Effects of rheumatoid arthritis associated transcriptional changes on osteoclast differentiation network in the synovium

Shilpa Harshan [1] , **Poulami Dey** [1, 2] , **Srivatsan Ragunathan** [Corresp. 1]

[1] Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

[2] Manipal Academy of Higher Education, Manipal, Karnataka, India

Corresponding Author: Srivatsan Ragunathan
Email address: srivatsan@ibab.ac.in

Background. Osteoclast differentiation in the inflamed synovium of rheumatoid arthritis affected joints leads to the formation of bone lesions. Reconstruction and analysis of protein interaction networks underlying specific disease phenotypes are essential for designing therapeutic interventions. In this study we have created a network that captures signal flow leading to osteoclast differentiation. Based on transcriptome analysis, we have indicated the potential mechanisms responsible for the phenotype in the rheumatoid arthritis affected synovium.

Method. We collected information on gene expression, pathways and protein interactions related to rheumatoid arthritis from literature and databases namely Gene Expression Omnibus, KEGG pathway and STRING. Based on these information, we created a network for the differentiation of osteoclasts. We identified the differentially regulated network genes and reported the signaling that are responsible for the process in the rheumatoid arthritis affected synovium.

Result. Our network reveals the mechanisms underlying the activation of the Neutrophil Cytosolic Factor complex in connection to osteoclastogenesis in rheumatoid arthritis. Additionally, the study reports the predominance of the canonical pathway of NF-κB activation in the diseased synovium. The network also confirms that the upregulation of T cell receptor signaling and downregulation of TGFβ signaling pathway favour osteoclastogenesis in Rheumatoid Arthritis. To the best of our knowledge, this is the first comprehensive protein-protein interaction network describing Rheumatoid Arthritis driven osteoclastogenesis in the synovium.

Discussion. This study provides information that can be used to build models of the signal flow involved in the process of osteoclast differentiation. The models can further be used to design therapies to ameliorate bone destruction in the Rheumatoid Arthritis affected joints.

1 **Effects of Rheumatoid Arthritis associated transcriptional**

2 **changes on osteoclast differentiation network in the**

3 **synovium**

4 Shilpa Harshan[1, +], Poulami Dey[1, 2, +], Srivatsan Raghunathan[1]*

5 [1]Institute of Bioinformatics and Applied Biotechnology (IBAB), Biotech Park, Electronic City

6 Phase I, Bengaluru 560 100, Karnataka, India.

7 [2] Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India.

8 * srivatsan@ibab.ac.in

9 + These authors contributed equally to this work

Abstract

Background

Osteoclast differentiation in the inflamed synovium of Rheumatoid Arthritis affected joints leads to the formation of bone lesions. Reconstruction and analysis of protein interaction networks underlying specific disease phenotypes are essential for designing therapeutic interventions. In this study we have created a network that captures signal flow leading to osteoclast differentiation. Based on transcriptome analysis, we have indicated the potential mechanisms responsible for the phenotype in the Rheumatoid Arthritis affected synovium.

Method

We collected information on gene expression, pathways and protein interactions related to Rheumatoid Arthritis from literature and databases namely Gene Expression Omnibus, KEGG pathway and STRING. Based on these information, we created a network for the differentiation of osteoclasts. We identified the differentially regulated network genes and reported the signaling that are responsible for the process in the Rheumatoid Arthritis affected synovium.

Result

Our network reveals the mechanisms underlying the activation of the Neutrophil Cytosolic Factor complex in connection to osteoclastogenesis in Rheumatoid Arthritis. Additionally, the study reports the predominance of the canonical pathway of NF-κB activation in the diseased synovium. The network also confirms that the upregulation of T cell receptor signaling and downregulation of TGFβ signaling pathway favour osteoclastogenesis in Rheumatoid Arthritis.

30   To the best of our knowledge, this is the first comprehensive protein-protein interaction network

31   describing Rheumatoid Arthritis driven osteoclastogenesis in the synovium.

32   **Discussion**

33   This study provides information that can be used to build models of the signal flow involved in

34   the process of osteoclast differentiation. The models can further be used to design therapies to

35   ameliorate bone destruction in the Rheumatoid Arthritis affected joints.

36   # Introduction

37   Rheumatoid arthritis (RA) is a systemic autoimmune disease that primarily affects synovial

38   joints. The disease is characterised by chronic inflammation in the joints, leading to synovial

39   hyperplasia (pannus formation), destruction of the cartilage and erosion of the underlying bone.

40   RA is a complex disease involving several molecular pathways across various cell types and

41   tissues. Thus in order to elucidate the underlying cause of a particular phenotype associated to

42   the disease, identification of the network consisting of differentially expressed genes (DEG) in

43   the interacting pathways is essential. Studies have used pathway analysis to identify affected

44   pathways from lists of DEGs (Hao et al., 2017; Wang et al., 2017; Lee et al., 2011; Wu et al.,

45   2010). The lists have also been used to create networks that are related to specific diseases or

46   conditions. Earlier work using RA samples has focused on generating networks of the genes

47   showing differential regulation (Hao et al., 2017; Wang et al., 2017) or the most enriched gene

48   ontology (GO) ("Expansion of the Gene Ontology knowledgebase and resources.," 2017)

49   category in the DEG lists (Lee et al., 2011). A comprehensive network describing molecular

50   interactions across various RA affected tissues was created using publicly available microarray

51   data by Wu *et al* (Wu et al., 2010). Other groups have created gene regulatory networks (GRN)

52   using *in vitro* data from cultured fibroblasts and macrophages (Kupfer et al., 2014; You et al.,

53   2014). Kupfer *et al* (Kupfer et al., 2014) used time series data generated from RA synovial

54   fibroblasts subjected to external stimulation to create a GRN. They simulated the network to

55   analyse the behaviour of genes involved in RA pathogenesis, in response to stimulation by RA

56   associated cytokines and growth factors. You *et al* (You et al., 2014) created a GRN and

57   identified the critical interactions responsible for synovial fibroblast invasiveness in RA

58   synovium. The creation of a detailed protein-protein interaction (PPI) network describing the

59   connections between various pathways involved in any specific RA process, at the level of the

60   synovial tissue, is yet to be attempted. In this study, using the publicly available gene expression

61   data for RA synovial tissue and protein interactions and pathway databases, we created and

62   analysed a detailed phenotype-specific PPI network. We used differentially regulated genes to

63   identify the altered pathways in the affected synovium. We identified the pathway of osteoclast

64   differentiation as a phenotype connected to many of the altered pathways in the RA synovium. It

65   is established that the RA synovium harbors osteoclasts, the cells responsible for bone

66   degradation in the affected joints (Schett, 2007). Therefore, a network of proteins participating in

67   the interacting pathways underlying the RA associated process of osteoclast differentiation in the

68   synovium was created for the first time. We report the upregulated signaling routes that drive

69   osteoclastogenesis via the generation of reactive oxygen species (ROS) by Neutrophil Cytosolic

70   Factor (NCF) complex in the RA synovium. We demonstrate the contribution of elevated T cell

71   receptor signaling in facilitating osteoclast differentiation in the affected tissue. In addition, we

72   describe the importance of the canonical pathway of NF-κB activation and the TGFβ pathway in

73   connection to the process. Finally, the network reports all the possible routes by which the

74   inflamed synovium promotes the differentiation of osteoclasts.

Materials and Methods

This study involved two major steps: selection of a phenotype exhibited by the RA synovium, and construction and analysis of a PPI network for the selected phenotype. Figure 1 shows the detailed workflow that was followed. Each step is described in detail in this section. The databases used in this study are summarized in Table 1.

Identification of DEGs using microarray data analysis

The DEGs were obtained by re-analysing the publicly available microarray datasets in Gene Expression Omnibus (GEO) (Edgar, Domrachev & Lash, 2002) database. The repository was searched for the data generated from synovial tissue in RA patients and healthy controls. The results were further narrowed down by considering only the data from Affymetrix platforms with at least four RA and four control samples. Datasets selected for the study are mentioned in Table 2.

Of the seven datasets, information regarding treatments received by the patients was not available for GSE77298 (Broeren et al., 2016) and GSE7307. Earlier, it was established that the differential regulation of the genes in these datasets was not under the influence of drug therapy (Dey, Panga & Raghunathan, 2016). The clinical information for the RA patients was available for the datasets GSE1919 (Ungethuem et al., 2010), GSE12021 (U133A) (Huber et al., 2008), GSE12021 (U133B) (Huber et al., 2008), GSE55235 (Woetzel et al., 2014) and GSE55457 (Woetzel et al., 2014). The erythrocyte sedimentation rate (ESR) and the concentration of C-reactive protein (CRP) reported for these datasets were higher than 40 mm hour^{-1} and 21 mg litre^{-1} respectively. These values indicated active inflammation in the synovium of the RA patients

96 (Wetteland et al., 1996 ; Otterness, 1994). For the datasets, GSE77298 and GSE 7307, the values

97 for these parameters were not available.

98 Raw data from the seven datasets along with their metadata was downloaded using the R

99 libraries GEOquery (Davis & Meltzer, 2007) and GEOmetadb (Zhu et al., 2008). The data was

100 analysed using the affy (Gautier et al., 2004) and simpleaffy (Wilson & Miller, 2005) libraries in

101 the Bioconductor package in R (R Core Team, 2017).

102 In this analysis, two algorithms, Robust Multiarray Average (RMA) and Microarray Suite 5.0

103 (MAS5) were used for the data normalization. The choice of data normalization algorithms affect

104 the final selection of the DEGs (Pepper et al., 2007). In order to reduce the algorithm specific

105 effects, both RMA and MAS5 were used in this study. In the case of MAS5, probesets having at

106 least one present call ("P") in control as well as treatment samples were considered. Probesets

107 were annotated with Entrez IDs using the Bioconductor as well as DAVID gene-ID conversion

108 tool (Huang, Sherman & Lempicki, 2009b; Huang, Sherman & Lempicki, 2009a). Welch t test

109 was applied to calculate the significance for differential expression between the RA and the

110 control samples. As per the recommendations by Huang *et al* (Huang, Sherman & Lempicki,

111 2009a), in our study a gene with a linear fold change of 2 (for up and downregulation) and a p

112 value <= 0.05 was considered to be differentially expressed. A final list of DEGs, from the seven

113 datasets, was obtained using the selection rules as described below:

114   A. In a dataset, a gene is considered to be upregulated if:

115   i.   It is upregulated in both RMA and MAS5 [+,+]

116   Or,

117   ii.   It is upregulated in one of the algorithm and not differentially expressed in the other

118   [+, 0] or [0, +].

119   B. Across the datasets, the gene is upregulated if:

120   i.   It is [+, +] in at least one of the seven datasets and no downregulation in any of the

121   datasets

122   Or,

123   ii.   It is [+, 0], in at least one dataset and [0, +] in at least one of the remaining datasets

124   while there is no downregulation in any one of them

125   The same procedure was repeated for the downregulated genes.

126   When the selection criteria was made more stringent by demanding the selection of a gene in at

127   least two datasets, the number of selected genes reduced by almost 50 percent as shown in Fig. 2.

128   Since the aim of the study was to identify all the important signaling associated with the RA-

129   associated process we decided to proceed with the selection criteria of presence in at least one

130   dataset.

131   Finally, we prepared a list of up and downregulated genes which we named as "common-up" and

132   "common-down" respectively.

133   **Pathway Analysis:**

134   The common-up and common-down gene lists were separately examined for the enrichment of

135   pathways listed in the Kyoto Encyclopedia Of Genes And Genomes (KEGG) database (Kanehisa

136   et al., 2017) using Database for Annotation, Visualization and Integrated Discovery (DAVID).

137   For the enrichment analysis, we created a custom background by combining the total probesets

138   present on all four microarray platforms and annotating them with Entrez IDs using the DAVID

139   gene ID conversion tool. The pathways which were significantly over-represented in the

140   common-up or common-down gene lists with an EASE score <= 0.05 and fold enrichment >=

141   1.5 were considered to be affected in RA (Huang, Sherman & Lempicki, 2009a). EASE score is

142   a modified one-tailed Fisher exact probability used in enrichment analysis (Hosack et al., 2003).

143   The pathways were grouped according to their KEGG categories. Those belonging to the

144   categories "human diseases", "metabolism" or the ones lacking protein-protein interactions were

145   not considered further. The category "human diseases" contains pathways that represent specific

146   disease conditions. These were excluded because they do not reflect the protein-protein

147   interaction in healthy conditions. "Metabolism" contains pathways that describe interconversions

148   of metabolites. Since this study focused on PPI involved in the RA affected synovium, we did

149   not consider the metabolic pathways. The remaining pathways were categorized based on their

150   functional specificity. The pathways which result in a specific function like platelet activation

151   were tagged as process pathways, whereas the pathways describing more general signaling

152   events like the activation of multiple transcription factors through T cell receptor signaling, were

153   considered as signaling pathways.

154   The list of genes present in each selected pathway was downloaded from KEGG using the

155   KEGGREST (Tenenbaum, 2017) package in R. For each one of the selected pathway, the list of

156   genes that were common between the pathway and the microarray platforms was created. We

157   name this as "S-list". By the pairwise intersection of the S-list of each process pathway with

158   every one of the non-disease pathway, we obtained the DEGs shared between the pairs of

159   process and non-disease pathways. The number of DEGs shared by each pair was examined. The

160   pairs of pathways sharing at least five upregulated or five downregulated genes were retained for

161   our study. Figure 3 shows the number of up or downregulated genes common to the pair of

162   pathways. It is evident from this figure that the above mentioned criterion did not result in a bias

163   towards pairs of larger pathways.

164 **Construction of network:**

165 We constructed a PPI network for the osteoclast differentiation. The osteoclast differentiation

166 pathway (ODP) proteins, obtained from KEGG were defined as a set of core proteins. Interactors

167 of the core proteins (first-shell interactors) were extracted from the STRING database (version

168 10) (Szklarczyk et al., 2015). The protein list was restricted by considering only experimentally

169 validated interactions with a score of >= 0.9. This score on a scale of 0 to 1 represents the

170 confidence of experimental validation with maximum confidence being 1. We obtained the

171 directions for these interactions from the literature references used in STRING, when available,

172 or with a separate literature search in Pubmed. The complete network was built in Cytoscape

173 (Shannon et al., 2003) using all the obtained interactions. The proteins corresponding to the

174 DEGs in RA synovium obtained from the microarray data analysis were indicated in this

175 network.

176 **Analysis of the network:**

177 The network was a mixed network consisting of the undirected protein binding edges and the

178 directed edges of activation, inhibition or the post-translational modification (PTM). The nodes

179 are labelled using the official gene symbols corresponding to the proteins used to create the

180 network. The interactions involved undirected protein-protein interactions or directional PTM

181 like phosphorylation, methylation, acetylation, ubiquitination *etc.* We included activation or

182 inhibition as an interaction whenever the reference mentioned that the target protein is activated

183 or inhibited as a result of the interaction. We created a version of this directed network without

184 UBC and its edges. We named this version as "directed ODP network".

185   We conducted GO enrichment analysis on the directed ODP network proteins using the GO

186   molecular function (GOMF) and GO biological process (GOBP) terms. Terms with an EASE

187   score <= 0.05 and fold enrichment >=1.5 were considered as enriched. We combined 23 enriched

188   GOMF terms to identify the proteins that bind to DNA. In the case of the GOBP terms, we

189   selected the enriched signaling terms that contained differentially regulated genes for T cell

190   receptor signaling, B cell receptor signaling and FC-ε receptor signaling. For NF-κB signaling,

191   TLR signaling and TGFβ signaling pathways we combined six, nine and three enriched terms

192   respectively and examined their differential regulation. Using the selected enriched signaling

193   terms, we extracted subnetworks corresponding to each signaling pathway from the directed

194   ODP network. All the subnetworks demonstrate the flow of information from the first-shell

195   interactor proteins to the core proteins of the ODP network. The details of the GOMF and GOBP

196   signaling terms enriched in the analysis is provided in the Tables S1 and S2 respectively.

197   The database DrugBank (Wishart et al., 2018) was explored to locate the target proteins of drugs

198   that are commonly used in the treatment of RA. We used the information to pinpoint the network

199   proteins which are targets of the RA drugs. The details regarding the drugs and their targets are

200   submitted in Table S3.

201   We converted all the edges of the directed ODP network to single undirected protein binding

202   edges to create an "undirected ODP network". We analysed this network using the Cytoscape

203   plugin NetworkAnalyzer (Assenov et al., 2008). We used the plugin MCODE v1.5.1. (Bader &

204   Hogue, 2003) to identify the clusters in the network.

205   # Results

206   **Differentially expressed genes in RA synovium**

207   We analysed seven Affymetrix microarray datasets from five different platforms. Out of 21,246

208   Entrez annotated genes measured, 1018 upregulated and 893 downregulated genes were

209   identified in the RA synovium compared to the healthy controls. The differentially regulated

210   genes are submitted in Table S4. Only three genes STAT1, IL7R and IGKC were upregulated in

211   all seven datasets. Interestingly, three AP1 proteins FOSB, JUN and JUNB were downregulated.

212   Table 3 shows the datasets in which the genes were downregulated.

213   **Diverse pathways are involved in the disease processes affecting the RA synovium**

214   Using the DEGs from the microarray analysis as the input, we found that 52 KEGG pathways

215   were enriched in the upregulated gene list, and 29 in the downregulated gene list. The EASE

216   scores of the selected pathways were much less that the cut-off of 0.05. The enrichment analysis

217   when performed with the combined DEG list (1018 up genes + 893 down genes) resulted in only

218   55 pathways. Among these, only two pathways were newly obtained when compared to the

219   previous list of 52 upregulated and 29 downregulated pathways. As the combined analysis

220   proved less informative, the 52 up and 29 downregulated pathways were considered for the

221   study. Three pathways, namely, Extra cellular matrix (ECM)-receptor interaction, Focal adhesion

222   and Proteoglycans in cancer occurred in both the up and downregulated pathway lists because

223   each pathway had a significant number of up and downregulated genes. The KEGG category-

224   wise distributions of the enriched pathways are shown in Fig. 4 and Fig. 5 and the detailed

225   results of the pathway analysis are given in the Table S5 and S6. 26 of the upregulated pathways

226   and four of the downregulated pathways belonged to the KEGG category "human diseases".

227   Among the upregulated non-disease pathways, the category "immune system" had the highest

228   number of enriched pathways. Most of the immune receptor signaling pathways in this category

229   were upregulated. Among the other signaling pathways, NF-κB and JAK-STAT signaling

230   pathways were upregulated in our analysis. All the signaling pathways which belong to the

231   KEGG categories, "immune system" and "signal transduction" that were enriched in the

232   upregulated gene list are shown in Table 4.

233   Specialized cells called osteoclasts which facilitate bone resorption are also present in the

234   invading pannus of the RA joints (Gravallese et al., 1998; Jung et al., 2014; Nevius, Gomes &

235   Pereira, 2016). All the five microarray datasets which provided information on the disease state,

236   GSE1919, GSE12021 (U133A), GSE12021 (U133B), GSE55235 and GSE55457, used tissue

237   from patients with more than ten years of disease. Patients from the other dataset GSE77298,

238   were at the end stage of the disease. Since osteoclast differentiation is reported in severely

239   inflamed RA synovium, the process is likely to be detected in the synovial tissue used for the

240   datasets. It is noteworthy that our analysis identified the pathway osteoclast differentiation as one

241   of the enriched pathways with a fold enrichment of 3.11 (EASE score of 2.41 e^{-8}) in the RA

242   synovium. In addition, our analysis detected the upregulation of two well-known osteoclast

243   markers Cathepsin K (CTSK) and tartarate resistant acid phosphatase (ACP5) in the synovium.

244   The category highly represented in the list of downregulated pathways was "signal transduction".

245   All the signal transduction pathways enriched in the downregulated gene list are given in Table

246   5. In contrast to the upregulated pathways, the downregulated pathways included several

247   metabolic pathways such as fatty acid degradation, fatty acid elongation *etc.* Some endocrine

248   system pathways like regulation of lipolysis in adipocytes, insulin signaling pathway, which are

249   closely related to metabolic regulation were also listed among the downregulated pathways.

250   **RA affected signaling pathways interact to orchestrate osteoclast differentiation in the**

251   **synovium**

252   We categorised the 26 upregulated and 25 downregulated non-disease pathways based on their

253   functional specificity. In this analysis we identified 12 process pathways and 19 signaling

254   pathways among the differentially regulated pathways. This list of 31 pathways includes the

255   process pathway focal adhesion which was differentially regulated in both directions. While four

256   and ten process and signaling pathways respectively were downregulated, nine process and

257   signaling pathways each were upregulated. The detailed information about the number of

258   upregulated, downregulated and total genes in each of the selected pathway is submitted as Table

259   S7 and S8.

260   We examined the overlap of the process pathways with all the non-disease pathways based on

261   the shared number of DEGs. The overlapping pathways are represented as a pathway interaction

262   network in Fig. 6. Several of the signaling pathways share DEGs with the process pathways

263   indicating that the process is influenced by these signaling pathways. The details of the DEGs

264   shared by each pathway pair is presented in the Table S9.

265   Figure 7 is a graphical representation of the number of genes shared between the Osteoclast

266   differentiation pathway and other non-disease pathways. Signaling pathways, represented by

267   striped bars, constitute eight out of the 13 non-disease pathways that interact with osteoclast

268   differentiation. All the pathways sharing genes with osteoclast differentiation pathways are

269   upregulated pathways. Similar graphs were created for all upregulated process pathways and are

270   available in the supplementary figures: Fig. S1 - Fig. S7. Figure 8 shows the overlap analysis for

271   the downregulated signaling pathways.

272   Among the downregulated process pathways shown in Fig. 8, vascular smooth muscle

273   contraction interacted with platelet activation and cGMP-PKG signaling pathway, via

274   downregulated genes in the RA synovium. Regulation of lipolysis in adipocytes interacted only

275 with down signaling pathways through downregulated genes. Finally, the pathway adherens

276 junction did not overlap with any other pathways.

277 Our study revealed that the processes of natural killer cell mediated cytotoxicity as well as

278 osteoclast differentiation involved a network of several interacting pathways in the RA

279 synovium. However, osteoclast differentiation was influenced by the highest number of signaling

280 pathways. This indicates that the differentiation of osteoclasts in the RA synovium is coordinated

281 by several signaling pathways. In order to understand the collective effect of these signaling

282 pathways on the osteoclast differentiation, we created a detailed PPI network for the process in

283 the RA synovium. Among all the interactions obtained from the STRING database, we used only

284 the experimentally validated ones published in literature, for the creation of the network. In this

285 network, we indicated the differentially regulated genes from the microarray analysis to show the

286 possible ways by which the altered signaling promotes osteoclastogenesis in RA synovium.

287 **A comprehensive PPI network for the differentiation of osteoclasts in RA synovium**

288 The PPI network, created in our study, had 433 proteins and 1790 interactions. The network

289 consisted of three connected components. The two smaller connected components were the

290 interactions between CD47 and SIRPA, and IL1A and S100A13.

291 The protein Ubiquitin which has the highest number of interactions in the network was found to

292 interact with 175 network proteins. This is expected, as ubiquitination is a very common PTM

293 that marks the proteins for proteasomal degradation. In our network, ubiquitination was

294 represented as interaction of a protein with ubiquitin as well as with ubiquitin ligases. We

295 removed ubiquitin from our network since most of the edges of ubiquitin and those of the

296 ubiquitin ligases were redundant. The resulting network had 432 proteins and 1595 interactions.

297   In this network, in addition to the two small connected components, four proteins, PPP3R1,

298   PPP3CA, PPIA and RCAN1 were disconnected from the main network and formed a new

299   connected component. The network now had four components: CD47-SIRPA, IL1A-S100A13,

300   PPP3R1-PPP3CA-PPIA-RCAN1, and one large component.

301   We removed the three smaller connected components from the main network. The large

302   connected component, consisting of 424 protein nodes and 1589 interactions, was used for

303   further analysis. Henceforth, we refer to this as directed ODP network. The directed ODP

304   network contains 82 core proteins belonging to the KEGG osteoclast differentiation pathway.

305   The portion of the directed ODP network containing the 82 core proteins and their 152

306   connections is termed as the "**core network"** (directed). The rest of the network consisting of the

307   first shell interactors and their edges is the **"shell network"** (directed). The core network

308   contains proteins which are directly involved in the osteoclast differentiation. The shell network

309   represents the protein milieu in the RA synovium facilitating the osteoclast differentiation. The

310   complete directed ODP network is provided as a supplementary file S1. A second supplementary

311   file S2 contains the information about the core and shell proteins of the ODP network.

312   **The DNA-binding proteins of the directed ODP network**

313   In a PPI network, the terminal responders of the signals are the DNA-binding proteins such as a

314   transcription factors (TF), coactivator *etc*. or the proteins that generate non-protein signaling

315   molecules like secondary messengers. Using the enriched terms in the category GOMF, we

316   classified 82 of the 424 nodes as DNA-binding proteins (Fig. S8). 18 DNA- binding proteins

317   which include STAT, NF-κB and AP1 TF belong to the core network. Along with the differential

318   expression of STAT1, JUN, JUNB and FOSB, the STAT protein STAT2 was upregulated in this

319   study.

320 In addition to STAT and AP1 proteins, the other DNA binding core protein PPARG was

321 observed to be downregulated which is in agreement with the results of Li *et al* (Li et al., 2017).

322 The attachment of the shell network resulted in the inclusion of nine differentially regulated

323 DNA-binding proteins in the directed ODP network. The possible roles of these proteins in

324 osteoclastogenesis are described in context of their GOBP terms.

325 The signaling pathways of the directed ODP network

326 GOBP over-representation analysis of the directed ODP network proteins identified several

327 immune signaling terms. These terms included five out of the eight upregulated KEGG signaling

328 pathways which were found to interact with the osteoclast differentiation (Fig. 7). These

329 pathways are: T cell receptor signaling pathway, B cell receptor signaling pathway, Fc-ε receptor

330 signaling pathway, NF-κB signaling pathway and Toll-like receptor (TLR) signaling pathway.

331 Among these, T cell receptor signaling pathway had the most number of DEGs, with 14 up and

332 one downregulated nodes. The proteins belonging to the T cell receptor signaling term were

333 extracted as a subnetwork (Fig. S9). In this subnetwork, the T cell surface molecules CD3E and

334 CD28 were upregulated whereas CD247 did not show differential regulation. The downstream

335 signaling molecules ZAP70, LCK, ITK, CSK, LAT, LCP2, FYB, PAG1, PIK3CD, MAPK1,

336 PLCG2 and INPP5D were upregulated. Among them, PIK3CD, MAPK1, LCK, LCP2 and

337 PLCG2 are the core network proteins.

338 The B cell receptor signaling pathway (Fig. S10) term shared six proteins with the core network.

339 Five of these core proteins were upregulated. In addition, ZAP70, LYN and PRKCB, which are

340 part of the shell network, were found to be upregulated.

341   The Fc-ε receptor signaling pathway (Fig. S11) showed the MAP Kinase, NF-κB, Rac signaling

342   components. The term showed nine upregulated genes which included the receptor FCER1G.

343   Six GOBP terms were combined to extract 75 proteins of the NF-κB signaling pathway from the

344   directed ODP network (Fig. S12). This NF-κB subnetwork contained five receptors including

345   one core protein (TNFRSF1A). Out of the 11 DEGs in the subnetwork, TNFSF11 (RANKL) and

346   STAT1 were the core proteins. REL, a component of NF-κB transcription factor dimers was

347   upregulated. The osteoclast differentiation and activation factor, RANKL, an activator of NF-κB

348   pathway was highly upregulated with a log2 fold change of 3.32. It is known that REL

349   participates in the canonical NF-κB signaling (Shih et al., 2011). Of the many possible signaling

350   routes leading to REL, we observed DEGs in the following pathways: TCR-PRKCQ-CARD11-

351   BCL10; IL1R1-MyD88-IRAK4; TNFSF11 (RANKL)-TRAF6-IKK; CD27-TRAF2-IKK. We

352   have extracted all these routes and created a subnetwork for activation of REL in the RA

353   synovium (Fig. 9).

354   We combined nine over-represented GOBP terms to extract the TLR signaling pathway proteins

355   (Fig. S13). The extracted subnetwork featured the signaling from the receptors TLR3 and TLR4

356   to the IκB kinase complex (IKK). Although the pathway is upregulated in the RA synovium, the

357   TLR receptors in the ODP network did not show any differential regulation.

358   Interestingly, the directed ODP network demonstrated a downregulation of all the DEGs

359   participating in the TGFβ signaling pathway. The TGFβ subnetwork (Fig. S14) showed the

360   presence of the ligands TGFB1, TGFB2 and TGFB3 and the receptors TGFBR1, TGFBR2 and

361   TGFBR3. Among the downstream SMADs, SMAD3 was downregulated.

362   **Protein clusters in the directed ODP network**

363   The analysis done using the MCODE application of the Cytoscape tool revealed 19 clusters

364   which may indicate functional protein complexes. The details of these 19 clusters are submitted

365   as Table S10. Among them, 12 clusters included a mixture of core and shell proteins. In three of

366   the clusters (Cluster2, Cluster 4 and Cluster 10), the core proteins were differentially regulated.

367   Cluster 2 comprised of three core proteins and six shell proteins. The cluster had two DEGs

368   which were DNA-binding proteins STAT1 and PPARG. Cluster 4 was a TGFβ cluster and

369   cluster 10 was an NCF cluster. The TGFβ cluster had two downregulated proteins, of which

370   SMAD3 was a shell protein. The NCF cluster had four core proteins, of which NCF2, NCF4 and

371   CYBA were upregulated. The upregulated shell protein in the NCF cluster was NCF1.

372   ## Discussion

373   This study is aimed at understanding the mechanisms involved in a specific RA-related

374   phenotype. We have used a large number of microarray studies and relaxed inclusion criteria for

375   differential expression across datasets, to obtain relatively large number of DEGs that are likely

376   to be involved in RA. We have combined this gene expression data with pathway analysis and

377   identified various process pathways and several signaling pathways to be affected by RA. In

378   systemic diseases like RA, pathways responsible for a particular phenotype operate in an

379   environment consisting of various other disrupted pathways. Thus it becomes important to

380   understand the effect of this environment on the pathway immediately responsible for the

381   phenotype. We attempted to achieve this by overlapping the various process pathways with the

382   enriched signaling pathways in the synovium. Interestingly, the process pathway osteoclast

383   differentiation overlapped with several of the enriched signaling pathways. In order to

384   understand the signaling involved in osteoclast differentiation in the RA synovium, for the first

385   time, we created a detailed PPI network responsible for the phenomenon. Each interaction in this

386   network was manually verified from literature enabling the inclusion of directions of the

387   interactions and specific post translational modifications whenever such information was

388   available. While creating the network using all the possible interactions available in STRING,

389   we found that some proteins in the repository have more number of interactions reported than the

390   others. We acknowledge that this might have led to a bias in the directed ODP network. The

391   network lacks the important non-protein molecules involved in triggering the ectopic

392   differentiation of osteoclasts in the inflamed synovium. In addition, gene expression regulation

393   resulting from activation or repression of transcription factors was not depicted in the network.

394   Since the RA specific data used in this network was only gene expression data, information on

395   the activation state of specific proteins that are known to be involved in the disease, e.g.,

396   phosphorylation state of STAT1 was missing. Though the network lacks these information, it is

397   the most comprehensive and informative PPI network till date describing the process of

398   osteoclast differentiation.

399   **The differentially regulated genes in the RA synovium**

400   In order to identify the DEGs in the RA synovium, seven microarray datasets generated by five

401   different studies were used. Among the seven datasets, the RA patients belonging to the datasets

402   GSE1919, GSE12021 (U133A), GSE12021 (U133B), GS55235 and GSE55457 had similar high

403   values for the inflammatory markers, ESR and CRP. Additionally, the tissue used in GSE77298

404   were described as end stage RA synovial biopsies. Therefore, we surmise that the RA tissues

405   were highly inflamed. However, we observed that few genes were differentially expressed across

406   most of the datasets (Fig. 2). Since the level of inflammation in RA tissues were comparable, we

407   attribute this lack of concordance between the datasets to the heterogeneity of the disease.

The enriched pathways in the RA synovium

408

409   KEGG pathway enrichment analysis of the upregulated genes (common-up) resulted in 26

410   upregulated disease pathways. As expected, Rheumatoid arthritis was one of these disease

411   pathways. Staphylococcus aureus infection and Tuberculosis, the two upregulated infectious

412   disease pathways in the results of the pathway enrichment analysis are known to be associated

413   with RA (Sams et al., 2015; Jeong et al., 2017). The upregulated infectious disease pathways

414   share several genes with immune system pathways. Differential regulation of the immune

415   pathways is expected in RA since it is an immune disorder. Therefore, the upregulation of the

416   immune genes explains the enrichment of the infectious disease pathways in the up pathway list.

417   Several pathways belonging to the category "immune system" were enriched in the common-up

418   genes. The enrichment of these immune pathways is likely as infiltration of activated immune

419   cells has been observed in the RA synovium (McInnes & Schett, 2011). In addition, the resident

420   cells of the inflamed pannus exhibit activation of several immune signaling pathways (McInnes

421   & Schett, 2007). Among the pathways from Table 4, the Fc-ε RI signaling pathway shows an

422   upregulation of the IgE receptor Fc-ε RI. Elevated presence of IgE and activated mast cells have

423   been detected in the RA synovium (Gruber, Ballan & Gorevic, 1988; Tetlow & Woolley, 1995).

424   It has been demonstrated by *in vitro* experiments that the RA synovial mast cells express Fc-ε RI

425   and can be activated via the Fc-ε RI signaling pathway (Lee et al., 2013). However, the

426   contribution of the pathway to the pathological alteration of synovial tissue function needs to be

427   addressed in future studies.

428   Among the enriched downregulated pathways, seven belong to the category "signal

429   transduction" (Table 5). The AMPK signaling pathway which is known for its anti-inflammatory

430   effect shows downregulation in our analysis (Speirs et al., 2018). The downregulation of the

431   FOXO signaling pathway is also observed in this study. This reflects the results of the earlier

432   studies which have shown downregulation and inactivation of the FOXO transcription factors in

433   the RA affected synovium (Ludikhuize et al., 2007; Grabiec et al., 2015). Additionally, the

434   FOXO proteins are inhibited by NF-κB signaling, and activated by AMPK as shown in the

435   FOXO signaling pathway listed in KEGG pathway database. The pathway enrichment results

436   from our analysis show upregulation in NF-κB signaling and downregulation in AMPK

437   signaling, which explains the downregulation of FOXO signaling. Further, the cGMP-PKG

438   signaling is known to be essential for the vascular smooth muscle response to the inflammatory

439   cytokines (Browner, Sellak & Lincoln, 2004). The downregulation of genes in this pathway

440   needs to be studied in detail to analyse the effects of RA synovial inflammation on the blood

441   vessels in the affected tissue. The cAMP signaling pathway is known to facilitate regulatory T

442   cell function and T cell anergy (Raker, Becker & Steinbrink, 2016). The downregulation in this

443   pathway might indicate the pro-inflammatory nature of the infiltrating T cells in the synovium. It

444   is difficult to explain the importance of downregulation in the Wnt signaling pathway and the

445   MAPK signaling pathway in the RA synovium. Wnt signaling, which is required for repair of

446   bone erosions, is suppressed in mouse models of inflammatory arthritis (Lories, Corr & Lane,

447   2013). However, Wnt signaling has also been linked to proinflammatory cytokine production in

448   the affected synovium (Miao et al., 2013). In our study, the KEGG Wnt signaling pathway shows

449   downregulation of the Wnt receptors FZD4 and LRP6, suggesting downregulation of the

450   pathway, while the downregulation of Wnt antagonists SFRP1, SFRP2 and SFRP4 suggests

451   activation of the pathway. Thus no conclusions can be made about the role of Wnt signaling

452   pathway from its presence in the list of downregulated pathways. Similarly, no clear picture can

453   be drawn about the role of MAPK signaling pathway in the RA synovium. This is because the

454 KEGG MAPK signaling pathway is very large, with 252 genes and several genes are grouped

455 together under terms like RTK (receptor tyrosine kinase) and GF (growth factor), making the

456 pathway extremely general.

Construction of the directed ODP network

458 Three small connected components were removed while creating the directed ODP network. One

459 of the components contained the proteins PPP3R1 and PPP3CA which are the subunits of

460 calcineurin, an important regulator of osteoclast differentiation. The immunosuppressant drug

461 cyclosporin used in the treatment of RA acts as an inhibitor of calcineurin by forming a ternary

462 complex with PPIA and calcineurin (Wang & Heitman, 2005). The remaining protein in the

463 connected component, RCAN1 is an inhibitor of calcineurin. Both in the KEGG pathway

464 database and in literature, these proteins are described as participating in the RANKL signaling

465 pathway which regulates the differentiation of osteoclasts. Calcineurin participates in the

466 RANKL signaling pathway downstream of the non-protein components inositol triphosphate and

467 calcium ions. Since our network is based on protein-protein interactions, it failed to capture the

468 connection of calcineurin and its adjacent proteins to the large connected component, in spite of

469 showing the connections of other RANKL pathway proteins.

Analysis of the directed ODP network

471 The directed ODP network was analysed to examine three main aspects. Firstly, the proteins

472 binding to DNA were explored because they are involved in regulating gene expression.

473 Secondly, the proteins involved in signal transduction were studied for their role in facilitating

474 the process of osteoclast differentiation in the RA affected synovium. Finally, the clusters of

475 highly interconnected proteins were identified in the network. Several proteins function as part of

476 protein complexes. The protein clusters we identified may represent the protein complexes. We

477 also examined the differential expression of the proteins in the clusters. Protein clusters with

478 differentially regulated proteins may represent complexes actively involved in the

479 osteoclastogenesis in RA synovium.

480 **DNA binding proteins in the directed ODP network**

481 Among the DNA-binding proteins in the directed ODP network, the downregulation of the AP1

482 proteins JUN and JUNB is in contrast to the earlier studies which reported their upregulation in

483 RA synovium (Kinne et al., 1995). As shown in Table 3, the AP1 downregulation was observed

484 in five of the seven datasets, with the three proteins FOSB, JUN and JUNB being downregulated

485 in three out of the seven datasets. The presence of consistent downregulation in the datasets from

486 different studies shows that the downregulation of these proteins is not due to dataset specific

487 factors. As mentioned earlier, four of the five datasets showing downregulation had patients with

488 similar clinical characteristics. This may suggest a connection between the stage of the disease

489 and the downregulation of the AP1 proteins. Further studies are required to explain the

490 downregulation of these AP1 proteins in the synovium of a subset of RA patients.

491 According to previous reports, another AP1 protein FOS is upregulated in the RA synovium

492 (Dooley et al., 1996). However, our analysis does not show any differential regulation of FOS.

493 FOS is an indispensable transcription factor for osteoclast differentiation (Grigoriadis et al.,

494 1994). In the ODP network, the upregulation of MAPK1 and RPS6KA1 presents a possible

495 mechanism for FOS activation. In addition, our network reveals the activation of FOS by the

496 upregulated TF STAT1. Along with STAT1, the TF STAT2, the kinase JAK2 and the receptor

497 IFNAR2 were upregulated in the IFN pathway of the ODP network. TNFSF11 (RANKL)

498 induces expression of IFNβ which serves as a feedback inhibitor of osteoclastogenesis via

499   STAT1 (Xiong et al., 2016). The upregulation of the IFNβ receptor, IFNAR2 and the TF STAT1

500   may indicate that the feedback inhibition is functional in the RA synovium.

Signaling pathways in the directed ODP network

502   The examination of the enriched GOBP terms of the network showed an involvement of T cell

503   receptor signaling pathway, B cell receptor signaling pathway, Fc-ε receptor signaling pathway

504   and NF-κB signaling pathway in the differentiation of osteoclasts in the RA synovium. Figure 9

505   depicts the signaling routes that lead to activation of REL in the RA synovium. The figure shows

506   activation of REL by PKC. The network shows T cell receptor mediated activation of PRKCQ

507   and its subsequent activation of CARD11. Based on earlier studies (Szamel, Bartels & Resch,

508   1993; Sommer et al., 2005), we speculate that another isoform of PKC, PRKCB is also involved

509   in the activation of CARD11 in the affected synovium. It is known that the activated CARD11

510   via the formation of a trimeric complex with BCL10 and MALT1 activates the IKK complex

511   (Turvey et al., 2014). The IKK complex in turn activates NF-κB proteins including REL by

512   removing their inhibition by NFKBIA and NFKBIB. Figure 9 shows that the ODP network has

513   captured the interactions of these proteins. In this pathway, the upregulation of PRKCB and REL

514   and the downregulation of NFKBIA indicate activation of REL via this pathway in the RA

515   affected synovium. It is also known that NF-κB proteins are activators of RANKL (TNFSF11)

516   gene expression in activated T cells (Fionda et al., 2007). Our network, through the presence of

517   DEGs in the RA synovium, shows how TCR signaling aids in activation of REL which leads to

518   RANKL expression and osteoclastogenesis. In addition, the upregulation of TNFSF11 and

519   downregulation of its competitive inhibitor TNFRSF11B marks another route to the activation of

520   REL via TRAF6. PPARG, a known inhibitor of TNFSF11-mediated osteoclastogenesis, was

521   downregulated in the ODP network. This result agrees with the findings of Li *et al* (Li et al.,

522   2017). The upregulation of CD27 receptor is in accordance with the reports of high levels of

523   CD27 in the synovial tissue of RA patients (Tak et al., 1996). As CD27 also activates TRAF6,

524   we speculate that the upregulation of CD27 contributes to the activation of REL via this route in

525   the RA synovium. Among the DEGs involved in the REL activation via IL1 pathway, CASP1

526   was upregulated whereas the inhibitor PELI1 was downregulated. This is balanced against the

527   upregulation of IL1RN, a competitive inhibitor of IL1R1. It is known that both the cytokines,

528   IL1 and TNF, activate NF-κB through TRAF6. However, we did not observe differential

529   regulation of the TNF pathway proteins leading to the activation of NF-κB. On the other hand,

530   the apoptosis related proteins such as FAS and CASP8 which are downstream to TNF receptor,

531   were upregulated. Earlier studies have established that the death signaling pathways are

532   antagonized by the activity of BIRC2, BIRC3 and XIAP (Vasudevan & Ryoo, 2015). The

533   upregulation of BIRC3 points to suppression of TNF mediated apoptosis and the activation of

534   NF-κB via TNF receptor signaling in the RA synovium. This reflects the possibility that TNF

535   signaling results in both CASP8 mediated apoptosis and BIRC3 mediated NF-κB activation in

536   different parts of the RA synovium. The network also captures the activation of IKK by the

537   upregulated NOD2 through MAP3K7. Studies have reported that RA synovial cells express high

538   levels of NOD2 (Franca et al., 2016). We hypothesize that NOD-dependent activation of NF-κB

539   also contributes to the osteoclastogenesis in RA synovium. All the routes that lead to activation

540   of NF-κB points to the canonical signaling. PEL1 is known to be a negative regulator of REL

541   (Chang et al., 2011). REL, which functions only in the canonical NF-κB pathway, is the only

542   NF-κB protein showing differential regulation in this analysis (Shih et al., 2011). The

543   upregulation of REL, BIRC3 and PRKCB and the downregulation of PELI1 also support the

544   predominance of canonical NF-κB signaling in RA synovium osteoclastogenesis (Lutzny et al.,

545   2013; Varfolomeev et al., 2007).

546   The GOBP analysis revealed enrichment of TGFβ signaling pathway terms. The role of TGFβ

547   signaling pathway in osteoclastogenesis is not captured in the KEGG osteoclast differentiation

548   pathway as it shows only the receptor-ligand interaction. Our network connects the TGFβ

549   receptors to downstream DNA-binding proteins, of which Forkhead box proteins, SMAD3 and

550   ATF3 were downregulated. The consistent downregulation of all the DEGs in the TGFβ pathway

551   implies their negative regulatory role in the osteoclastogenesis. Although the published studies

552   support a mixed role of TGFβ pathway in the RA synovium, Karst *et al* showed that high

553   concentration of TGFβ is involved in the inhibition of osteoclastogenesis (Karst et al., 2004).

554   Therefore, we hypothesize that the downregulation of the TGFβ pathway produces a favourable

555   environment for osteoclastogenesis in the RA synovium.

556   **A proposed model of enhanced ROS production mediating osteoclastogenesis in RA**

557   **synovium**

558   It is known that the Nox2 complex generates ROS which act as secondary messengers during

559   osteoclast differentiation (Kang & Kim, 2016). ROS are also known to cause activation of

560   canonical NF-κB pathway (Gloire, Legrand-Poels & Piette, 2006). In the ODP network, the

561   upregulation of NCF1, NCF2, NCF4 and CYBA, the four components of Nox2 complex, may

562   indicate the osteoclast differentiation as well as oxidative burst by phagocytic cells in the RA

563   synovium (Rosen et al., 1995). The NCF cluster, selected by MCODE analysis, demonstrates the

564   activation of the core proteins NCF2, NCF4, CYBA and RAC1 by the shell proteins NCF1,

565   PRKCZ and PARD6G. It is known that PRKCZ is activated by T cell receptor (Bertrand et al.,

566   2010). In our analysis, it was observed that multiple T cell receptor signaling molecules were

567   upregulated. We hypothesize that the TCR signaling via PRKCZ activates NCF complex which

568   may subsequently generate ROS in the RA synovium. It is known that ROS can diffuse across

569   cell membranes to take part in intracellular signaling (Fisher, 2009). We believe that the

570   activation of PRKCZ leading to the generation of ROS is one of the routes facilitating osteoclast

571   differentiation in RA synovium. Our network also illustrated an upregulation of the catalytic

572   subunit of PI3K, PIK3CD which is another downstream molecule of T cell receptor signaling.

573   This is supported by the findings of Bartok et al (Bartok et al., 2012). PI3K-AKT pathway results

574   in inhibition of the Forkhead box proteins, FOXO1 and FOXO3 (Patel & Mohan, 2005). It was

575   shown that FOXO upregulates antioxidant enzymes that inhibit osteoclastogenesis (Bartell et al.,

576   2014). In the ODP network, the downregulation of FOXO1 and FOXO3 is a possible indication

577   of the presence of ROS mediated osteoclastogenesis in RA synovium. Bartell *et al* (Bartell et al.,

578   2014), experimentally proved the role of RANKL in the downregulation of the Forkhead box

579   proteins. In accordance with this, our analysis showed an upregulation of RANKL in the ODP

580   network. Finally, our analysis reported the upregulation of the cytokine receptor CSF1R which is

581   also required for osteoclastogenesis via PI3K-AKT pathway. The proposed model consisting of

582   all the signaling routes promoting osteoclastogenesis via generation of ROS in the RA synovium

583   is summarized in Fig. 10.

584 **Conclusion**

585   In this study, we have created a PPI network for osteoclast differentiation in the RA synovium

586   for the first time using gene expression under RA conditions from microarray experiments,

587   pathway enrichment analysis and protein protein interaction data. This network captures all the

588   signaling routes that lead to osteoclastogenesis in the synovium and depicts the roles of T cell

589   receptor signaling, canonical NF-κB pathway and ROS generation.

590   **Data Availability:** All the data generated and used for the analysis are included in this

591   manuscript and in the Supplementary information. Abbreviations used are included in Table S11.

592   References

593   Assenov Y., Ramirez F., Schelhorn S-E., Lengauer T., Albrecht M. 2008. Computing topological

594   parameters of biological networks. *Bioinformatics (Oxford, England)* 24:282–284. DOI:

595   10.1093/bioinformatics/btm554.

596   Bader GD., Hogue CWV. 2003. An automated method for finding molecular complexes in large protein

597   interaction networks. *BMC bioinformatics* 4:2.

598   Bartell SM., Kim H-N., Ambrogini E., Han L., Iyer S., Serra Ucer S., Rabinovitch P., Jilka RL., Weinstein RS.,

599   Zhao H., O'Brien CA., Manolagas SC., Almeida M. 2014. FoxO proteins restrain

600   osteoclastogenesis and bone resorption by attenuating H2O2 accumulation. *Nature*

601   *communications* 5:3773. DOI: 10.1038/ncomms4773.

602   Bartok B., Boyle DL., Liu Y., Ren P., Ball ST., Bugbee WD., Rommel C., Firestein GS. 2012. PI3 kinase delta

603   is a key regulator of synoviocyte function in rheumatoid arthritis. *The American journal of*

604   *pathology* 180:1906–1916. DOI: 10.1016/j.ajpath.2012.01.030.

605   Bertrand F., Esquerre M., Petit A-E., Rodrigues M., Duchez S., Delon J., Valitutti S. 2010. Activation of the

606   ancestral polarity regulator protein kinase C zeta at the immunological synapse drives

607   polarization of Th cell secretory machinery toward APCs. *Journal of immunology (Baltimore,*

608   *Md. : 1950)* 185:2887–2894. DOI: 10.4049/jimmunol.1000739.

609   Broeren MGA., de Vries M., Bennink MB., Arntz OJ., Blom AB., Koenders MI., van Lent PLEM., van der

610   Kraan PM., van den Berg WB., van de Loo FAJ. 2016. Disease-Regulated Gene Therapy with Anti-

611   Inflammatory Interleukin-10 Under the Control of the CXCL10 Promoter for the Treatment of

612   Rheumatoid Arthritis. *Human gene therapy* 27:244–254. DOI: 10.1089/hum.2015.127.

613   Browner NC., Sellak H., Lincoln TM. 2004. Downregulation of cGMP-dependent protein kinase

614   expression by inflammatory cytokines in vascular smooth muscle cells. *American journal of*

615   *physiology. Cell physiology* 287:C88-96. DOI: 10.1152/ajpcell.00039.2004.

616   Chang M., Jin W., Chang J-H., Xiao Y., Brittain G., Yu J., Zhou X., Wang Y-H., Cheng X., Li P., Rabinovich

617   BA., Hwu P., Sun S-C. 2011. Peli1 negatively regulates T-cell activation and prevents

618   autoimmunity. *Nature immunology* 12:1002–1009. DOI: 10.1038/ni.2090.

619   Davis S., Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and

620   BioConductor. *Bioinformatics (Oxford, England)* 23:1846–1847. DOI:

621   10.1093/bioinformatics/btm254.

622   Dey P., Panga V., Raghunathan S. 2016. A Cytokine Signalling Network for the Regulation of Inducible

623   Nitric Oxide Synthase Expression in Rheumatoid Arthritis. *PloS one* 11:e0161306. DOI:

624   10.1371/journal.pone.0161306.

625   Dooley S., Herlitzka I., Hanselmann R., Ermis A., Henn W., Remberger K., Hopf T., Welter C. 1996.

626   Constitutive expression of c-fos and c-jun, overexpression of ets-2, and reduced expression of

627   metastasis suppressor gene nm23-H1 in rheumatoid arthritis. *Annals of the rheumatic diseases*

628   55:298–304.

629   Edgar R., Domrachev M., Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and

630   hybridization array data repository. *Nucleic acids research* 30:207–210.

631   Expansion of the Gene Ontology knowledgebase and resources. 2017. *Nucleic acids research* 45:D331–

632   D338. DOI: 10.1093/nar/gkw1108.

633   Fionda C., Nappi F., Piccoli M., Frati L., Santoni A., Cippitelli M. 2007. 15-Deoxy-$\Delta^{12,14}$-Prostaglandin J_2

634   Negatively Regulates rankl Gene Expression in Activated T Lymphocytes: Role of NF-

635   κB and Early Growth Response Transcription Factors. *The Journal of Immunology* 178:4039. DOI:

636   10.4049/jimmunol.178.7.4039.

637 Fisher AB. 2009. Redox Signaling Across Cell Membranes. *Antioxidants & Redox Signaling* 11:1349–1356.

638   DOI: 10.1089/ars.2008.2378.

639 Franca R., Vieira SM., Talbot J., Peres RS., Pinto LG., Zamboni DS., Louzada-Junior P., Cunha FQ., Cunha

640   TM. 2016. Expression and activity of NOD1 and NOD2/RIPK2 signalling in mononuclear cells

641   from patients with rheumatoid arthritis. *Scandinavian journal of rheumatology* 45:8–12. DOI:

642   10.3109/03009742.2015.1047403.

643 Gautier L., Cope L., Bolstad BM., Irizarry RA. 2004. affy--analysis of Affymetrix GeneChip data at the

644   probe level. *Bioinformatics (Oxford, England)* 20:307–315. DOI: 10.1093/bioinformatics/btg405.

645 Gloire G., Legrand-Poels S., Piette J. 2006. NF-kappaB activation by reactive oxygen species: fifteen years

646   later. *Biochemical pharmacology* 72:1493–1505. DOI: 10.1016/j.bcp.2006.04.011.

647 Grabiec AM., Angiolilli C., Hartkamp LM., van Baarsen LGM., Tak PP., Reedquist KA. 2015. JNK-

648   dependent downregulation of FoxO1 is required to promote the survival of fibroblast-like

649   synoviocytes in rheumatoid arthritis. *Annals of the rheumatic diseases* 74:1763–1771. DOI:

650   10.1136/annrheumdis-2013-203610.

651 Gravallese EM., Harada Y., Wang JT., Gorn AH., Thornhill TS., Goldring SR. 1998. Identification of cell

652   types responsible for bone resorption in rheumatoid arthritis and juvenile rheumatoid arthritis.

653   *The American journal of pathology* 152:943–951.

654 Grigoriadis AE., Wang ZQ., Cecchini MG., Hofstetter W., Felix R., Fleisch HA., Wagner EF. 1994. c-Fos: a

655   key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science*

656   *(New York, N.Y.)* 266:443–448.

657   Gruber B., Ballan D., Gorevic PD. 1988. IgE rheumatoid factors: quantification in synovial fluid and ability

658   to induce synovial mast cell histamine release. *Clinical and Experimental Immunology* 71:289–

659   294.

660   Hao R., Du H., Guo L., Tian F., An N., Yang T., Wang C., Wang B., Zhou Z. 2017. Identification of

661   dysregulated genes in rheumatoid arthritis based on bioinformatics analysis. *PeerJ* 5:e3078. DOI:

662   10.7717/peerj.3078.

663   Hosack DA., Dennis G., Sherman BT., Lane HC., Lempicki RA. 2003. Identifying biological themes within

664   lists of genes with EASE. *Genome Biology* 4:P4. DOI: 10.1186/gb-2003-4-6-p4.

665   Huang DW., Sherman BT., Lempicki RA. 2009a. Systematic and integrative analysis of large gene lists

666   using DAVID bioinformatics resources. *Nature protocols* 4:44–57. DOI: 10.1038/nprot.2008.211.

667   Huang DW., Sherman BT., Lempicki RA. 2009b. Bioinformatics enrichment tools: paths toward the

668   comprehensive functional analysis of large gene lists. *Nucleic acids research* 37:1–13. DOI:

669   10.1093/nar/gkn923.

670   Huber R., Hummert C., Gausmann U., Pohlers D., Koczan D., Guthke R., Kinne RW. 2008. Identification of

671   intra-group, inter-individual, and gene-specific variances in mRNA expression profiles in the

672   rheumatoid arthritis synovial membrane. *Arthritis research & therapy* 10:R98. DOI:

673   10.1186/ar2485.

674   Jeong H., Baek SY., Kim SW., Eun YH., Kim IY., Kim H., Lee J., Koh E-M., Cha H-S. 2017. Comorbidities of

675   rheumatoid arthritis: Results from the Korean National Health and Nutrition Examination

676   Survey. *PloS one* 12:e0176260. DOI: 10.1371/journal.pone.0176260.

677   Jung SM., Kim KW., Yang C-W., Park S-H., Ju JH. 2014. Cytokine-Mediated Bone Destruction in

678   Rheumatoid Arthritis. *Journal of Immunology Research* 2014. DOI: 10.1155/2014/263625.

679   Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K. 2017. KEGG: new perspectives on

680   genomes, pathways, diseases and drugs. *Nucleic acids research* 45:D353–D361. DOI:

681   10.1093/nar/gkw1092.

682   Kang IS., Kim C. 2016. NADPH oxidase gp91(phox) contributes to RANKL-induced osteoclast

683   differentiation by upregulating NFATc1. *Scientific reports* 6:38014. DOI: 10.1038/srep38014.

684   Karst M., Gorny G., Galvin RJS., Oursler MJ. 2004. Roles of stromal cell RANKL, OPG, and M-CSF

685   expression in biphasic TGF-beta regulation of osteoclast differentiation. *Journal of cellular*

686   *physiology* 200:99–106. DOI: 10.1002/jcp.20036.

687   Kinne RW., Boehm S., Iftner T., Aigner T., Vornehm S., Weseloh G., Bravo R., Emmrich F., Kroczek RA.

688   1995. Synovial fibroblast-like cells strongly express jun-B and C-fos proto-oncogenes in

689   rheumatoid- and osteoarthritis. *Scandinavian journal of rheumatology. Supplement* 101:121–

690   125.

691   Kupfer P., Huber R., Weber M., Vlaic S., Haupl T., Koczan D., Guthke R., Kinne RW. 2014. Novel

692   application of multi-stimuli network inference to synovial fibroblasts of rheumatoid arthritis

693   patients. *BMC medical genomics* 7:40. DOI: 10.1186/1755-8794-7-40.

694   Lee H., Kashiwakura J., Matsuda A., Watanabe Y., Sakamoto-Sasaki T., Matsumoto K., Hashimoto N.,

695   Saito S., Ohmori K., Nagaoka M., Tokuhashi Y., Ra C., Okayama Y. 2013. Activation of human

696   synovial mast cells from rheumatoid arthritis or osteoarthritis patients in response to

697   aggregated IgG through Fcgamma receptor I and Fcgamma receptor II. *Arthritis and rheumatism*

698   65:109–119. DOI: 10.1002/art.37741.

699   Lee H-M., Sugino H., Aoki C., Shimaoka Y., Suzuki R., Ochi K., Ochi T., Nishimoto N. 2011. Abnormal

700   networks of immune response-related molecules in bone marrow cells from patients with

701   rheumatoid arthritis as revealed by DNA microarray analysis. *Arthritis research & therapy*

702   13:R89. DOI: 10.1186/ar3364.

703   Li X-F., Sun Y-Y., Bao J., Chen X., Li Y-H., Yang Y., Zhang L., Huang C., Wu B-M., Meng X-M., Li J. 2017.

704   Functional role of PPAR-gamma on the proliferation and migration of fibroblast-like

705   synoviocytes in rheumatoid arthritis. *Scientific reports* 7:12671. DOI: 10.1038/s41598-017-

706   12570-6.

707   Lories RJ., Corr M., Lane NE. 2013. To Wnt or not to Wnt: the bone and joint health dilemma. *Nature*

708   *reviews. Rheumatology* 9:328–339. DOI: 10.1038/nrrheum.2013.25.

709   Ludikhuize J., de Launay D., Groot D., Smeets TJM., Vinkenoog M., Sanders ME., Tas SW., Tak PP.,

710   Reedquist KA. 2007. Inhibition of forkhead box class O family member transcription factors in

711   rheumatoid synovial tissue. *Arthritis and rheumatism* 56:2180–2191. DOI: 10.1002/art.22653.

712   Lutzny G., Kocher T., Schmidt-Supprian M., Rudelius M., Klein-Hitpass L., Finch AJ., Dürig J., Wagner M.,

713   Haferlach C., Kohlmann A., Schnittger S., Seifert M., Wanninger S., Zaborsky N., Oostendorp R.,

714   Ruland J., Leitges M., Kuhnt T., Schäfer Y., Lampl B., Peschel C., Egle A., Ringshausen I. 2013.

715   Protein Kinase C-β-Dependent Activation of NF-κB in Stromal Cells Is Indispensable for the

716   Survival of Chronic Lymphocytic Leukemia B Cells In Vivo. *Cancer Cell* 23:77–92. DOI:

717   10.1016/j.ccr.2012.12.003.

718   McInnes IB., Schett G. 2007. Cytokines in the pathogenesis of rheumatoid arthritis. *Nature reviews.*

719   *Immunology* 7:429–442. DOI: 10.1038/nri2094.

720   McInnes IB., Schett G. 2011. The pathogenesis of rheumatoid arthritis. *The New England journal of*

721   *medicine* 365:2205–2219. DOI: 10.1056/NEJMra1004965.

722   Miao C., Yang Y., He X., Li X., Huang C., Huang Y., Zhang L., Lv X-W., Jin Y., Li J. 2013. Wnt signaling

723   pathway in rheumatoid arthritis, with special emphasis on the different roles in synovial

724   inflammation and bone remodeling. *Cellular signalling* 25:2069–2078. DOI:

725   10.1016/j.cellsig.2013.04.002.

726   Nevius E., Gomes AC., Pereira JP. 2016. A comprehensive review of inflammatory cell migration in

727   rheumatoid arthritis. *Clinical reviews in allergy & immunology* 51:59–78. DOI: 10.1007/s12016-

728   015-8520-9.

729   Otterness IG. 1994. The value of C-reactive protein measurement in rheumatoid arthritis. *Seminars in*

730   *arthritis and rheumatism* 24:91–104.

731   Patel RK., Mohan C. 2005. PI3K/AKT signaling and systemic autoimmunity. *Immunologic research* 31:47–

732   55. DOI: 10.1385/IR:31:1:47.

733   Pepper SD., Saunders EK., Edwards LE., Wilson CL., Miller CJ. 2007. The utility of MAS5 expression

734   summary and detection call algorithms. *BMC Bioinformatics* 8:273. DOI: 10.1186/1471-2105-8-

735   273.

736   R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical

737   Computing, Vienna, Austria. URL https://www.R-project.org/.

738   Raker VK., Becker C., Steinbrink K. 2016. The cAMP Pathway as Therapeutic Target in Autoimmune and

739   Inflammatory Diseases. *Frontiers in Immunology* 7. DOI: 10.3389/fimmu.2016.00123.

740   Rosen GM., Pou S., Ramos CL., Cohen MS., Britigan BE. 1995. Free radicals and phagocytic cells. *FASEB*

741   *journal : official publication of the Federation of American Societies for Experimental Biology*

742   9:200–209.

743   Sams M., Olsen MA., Joshi R., Ranganathan P. 2015. Staphylococcus aureus sepsis in rheumatoid

744   arthritis. *Rheumatology international* 35:1503–1510. DOI: 10.1007/s00296-015-3239-8.

745   Schett G. 2007. Cells of the synovium in rheumatoid arthritis. Osteoclasts. *Arthritis research & therapy*

746   9:203. DOI: 10.1186/ar2110.

747   Shannon P., Markiel A., Ozier O., Baliga NS., Wang JT., Ramage D., Amin N., Schwikowski B., Ideker T.

748   2003. Cytoscape: a software environment for integrated models of biomolecular interaction

749   networks. *Genome research* 13:2498–2504. DOI: 10.1101/gr.1239303.

750   Shih VF-S., Tsui R., Caldwell A., Hoffmann A. 2011. A single NFκB system for both canonical and non-

751   canonical signaling. *Cell Research* 21:86–102. DOI: 10.1038/cr.2010.161.

752   Sommer K., Guo B., Pomerantz JL., Bandaranayake AD., Moreno-Garcia ME., Ovechkina YL., Rawlings DJ.

753   2005. Phosphorylation of the CARMA1 linker controls NF-kappaB activation. *Immunity* 23:561–

754   574. DOI: 10.1016/j.immuni.2005.09.014.

755   Speirs C., Williams JJL., Riches K., Salt IP., Palmer TM. 2018. Linking energy sensing to suppression of JAK-

756   STAT signalling: A potential route for repurposing AMPK activators? *Pharmacological research*

757   128:88–100. DOI: 10.1016/j.phrs.2017.10.001.

758   Szamel M., Bartels F., Resch K. 1993. Cyclosporin A inhibits T cell receptor-induced interleukin-2

759   synthesis of human T lymphocytes by selectively preventing a transmembrane signal

760   transduction pathway leading to sustained activation of a protein kinase C isoenzyme, protein

761   kinase. *European journal of immunology* 23:3072–3081. DOI: 10.1002/eji.1830231205.

762   Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A.,

763   Santos A., Tsafou KP., Kuhn M., Bork P., Jensen LJ., von Mering C. 2015. STRING v10: protein-

764   protein interaction networks, integrated over the tree of life. *Nucleic acids research* 43:D447-

765   452. DOI: 10.1093/nar/gku1003.

766   Tak PP., Hintzen RQ., Teunissen JJ., Smeets TJ., Daha MR., van Lier RA., Kluin PM., Meinders AE., Swaak

767   AJ., Breedveld FC. 1996. Expression of the activation antigen CD27 in rheumatoid arthritis.

768   *Clinical immunology and immunopathology* 80:129–138.

769   Tenenbaum D. 2017. KEGGREST: Client-side REST access to KEGG. R   package version 1.14.1.

770   Tetlow LC., Woolley DE. 1995. Mast cells, cytokines, and metalloproteinases at the rheumatoid lesion:

771   dual immunolocalisation studies. *Annals of the rheumatic diseases* 54:896–903.

772   Turvey SE., Durandy A., Fischer A., Fung S-Y., Geha RS., Gewies A., Giese T., Greil J., Keller B., McKinnon

773   ML., Neven B., Rozmus J., Ruland J., Snow AL., Stepensky P., Warnatz K. 2014. The CARD11-

774   BCL10-MALT1 (CBM) signalosome complex: Stepping into the limelight of human primary

775   immunodeficiency. *The Journal of allergy and clinical immunology* 134:276–284. DOI:

776   10.1016/j.jaci.2014.06.015.

777 Ungethuem U., Haeupl T., Witt H., Koczan D., Krenn V., Huber H., von Helversen TM., Drungowski M.,

778   Seyfert C., Zacher J., Pruss A., Neidel J., Lehrach H., Thiesen HJ., Ruiz P., Blass S. 2010. Molecular

779   signatures and new candidates to target the pathogenesis of rheumatoid arthritis. *Physiological*

780   *genomics* 42A:267–282. DOI: 10.1152/physiolgenomics.00004.2010.

781 Varfolomeev E., Blankenship JW., Wayson SM., Fedorova AV., Kayagaki N., Garg P., Zobel K., Dynek JN.,

782   Elliott LO., Wallweber HJA., Flygare JA., Fairbrother WJ., Deshayes K., Dixit VM., Vucic D. 2007.

783   IAP antagonists induce autoubiquitination of c-IAPs, NF-kappaB activation, and TNFalpha-

784   dependent apoptosis. *Cell* 131:669–681. DOI: 10.1016/j.cell.2007.10.030.

785 Vasudevan D., Ryoo HD. 2015. Regulation of Cell Death by IAPs and Their Antagonists. *Current topics in*

786   *developmental biology* 114:185–208. DOI: 10.1016/bs.ctdb.2015.07.026.

787 Wang H., Guo J., Jiang J., Wu W., Chang X., Zhou H., Li Z., Zhao J. 2017. New genes associated with

788   rheumatoid arthritis identified by gene expression profiling. *International journal of*

789   *immunogenetics* 44:107–113. DOI: 10.1111/iji.12313.

790 Wang P., Heitman J. 2005. The cyclophilins. *Genome Biology* 6:226. DOI: 10.1186/gb-2005-6-7-226.

791 Wetteland P., Roger M., Solberg HE., Iversen OH. 1996. Population-based erythrocyte sedimentation

792   rates in 3910 subjectively healthy Norwegian adults. A statistical study based on men and

793   women from the Oslo area. *Journal of internal medicine* 240:125–131.

794 Wilson CL., Miller CJ. 2005. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data

795   analysis. *Bioinformatics (Oxford, England)* 21:3683–3685. DOI: 10.1093/bioinformatics/bti605.

796 Wishart DS., Feunang YD., Guo AC., Lo EJ., Marcu A., Grant JR., Sajed T., Johnson D., Li C., Sayeeda Z.,

797   Assempour N., Iynkkaran I., Liu Y., Maciejewski A., Gale N., Wilson A., Chin L., Cummings R., Le

798   D., Pon A., Knox C., Wilson M. 2018. DrugBank 5.0: a major update to the DrugBank database for

799   2018. *Nucleic acids research* 46:D1074–D1082. DOI: 10.1093/nar/gkx1037.

800   Woetzel D., Huber R., Kupfer P., Pohlers D., Pfaff M., Driesch D., Haupl T., Koczan D., Stiehl P., Guthke R.,

801   Kinne RW. 2014. Identification of rheumatoid arthritis and osteoarthritis patients by

802   transcriptome-based rule set generation. *Arthritis research & therapy* 16:R84. DOI:

803   10.1186/ar4526.

804   Wu G., Zhu L., Dent JE., Nardini C. 2010. A comprehensive molecular interaction map for rheumatoid

805   arthritis. *PloS one* 5:e10137. DOI: 10.1371/journal.pone.0010137.

806   Xiong Q., Zhang L., Ge W., Tang P. 2016. The roles of interferons in osteoclasts and osteoclastogenesis.

807   *Joint, bone, spine : revue du rhumatisme* 83:276–281. DOI: 10.1016/j.jbspin.2015.07.010.

808   You S., Yoo S-A., Choi S., Kim J-Y., Park S-J., Ji JD., Kim T-H., Kim K-J., Cho C-S., Hwang D., Kim W-U. 2014.

809   Identification of key regulators for the migration and invasion of rheumatoid synoviocytes

810   through a systems approach. *Proceedings of the National Academy of Sciences of the United*

811   *States of America* 111:550–555. DOI: 10.1073/pnas.1311239111.

812   Zhu Y., Davis S., Stephens R., Meltzer PS., Chen Y. 2008. GEOmetadb: powerful alternative search engine

813   for the Gene Expression Omnibus. *Bioinformatics (Oxford, England)* 24:2798–2800. DOI:

814   10.1093/bioinformatics/btn520.

815

The details of the databases used in this study.

The name and the reference of the database is listed in the column named "Database". The type of data used from the database and the features of the database are described in the table. The rationale for using the data from each of the databases in our study is included in the table.

Database	Type of data obtained for this study	Features	Rationale
GEO (Edgar, Domrachev & Lash, 2002)	microarray gene expression data	GEO is a public repository with easy access to high throughput data, including microarray data and related metadata such as tissue type, disease state etc.	Microarray data from GEO database was used to identify DEGs in the RA synovium
DAVID - Gene ID conversion tool (Huang, Sherman & Lempicki, 2009b) (Huang, Sherman & Lempicki, 2009a)	Gene ID types	The DAVID knowledge base supports conversion between more than 20 gene ID types, including Affymetrix probe IDs.	The DAVID gene ID conversion table was used to convert Affymetrix probe IDs to Entrez IDs.
KEGG pathway (Kanehisa et al., 2017)	molecular pathways	KEGG pathways are manually drawn and frequently updated. References are provided for each pathway.	The KEGG pathway database was used to identify the enriched pathways from the list of DEGs.
String (v10) (Szklarczyk et al., 2015)	protein-protein interaction	String integrates information from several sources including experimental results from literature, and provides a confidence score for the interaction.	The String database was used to identify the proteins that directly interacted with KEGG osteoclast differentiation pathway proteins. These proteins (KEGG osteoclast differentiation pathway proteins and their interactors) were used to create a PPI network for osteoclast differentiation.
Gene Ontology ("Expansion of the Gene Ontology knowledgebase and	functional annotation	Gene Ontology annotation associates genes to specific functional terms. A	GO enrichment was used to identify the functions associated with the proteins in the osteoclast

resources.," 2017)		GO enrichment analysis provides information about functions that a set of genes may be involved in.	differentiation PPI network.
DrugBank (Wishart et al., 2018)	drug targets	DrugBank provides comprehensive information of drug targets and drug types (small molecule, biologics *etc*.)	DrugBank was used to identify proteins in the network that are targets of RA drugs.

The details of the microarray datasets from GEO repository used in our study

The accession number and the reference of the datasets is listed in the column named "Accession Number". The title of the Affymetrix platform used by the particular dataset in included in the "Platform" column of the table. The information on the number of the RA patients and healthy controls used in each dataset are included in this table. Gender distribution is provided for the datasets GSE12021(U133A) and 12021 (U133B).

Accession Number	Platform	Rheumatoid arthritis samples	Healthy control samples
GSE1919(Ungethuem et al., 2010)	Affymetrix Human Genome U95A Array	5	5
GSE7307	Affymetrix Human Genome U133 Plus 2.0 Array	5	9
GSE12021(Huber et al., 2008)	Affymetrix Human Genome U133A Array	12 [3m/9f]	9 [7m/2f]
	Affymetrix Human Genome U133B Array	12 [3m/9f]	4 [3m/1f]
GSE55235(Woetzel et al., 2014)	Affymetrix Human Genome U133A Array	10	10
GSE55457(Woetzel et al., 2014)	Affymetrix Human Genome U133A Array	13	10
GSE77298(Broeren et al., 2016)	Affymetrix Human Genome U133 Plus 2.0 Array	16	7

Datasets showing downregulation of AP1 proteins in the RA synovium.

This table includes the information about the downregulation of the AP1 proteins as obtained by different microarray datasets used in our study.

Dataset	FOSB	JUN	JUNB
GSE1919	Downregulation	No differential regulation	No differential regulation
GSE7307	No differential regulation	Downregulation	No differential regulation
GSE12021 [U133A]	Downregulation	Downregulation	Downregulation
GSE55235	Downregulation	Downregulation	Downregulation
GSE55457	Downregulation	Downregulation	Downregulation

Upregulated signaling pathways in the RA synovium and their KEGG categories.

The numbers prefixed with "hsa" are the KEGG identifiers for each pathway. The KEGG pathway term name along with the KEGG pathway identifier is listed in the column named "Pathways". The "Fold enrichment" is calculated based on the number of upregulated genes which belong to the KEGG pathway term ("Count"), total number of upregulated genes which are part of the KEGG pathway database ("List Total"), total number of genes that belong to the KEGG Pathway database ("Population Total") and number of the KEGG Pathway database genes which are part of the KEGG pathway term ("Population Hits"). Precisely, the "Fold enrichment" is the proportion of "Count"/ "List Total" to "Population Hits"/ "Population Total" The "EASE score" represents the significance of obtaining "Count" genes in the "List Total" for the given KEGG pathway term which has "Population Hits" genes in the background "Population Total".

Pathways	KEGG Categories	Fold enrichment	EASE score
hsa04064:NF-κB signaling pathway	Signal transduction	3.92	$6.25\ e^{-9}$
hsa04664:Fc-ε RI signaling pathway	Immune system	3.19	$8.96\ e^{-5}$
hsa04621:NOD-like receptor signaling pathway	Immune system	3.11	$6.68\ e^{-4}$
hsa04662:B cell receptor signaling pathway	Immune system	3.05	$1.53\ e^{-4}$
hsa04620:Toll-like receptor signaling pathway	Immune system	2.89	$1.35\ e^{-5}$
hsa04062:Chemokine signaling pathway	Immune system	2.82	$4.96\ e^{-9}$
hsa04660:T cell receptor signaling pathway	Immune system	2.30	$1.69\ e^{-3}$
hsa04630:JAK-STAT signaling pathway	Signal transduction	1.96	$4.45\ e^{-3}$

Downregulated signaling pathways in the RA synovium and their KEGG categories.

The numbers prefixed with "hsa" are the KEGG identifiers for each pathway. The KEGG pathway term name along with the KEGG pathway identifier is listed in the column named "Pathways". The "Fold enrichment" is calculated based on the number of downregulated genes which belong to the KEGG pathway term ("Count"), total number of downregulated genes which are part of the KEGG pathway database ("List Total"), total number of genes that belong to the KEGG Pathway database ("Population Total") and number of the KEGG Pathway database genes which are part of the KEGG pathway term ("Population Hits"). Precisely, the "Fold enrichment" is the proportion of "Count"/ "List Total" to "Population Hits"/ "Population Total" The "EASE score" represents the significance of obtaining "count" genes in the "List Total" for the given KEGG pathway term which has "Population Hits" genes in the background "Population Total".

Pathways	KEGG Categories	Fold enrichment	EASE score
hsa04350:TGF-beta signaling pathway	Signal transduction	2.42	9.49 e^{-3}
hsa04152:AMPK signaling pathway	Signal transduction	2.30	2.56 e^{-3}
hsa04068:FoxO signaling pathway	Signal transduction	2.09	6.57 e^{-3}
hsa04022:cGMP-PKG signaling pathway	Signal transduction	1.90	9.85 e^{-3}
hsa04310:Wnt signaling pathway	Signal transduction	1.83	3.26 e^{-2}
hsa04024:cAMP signaling pathway	Signal transduction	1.62	4.37 e^{-2}
hsa04010:MAPK signaling pathway	Signal transduction	1.55	3.27 e^{-2}

Figure 1

The workflow.

(A). Selection of a KEGG pathway representing a phenotype exhibited by the RA synovium. The differentially expressed genes in the RA synovium were identified from the analysis of the microarray datasets obtained from GEO database. A KEGG pathway enrichment analysis was performed on the differentially expressed gene lists. The enriched pathways were categorized into process and signaling pathways. Based on the shared differentially expressed genes, a pathway overlap network was created. Osteoclast differentiation pathway was selected as the pathway of interest as it overlapped with most number of signaling pathways. **(B). Construction and analysis of the osteoclast differentiation network.** The proteins belonging to the KEGG osteoclast differentiation pathway were termed as "core proteins". The core proteins were used as an input to the String DB to obtain the proteins interacting with them (first shell proteins). The interactions among all the proteins (core and first shell) were also extracted from String DB. The interactions were validated using PubMed. The directions of the protein interactions were also obtained. The proteins and the interactions were used to construct a network for osteoclast differentiation. The gene ontology term enrichment analysis was performed on the network. Differentially regulated genes were indicated in the network. The gene ontology term enrichment analysis and the differentially regulated genes were used to identify the important protein interactions that lead to osteoclast differentiation in the RA synovium. The protein targets of the drugs used in RA treatment were obtained from DrugBank database and were indicated in the network.

A

B
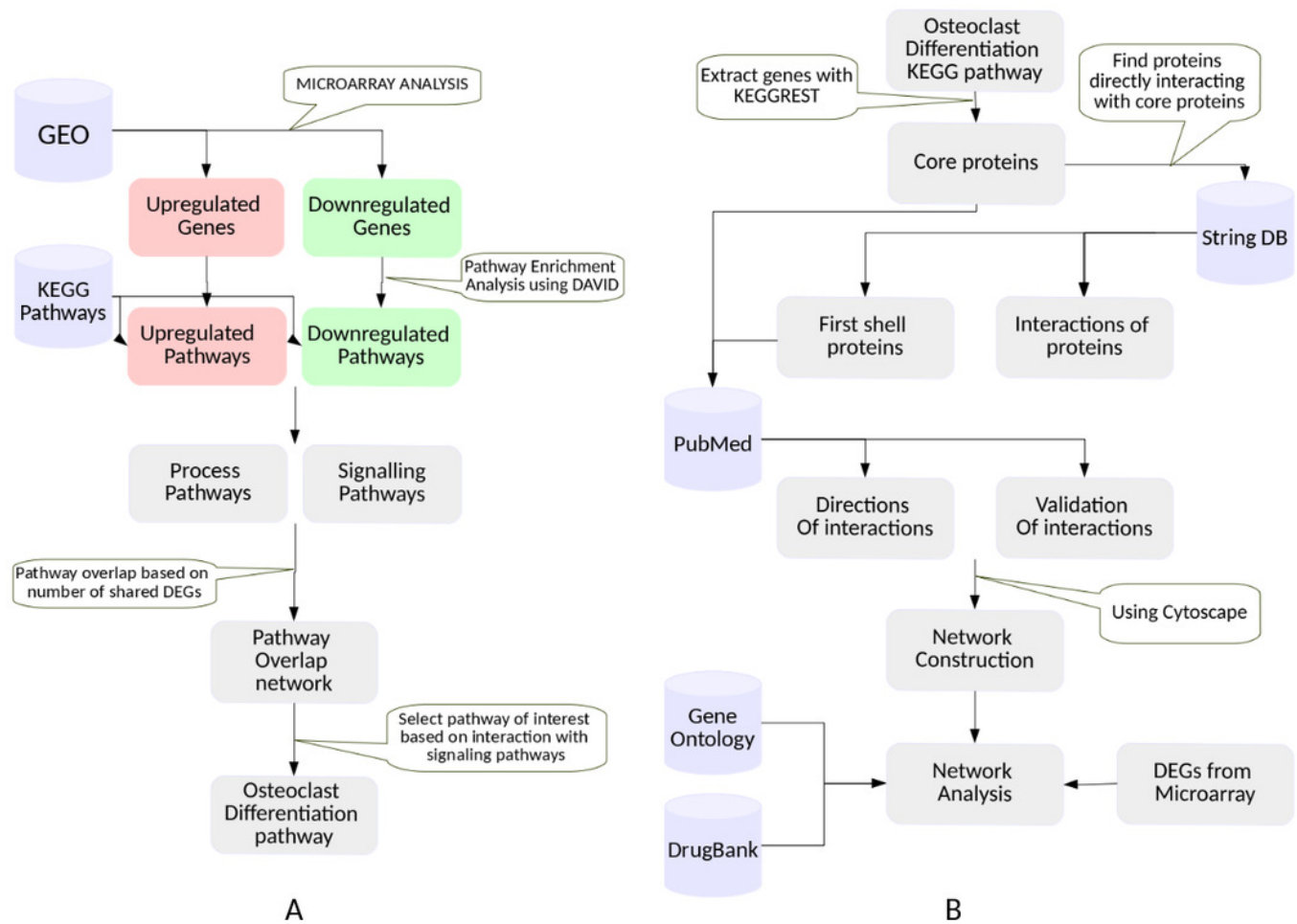
Figure 2

Number of up and downregulated genes obtained based on the selection criteria.

The numbers on x-axis represent the following: "1": selection of the genes at least once by the algorithm MAS5 and RMA, "2": selection of the genes at least twice by both the algorithms, "3": selection of the genes at least thrice by both the algorithms and so on. The y-axis represents the number of selected up and downregulated genes. The number of differentially expressed genes selected at least once by both the algorithms are used in this study.
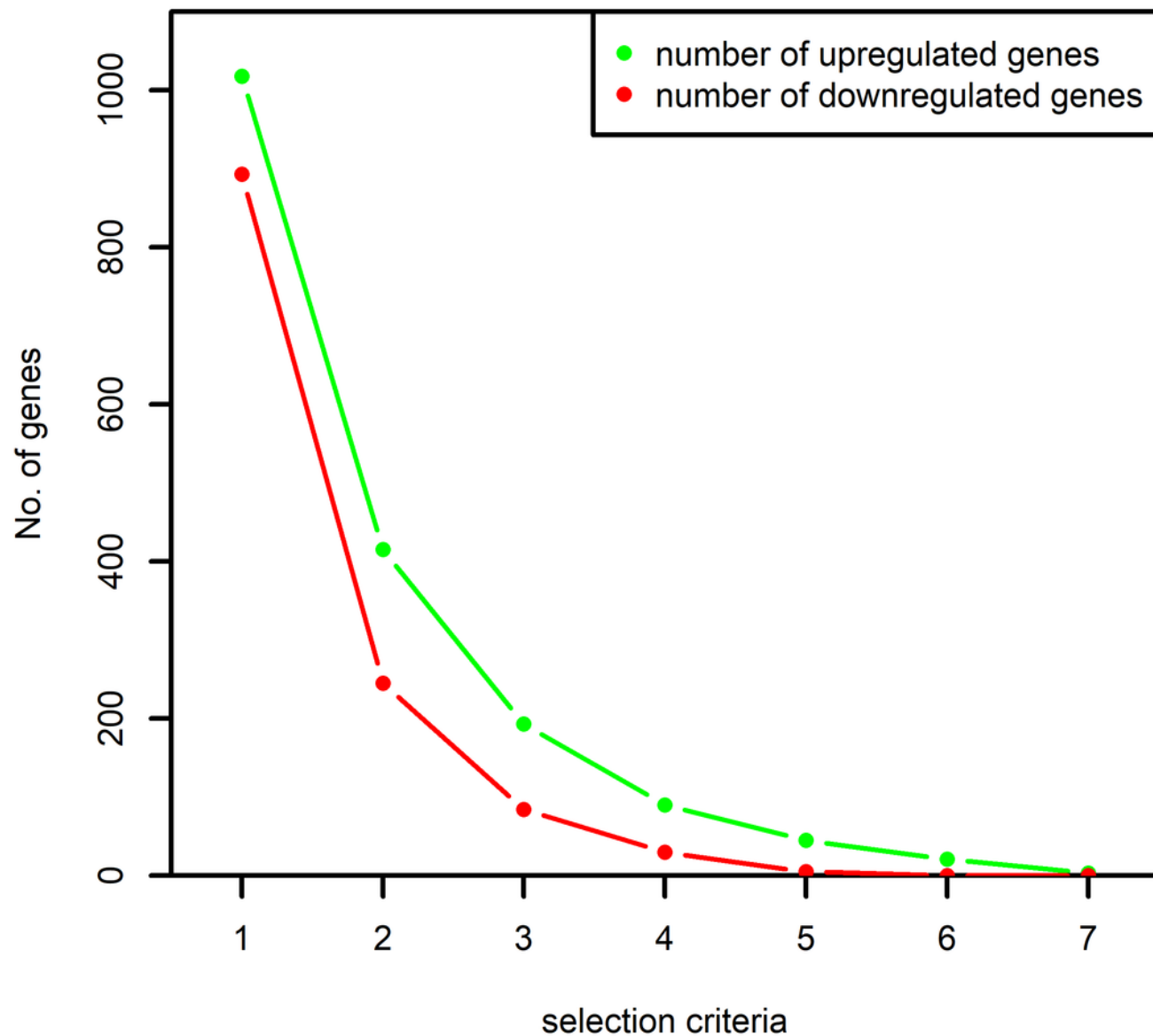
Figure 3

Number of up and downregulated genes shared by pairs of pathways.

The numbers on x-axis represent the sum of all the genes belonging to the pathway pairs. The y-axis represents the number of up and downregulated genes common to the pairs. The pathway pairs sharing at least 5 up or 5 downregulated genes are used in this study.
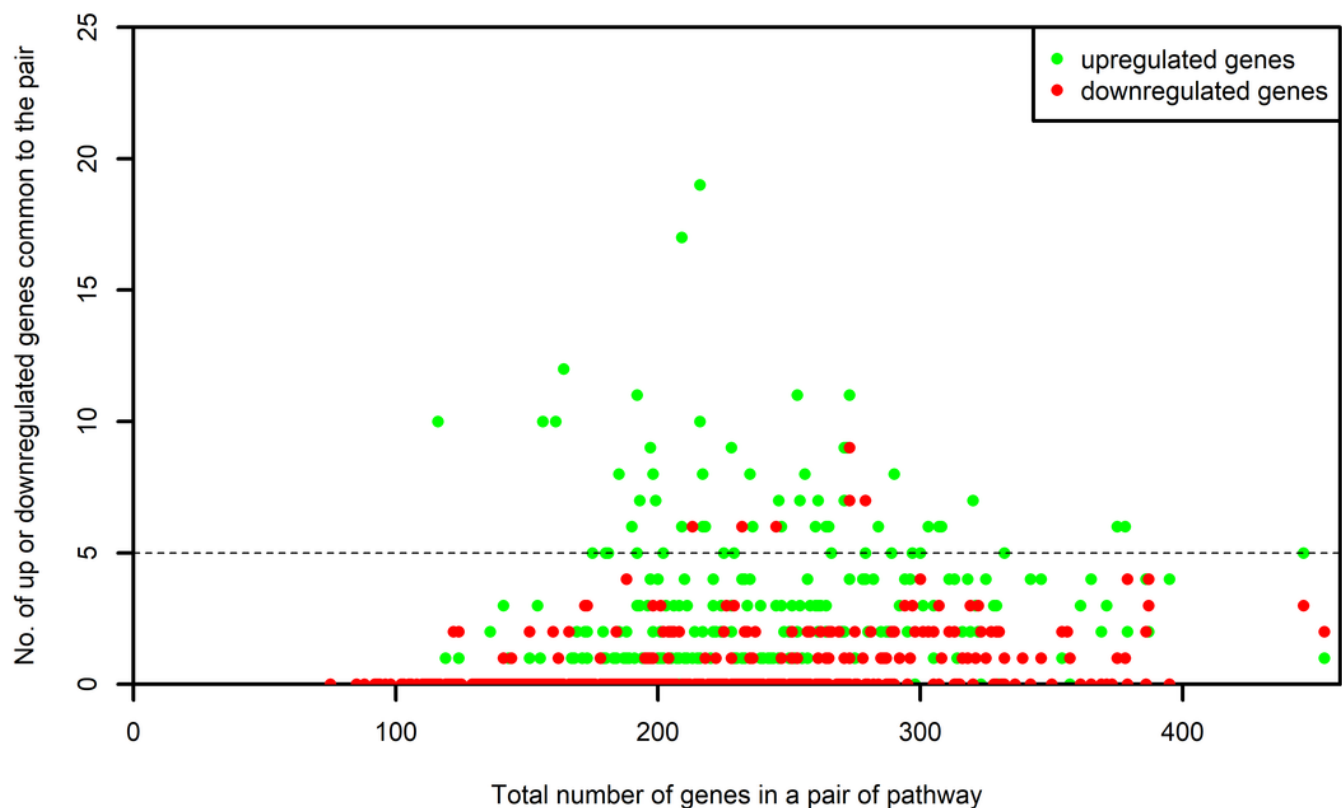
Figure 4

The KEGG category-wise distribution of the enriched pathways in the upregulated genes of the RA synovium.

The x-axis represents the KEGG categories as listed in the KEGG database whereas the y-axis denotes the number of enriched pathways from our study belonging to the KEGG categories.
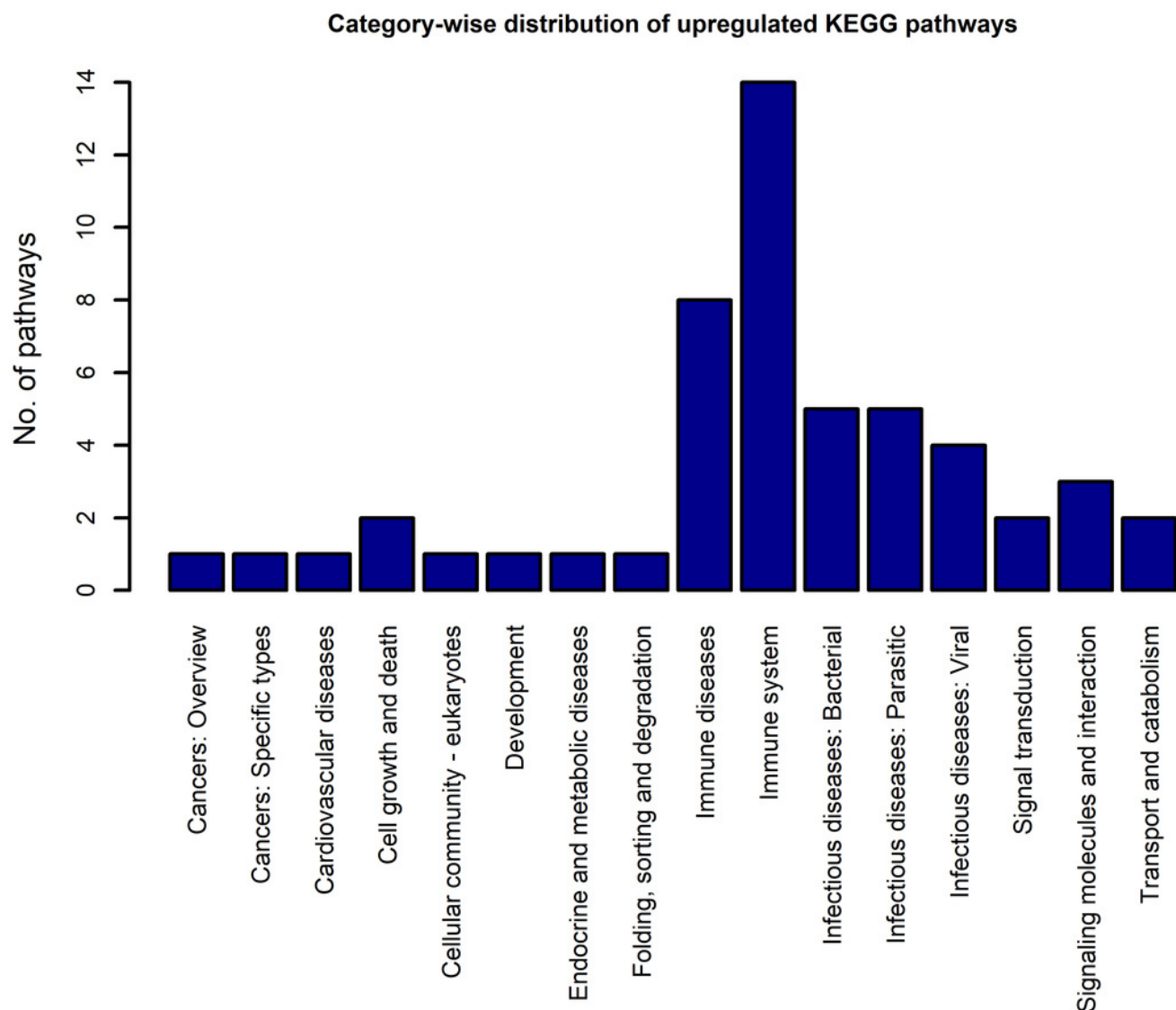
Figure 5

The KEGG category-wise distribution of the enriched pathways in the downregulated genes of the RA synovium.

The x-axis represents the KEGG categories as listed in the KEGG database whereas the y-axis denotes the number of enriched pathways from our study belonging to the KEGG categories.

Category-wise distribution of downregulated KEGG pathways
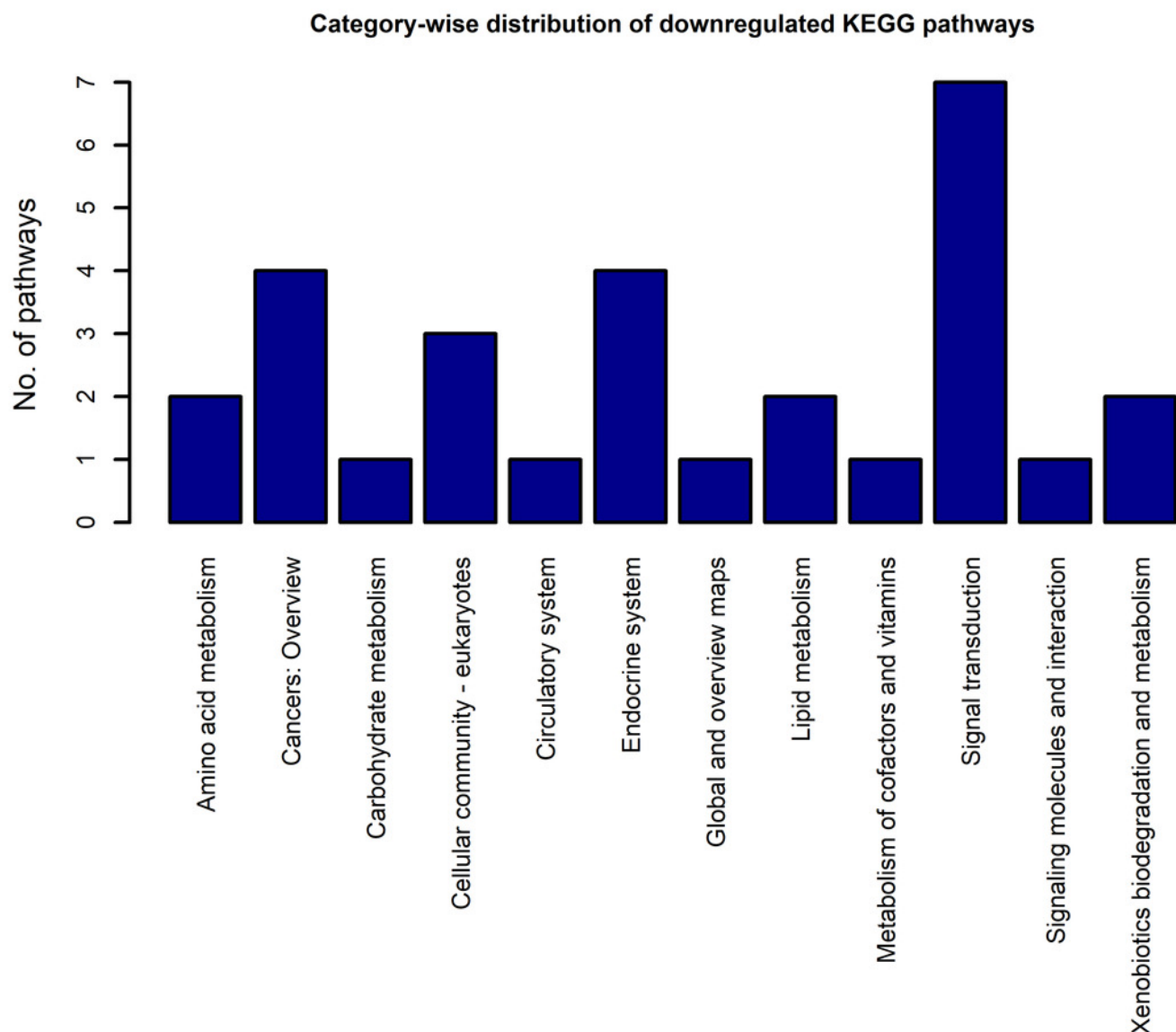
Figure 6

The interaction network of overlapping pathways.

The blue and red pathway nodes denote the process and the signaling pathways respectively. The white pathway nodes denote the other non-disease pathways. The connection between the pathway nodes represents sharing of at least five upregulated or five downregulated genes by the pair of pathway nodes. The size of the node represents the number of genes present in the pathway. Larger nodes show pathways having more number of genes, and smaller nodes represents pathways with lower number of genes.
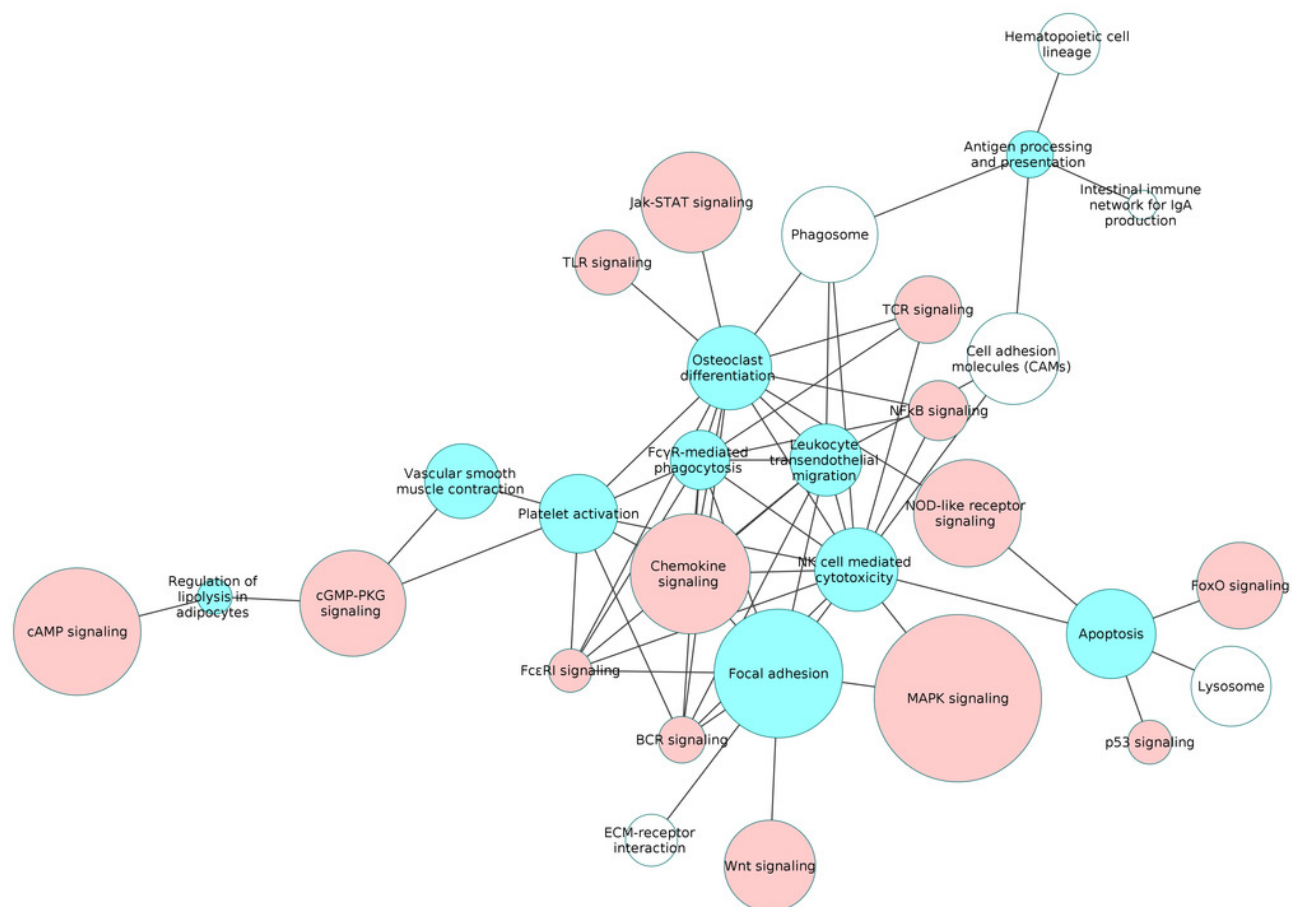
Figure 7

Overlap analysis of the upregulated process pathway osteoclast differentiation with other enriched non-disease pathways in the RA synovium.

The x-axis represents the enriched non-disease pathways with which the osteoclast differentiation pathway share DEGs. The signaling pathways are represented by striped bars whereas the non-signaling pathways are represented by solid bars. The y-axis denotes the number of DEGs shared between the osteoclast differentiation pathway and each of the enriched non-disease pathways.

Osteoclast differentiation with up pathways

No. of shared DEGs

Legend:
- ■ up genes (dark green)
- ■ down genes (red)
- ▨ up signaling pathway genes (light green hatched)
- ▨ down signaling pathway genes (red hatched)

Categories (x-axis):
- Phagosome
- Chemokine signaling pathway
- Natural killer cell mediated cytotoxicity
- NF-kappa B signaling pathway
- Fc gamma R-mediated phagocytosis
- Leukocyte transendothelial migration
- Toll-like receptor signaling pathway
- Fc epsilon RI signaling pathway
- B cell receptor signaling pathway
- NOD-like receptor signaling pathway
- Platelet activation
- T cell receptor signaling pathway
- Jak-STAT signaling pathway

Figure 8

Overlap analysis of the enriched downregulated process pathways in the RA synovium.

The x-axis represents the enriched non-disease pathways with which the downregulated process pathways share DEGs. The signaling pathways are represented by striped bars whereas the non-signaling pathways are represented by solid bars. The y-axis denotes the number of DEGs shared between the downregulated process pathways and each of the enriched non-disease pathways.
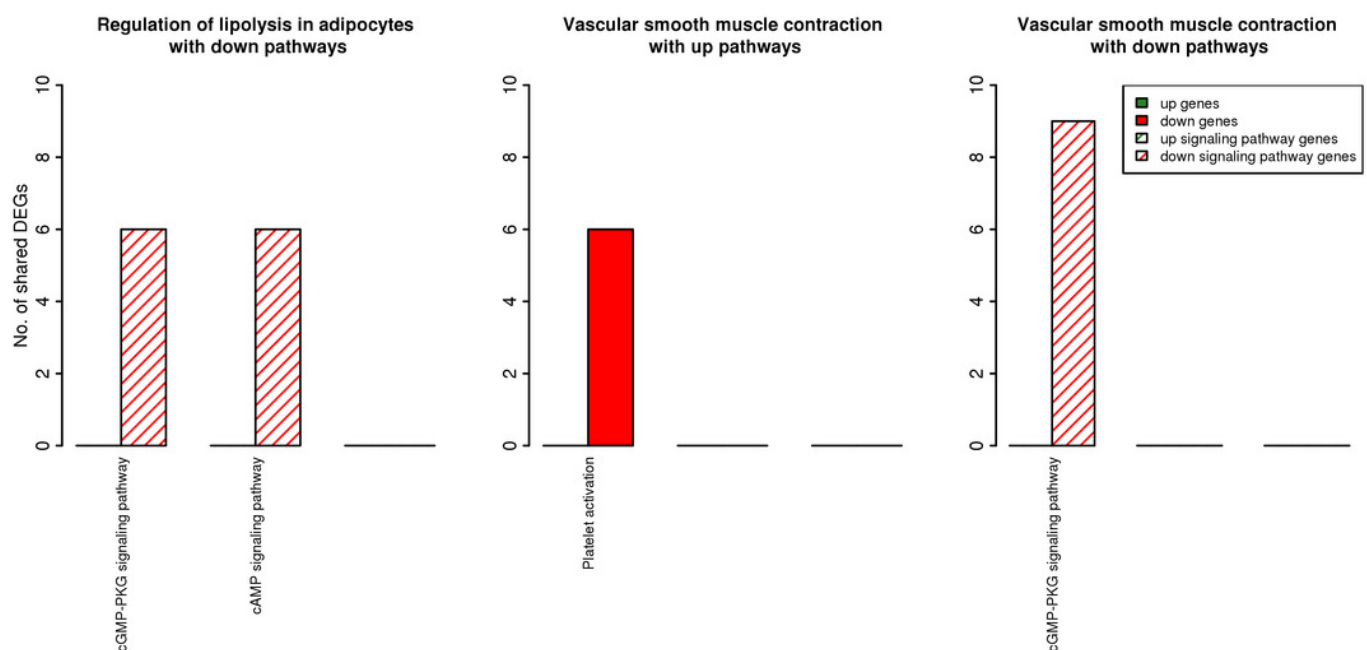
Figure 9

Activation of NF- κB proteins in the osteoclast differentiation network.

The core proteins are represented by thick borders and the shell proteins by thin borders. Rectangle nodes represent the DNA-binding proteins. The RA drug targets are indicated by red borders. The degree of differential regulation of the nodes is denoted as follows: red to grey - downregulation and green – upregulation. The edges of PRKCB are included from literature.
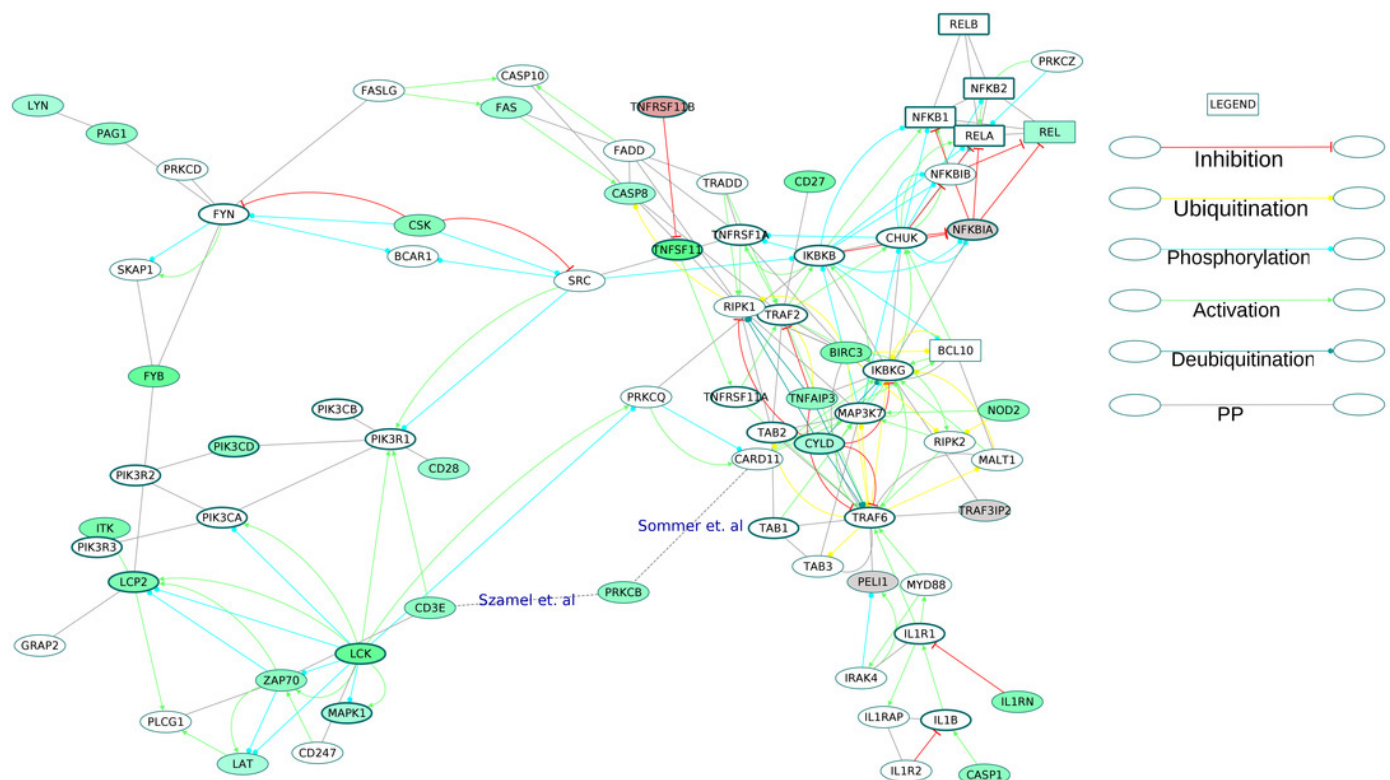
Figure 10

A schematic of enhanced ROS production mediating differentiation of osteoclasts in the RA synovium.

The degree of differential regulation of the nodes in the RA synovium is denoted as follows: red - downregulation and green – upregulation. The cytokine (TNFSF11) and the membrane receptors (CSF1R and CD3E) are above the horizontal line. Genes corresponding to the Nox2 complex and the Forkhead Box proteins are within rectangular boxes. The colors of the edges denote the following: green – activation, red – inhibition, blue – phosphorylation and black – PP. The dotted edges represent the interactions which involve other intermediate molecules. The inhibition of Forkhead box proteins by TNFSF11 are included from literature.