hyperplasia, inflammatory infiltration and erosions to the bone and cartilage); finally, in the late stage, there is total distortion of tissue architecture, with extensive bone loss and cartilage destruction and high immune cell infiltration.

**Supplementary Figure 2** Experimental scheme and sequencing statistics. (A) Drawing illustrating experimental set-up of this project. (B) Quality scores attained in (1) RNA, (2) MethylCap and (3) H3K4me3 sequencing. (C) Table summarising sequencing depth statistics for each biological replicate and sequencing effort.

**Supplementary Figure 3** The transcriptional networks of the three disease stages, as identified by RNEA. (A) Early; (B) established; (C) late. Colour bar indicates node centrality; circle size indicates number of edges from the node, i.e. number of targets. The top five genes as ranked by node centrality are named in each network.

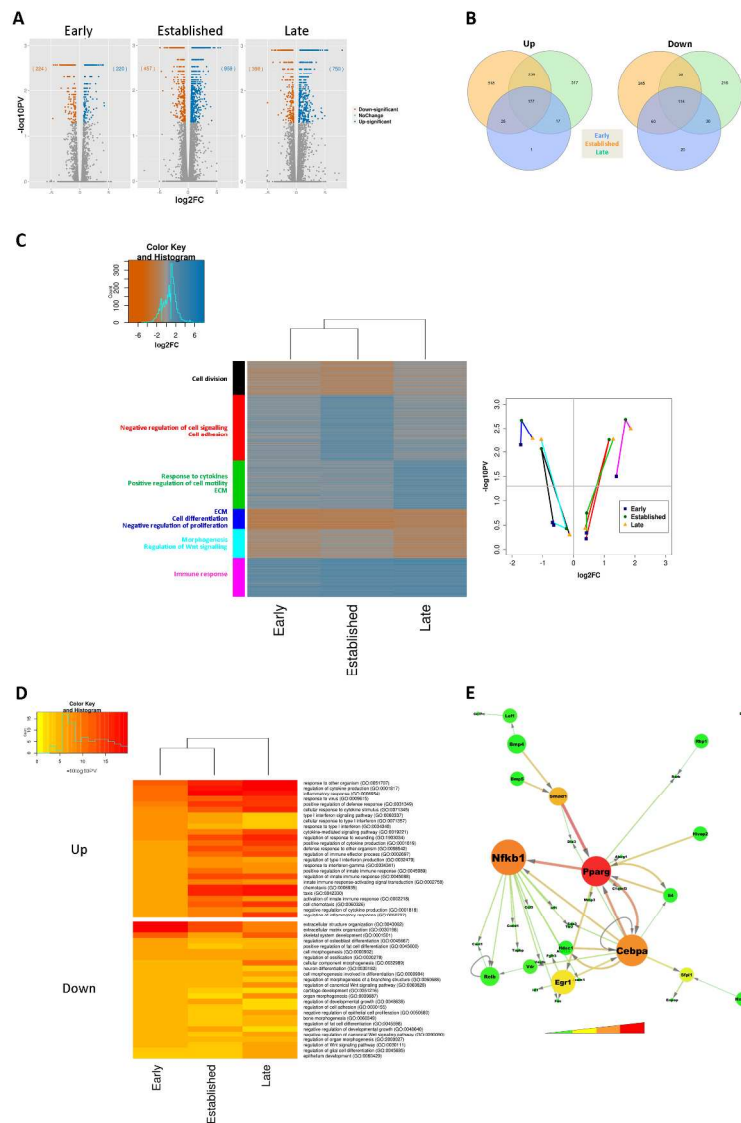


Figure 1  
Figure 1  
209x297mm (300 x 300 DPI)

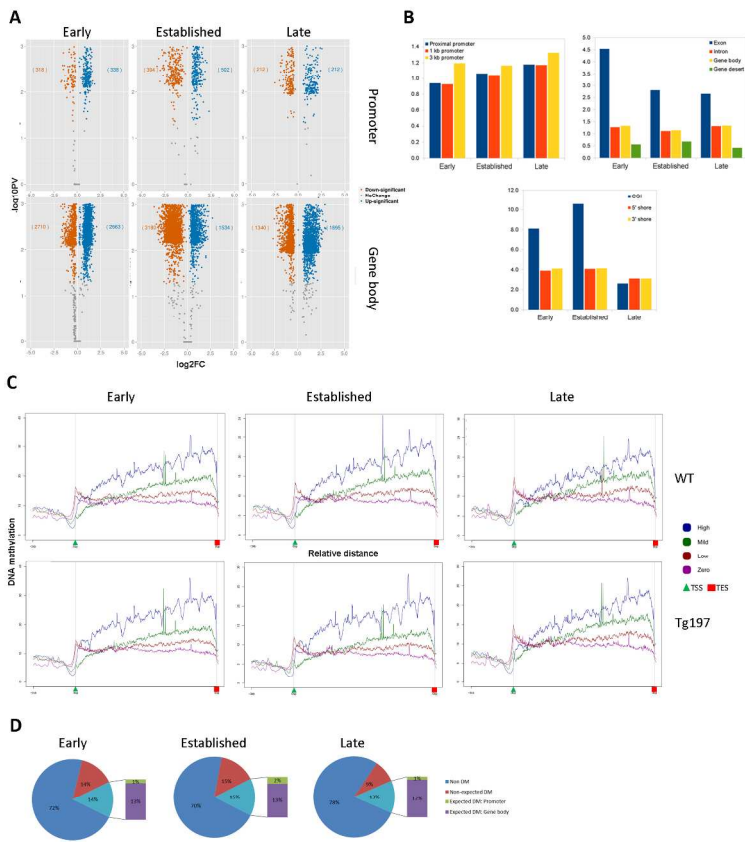


Figure 2  
Figure 2  
209x297mm (300 x 300 DPI)



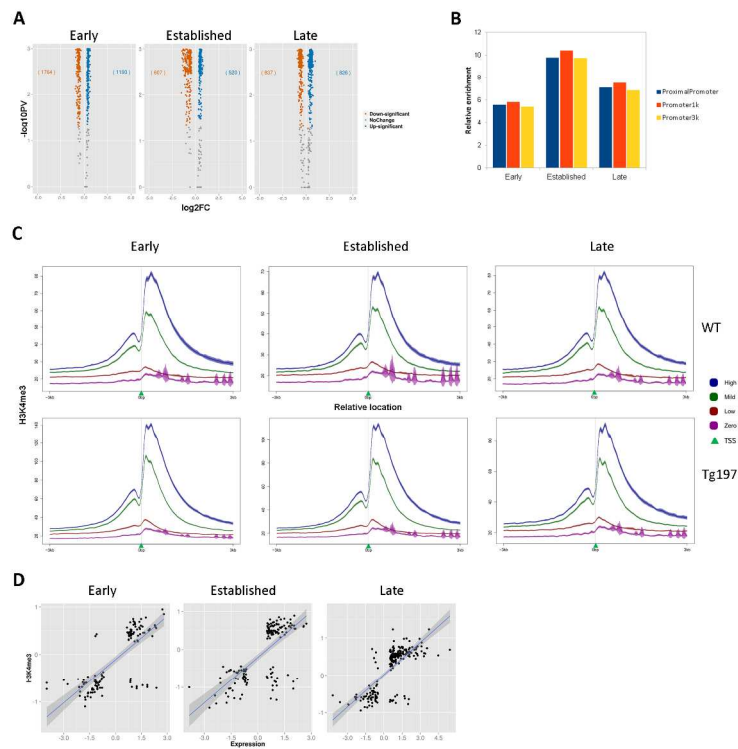


Figure 3  
Figure 3  
209x297mm (300 x 300 DPI)

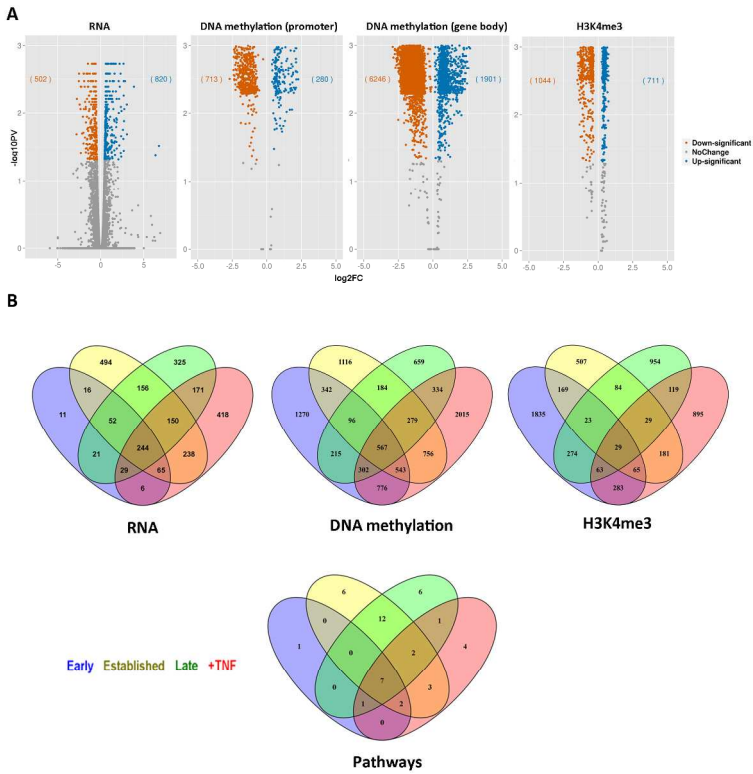


Figure 4  
Figure 4  
209x297mm (300 x 300 DPI)

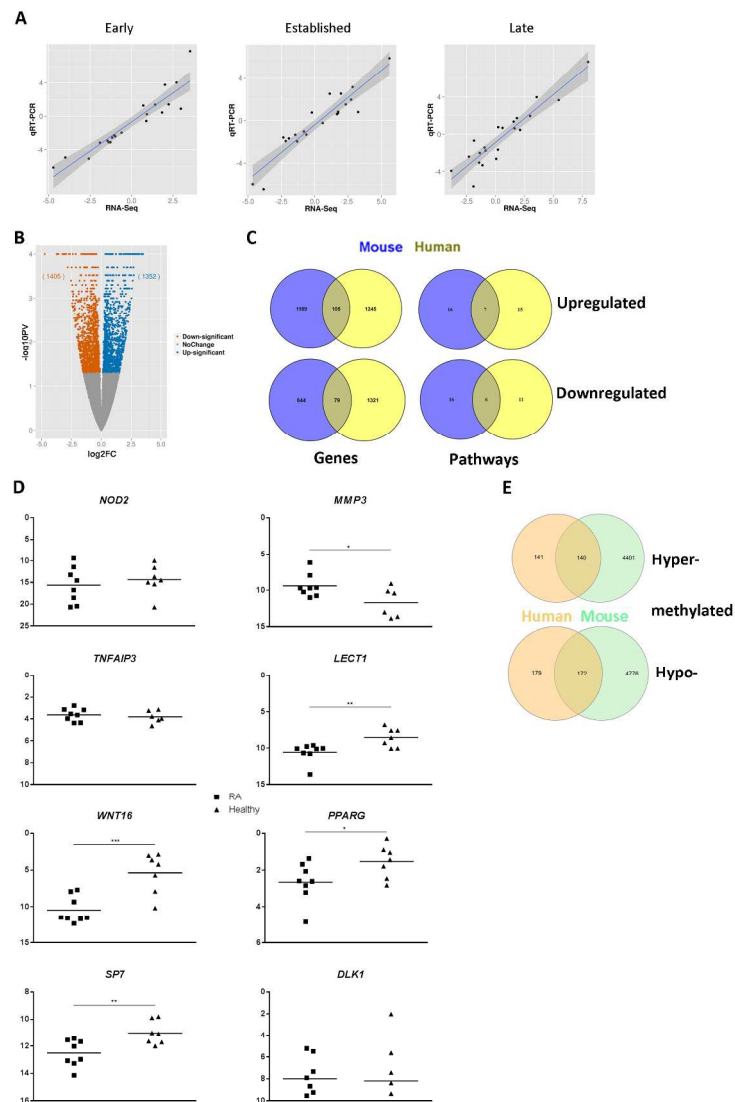
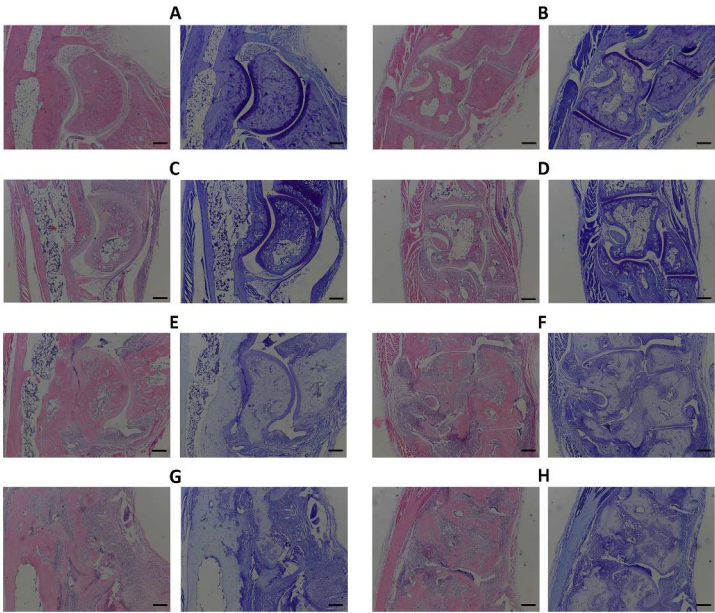
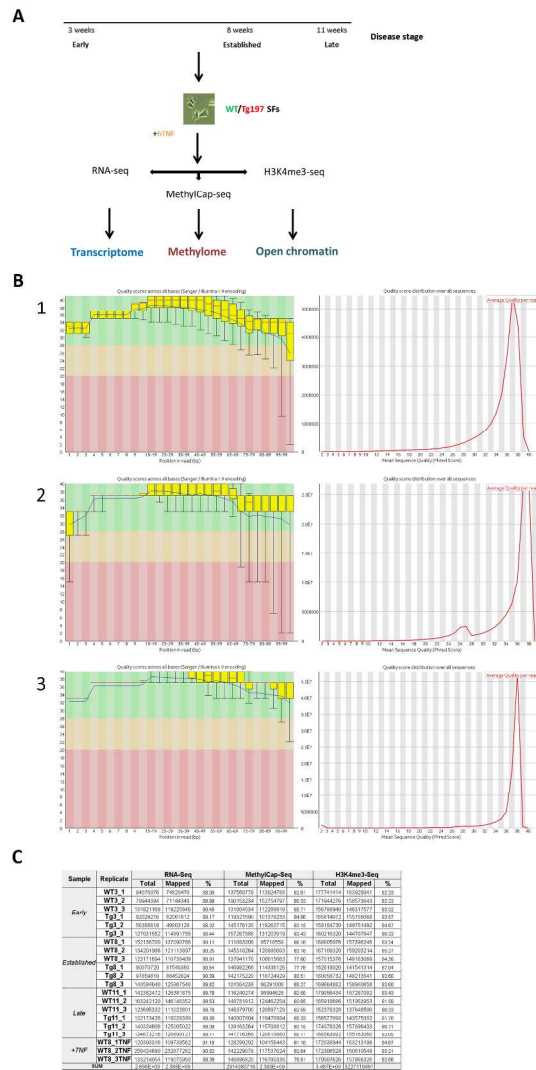


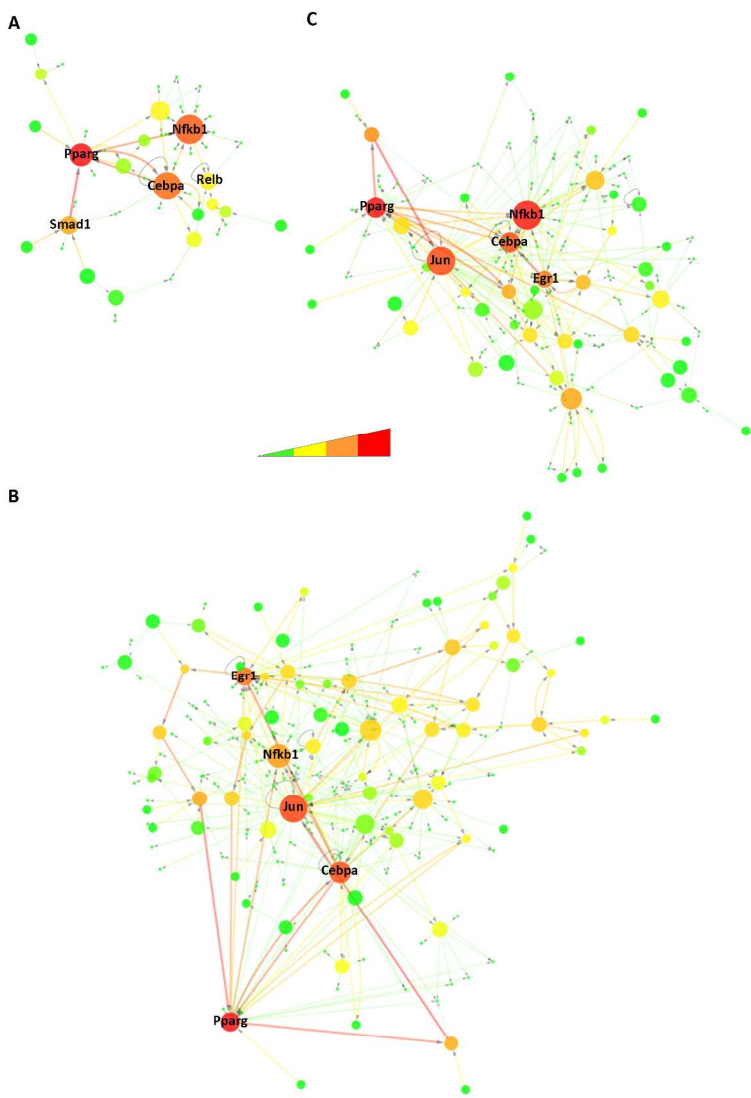
Figure 5  
Figure 5  
209x297mm (300 x 300 DPI)



209x297mm (300 x 300 DPI)



209x297mm (300 x 300 DPI)



209x297mm (300 x 300 DPI)

AC

Sample type	Diagnosis	Sex	Age (years)	CRP (mg/l)	Disease duration (years)	Treatment
RA	RA	F	71	17	51	2 cDMARDs, corticosteroid
RA	RA	F	54	47	16	1 cDMARD, 1 bDMARD, corticosteroid
RA	RA	F	64	<8	7	1 cDMARD, 1 bDMARD
RA	RA	F	52	<8	20	2 cDMARDs, corticosteroid
RA	RA	M	61	<8	9	1 cDMARD, 1 bDMARD, corticosteroid
RA	RA	F	70	26	<1	-
RA	RA	M	46	<8	3	-
Healthy	Arthralgia	M	23	NA		
Healthy	Arthralgia	M	44	NA		
Healthy	Arthralgia	F	49	NA		
Healthy	Arthralgia	M	42	NA		
Healthy	Arthralgia	F	38	NA		
Healthy	Arthralgia	F	46	NA		
Healthy	Arthralgia	M	34	NA		

CRP = C-reactive protein; RA = rheumatoid arthritis; F = female, M= male; NA = not assessed; cDMARD = conventional disease modifying anti-rheumatic drug, bDMARD = biological DMARD.

Gene	Forward primer	Reverse primer
<i>NOD2</i>	TTCTCCGGGTTGTGAAATGT	CTCCTCTGTGCCTGAAAAGC
<i>MMP3</i>	GGGCTATCAGAGGAAATGAG	CACGGTTGGAGGGAAACCTA
<i>TNFAIP3</i>	AAGGACAGTGGGCCTGAAATC	TTCCCCGGTCTCTGTTAACAAG
<i>LECT1</i>	CCTTATCATCAGGAAGGGGAAAG	GTGATCCAGTCTAGGGTCGAA
<i>WNT16</i>	TCACAGGGGCTTCTCAAAAGA	AGGAGGCAATGCCCAACC
<i>PPARG</i>	CGAGAGTCAGCCTTTAACGAAAT	ATGGCATCTCTGTGTCAACCA
<i>SP7</i>	CCCAGCAGCCCCGGA	CCCGCCGCCTCAGAAG
<i>DLK1</i>	CACCTATGGGGCTGAATGCT	AGAATCCATTTTGGGGGTTGC