Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Network analysis of synovial RNA sequencing identifies gene-gene interactions predictive of response in rheumatoid arthritis	
Article Sub-Title		
Article CopyRight	The Author(s) (This will be the copyright line in the final PDF)	
Journal Name	Arthritis Research & Therapy	
Corresponding Author	FamilyName	**Pitzalis**
	Particle	
	Given Name	**Costantino**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Phone	
	Fax	
	Email	c.pitzalis@qmul.ac.uk
	URL	
	ORCID	
Corresponding Author	FamilyName	**Lewis**
	Particle	
	Given Name	**Myles J.**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Division	
	Organization	Digital Environment Research Institute, Queen Mary University of London
	Address	London, UK
	Phone	
	Fax	
	Email	myles.lewis@qmul.ac.uk
	URL	
	ORCID	
Author	FamilyName	**Sciacca**
	Particle	
	Given Name	**Elisabetta**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Division	Centre for Translational Bioinformatics
	Organization	William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London
	Address	London, UK
	Phone	

	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Surace**
	Particle	
	Given Name	**Anna E. A.**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Division	Centre for Translational Bioinformatics
	Organization	William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London
	Address	London, UK
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Alaimo**
	Particle	
	Given Name	**Salvatore**
	Suffix	
	Division	Department of Clinical and Experimental Medicine
	Organization	University of Catania
	Address	Catania, Italy
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Pulvirenti**
	Particle	
	Given Name	**Alfredo**
	Suffix	
	Division	Department of Clinical and Experimental Medicine
	Organization	University of Catania
	Address	Catania, Italy
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Rivellese**
	Particle	
	Given Name	**Felice**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Phone	

	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Goldmann**
	Particle	
	Given Name	**Katriona**
	Suffix	
	Division	Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry
	Organization	Queen Mary University of London
	Address	London, UK
	Division	Centre for Translational Bioinformatics
	Organization	William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London
	Address	London, UK
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Ferro**
	Particle	
	Given Name	**Alfredo**
	Suffix	
	Division	Department of Clinical and Experimental Medicine
	Organization	University of Catania
	Address	Catania, Italy
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	FamilyName	**Latora**
	Particle	
	Given Name	**Vito**
	Suffix	
	Division	School of Mathematical Sciences
	Organization	Queen Mary University of London
	Address	London, UK
	Division	Dipartimento di Fisica ed Astronomia
	Organization	Università di Catania and INFN
	Address	I-95123, Catania, Italy
	Phone	
	Fax	
	Email	
	URL	
	ORCID	

Schedule	Received	19 Jan 2022
	Revised	
	Accepted	4 May 2022

Abstract

Background:
To determine whether gene-gene interaction network analysis of RNA sequencing (RNA-Seq) of synovial biopsies in early rheumatoid arthritis (RA) can inform our understanding of RA pathogenesis and yield improved treatment response prediction models.

Methods:
We utilized four well curated pathway repositories obtaining 10,537 experimentally evaluated gene-gene interactions. We extracted specific gene-gene interaction networks in synovial RNA-Seq to characterize histologically defined pathotypes in early RA and leverage these synovial specific gene-gene networks to predict response to methotrexate-based disease-modifying anti-rheumatic drug (DMARD) therapy in the Pathobiology of Early Arthritis Cohort (PEAC). Differential interactions identified within each network were statistically evaluated through robust linear regression models. Ability to predict response to DMARD treatment was evaluated by receiver operating characteristic (ROC) curve analysis.
Results:
Analysis comparing different histological pathotypes showed a coherent molecular signature matching the histological changes and highlighting novel pathotype-specific gene interactions and mechanisms. Analysis of responders vs non-responders revealed higher expression of apoptosis regulating gene-gene interactions in patients with good response to conventional synthetic DMARD. Detailed analysis of interactions between pairs of network-linked genes identified the *SOCS2/STAT2* ratio as predictive of treatment success, improving ROC area under curve (AUC) from 0.62 to 0.78. We identified a key role for angiogenesis, observing significant statistical interactions between *NOS3* (eNOS) and both *CAMK1* and eNOS activator *AKT3* when comparing responders and non-responders. The ratio of *CAMKD2/NOS3* enhanced a prediction model of response improving ROC AUC from 0.624 to 0.726.
Conclusions:
We demonstrate a novel, powerful method which harnesses gene interaction networks for leveraging biologically relevant gene-gene interactions leading to improved models for predicting treatment response.

Keywords (separated by '-')	Rheumatoid arthritis - RNA sequencing - Synovial biopsy - Network analysis - Pathobiology of Early Arthritis Cohort study (PEAC)
Footnote Information	Costantino Pitzalis and Myles J. Lewis are co-senior authors.The online version contains supplementary material available at https://doi.org/10.1186/s13075-022-02803-z.

Arthritis Research & Therapy

RESEARCH Open Access

Network analysis of synovial RNA sequencing identifies gene-gene interactions predictive of response in rheumatoid arthritis

Elisabetta Sciacca[1,2], Anna E. A. Surace[1,2], Salvatore Alaimo[3], Alfredo Pulvirenti[3], Felice Rivellese[1], Katriona Goldmann[1,2], Alfredo Ferro[3], Vito Latora[4,5], Costantino Pitzalis[1*] and Myles J. Lewis[1,6*]

Abstract

Background: To determine whether gene-gene interaction network analysis of RNA sequencing (RNA-Seq) of synovial biopsies in early rheumatoid arthritis (RA) can inform our understanding of RA pathogenesis and yield improved treatment response prediction models.

Methods: We utilized four well curated pathway repositories obtaining 10,537 experimentally evaluated gene-gene interactions. We extracted specific gene-gene interaction networks in synovial RNA-Seq to characterize histologically defined pathotypes in early RA and leverage these synovial specific gene-gene networks to predict response to methotrexate-based disease-modifying anti-rheumatic drug (DMARD) therapy in the Pathobiology of Early Arthritis Cohort (PEAC). Differential interactions identified within each network were statistically evaluated through robust linear regression models. Ability to predict response to DMARD treatment was evaluated by receiver operating characteristic (ROC) curve analysis.

Results: Analysis comparing different histological pathotypes showed a coherent molecular signature matching the histological changes and highlighting novel pathotype-specific gene interactions and mechanisms. Analysis of responders vs non-responders revealed higher expression of apoptosis regulating gene-gene interactions in patients with good response to conventional synthetic DMARD. Detailed analysis of interactions between pairs of network-linked genes identified the *SOCS2/STAT2* ratio as predictive of treatment success, improving ROC area under curve (AUC) from 0.62 to 0.78. We identified a key role for angiogenesis, observing significant statistical interactions between *NOS3* (eNOS) and both *CAMK1* and eNOS activator *AKT3* when comparing responders and non-responders. The ratio of *CAMKD2/NOS3* enhanced a prediction model of response improving ROC AUC from 0.624 to 0.726.

Conclusions: We demonstrate a novel, powerful method which harnesses gene interaction networks for leveraging biologically relevant gene-gene interactions leading to improved models for predicting treatment response.

[†]Costantino Pitzalis and Myles J. Lewis are co-senior authors.

*Correspondence: c.pitzalis@qmul.ac.uk; myles.lewis@qmul.ac.uk

[1] Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK
[6] Digital Environment Research Institute, Queen Mary University of London, London, UK
Full list of author information is available at the end of the article

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

Keywords: Rheumatoid arthritis, RNA sequencing, Synovial biopsy, Network analysis, Pathobiology of Early Arthritis Cohort study (PEAC)

Background

Differential gene expression analysis is a common starting point for many gene expression studies. However, this only reveals differences at the level of individual genes. The identification of statistical interactions between pairs of genes can enhance understanding of biological processes and functional mechanisms which are active within tissues. However, the large number (20,000-50,000) of expressed genes detectable by RNA-Seq renders analysis of all possible gene-gene correlations (of the order of 10^9) computationally time consuming and confounded by substantial numbers of false positive gene-gene pairs which are biologically and functionally unrelated.

In the current study, we developed a novel network tool integrating information from four pathway repositories [1–4] obtaining 10,537 gene-gene interactions. The gene-gene interactions include protein-protein interactions which have been reported from experiments in the literature including co-immunoprecipitation, yeast-2-hybrid and direct molecular biology studies, as well as gene-gene interactions based on integration of microRNA and transcriptome data [5]. This network tool was applied to RNA-Seq data on synovial biopsies and blood samples from early rheumatoid arthritis (RA) patients to characterize statistical differences in gene-gene interactions between histologically defined RA subgroups known as pathotypes [6, 7] and between responders and non-responders to conventional synthetic disease-modifying anti-rheumatic drugs (csDMARD).

Patients treated with csDMARD are often subject to lack of treatment efficacy [8, 9]. Several studies have tried to predict patients' responsiveness based on synovial gene expression, mostly from joint replacement tissue. However, the presence of concomitant immunosuppressive medications and use of microarrays are major limitations of these studies [10–13].

For these reasons in the present work, we used RNA-seq data from the Pathobiology of Early Arthritis Cohort [6] where synovial biopsies and blood samples were taken from a cohort of 94 early, treatment-naïve RA patients. In this cohort, Lewis et al. [6] identified three histological and molecular subgroups characterized by (i) B cell infiltration (*lympho-myeloid* pathotype), (ii) macrophage infiltration (*diffuse-myeloid* pathotype), and (iii) absence of immune cells with stromal cell predominance (*pauci-immune fibroid* pathotype). In the present study, we use a novel network approach to identify functionally relevant gene-gene interactions that were not highlighted before and which are for the first time associated with response to csDMARD at 6 months through robust linear modeling incorporating interaction terms. While the previous study could not derive any prediction model using single gene expressions, here we demonstrate that the use of a new, network-based tool can detect significant gene-gene pairs that improved predictive models of response to csDMARD treatment as tested by receiver operating characteristic (ROC) curve analysis.

Methods

This study used the dataset described in Lewis et al. [6, 7] where RNA-Seq data from 94 early, treatment-naïve RA patients fulfilling the 2010 ACR/EULAR criteria was collected. Eleven samples had ungraded histopathology or were removed due to poor RNA quality, leaving 83 samples with RNA-Seq and matched histology in the present study (Table 1). Patients were stratified following the same histopathological classification described in the previous work [5]: lympho-myeloid, diffuse-myeloid, and pauci-immune. After a baseline synovial biopsy and blood sample collection (treatment-naïve), patients underwent 6 months of methotrexate-based csDMARD therapy. Responsiveness was assessed according to DAS28 EULAR criteria. Our study compared both histopathological and treatment response groups, with separate analyses run for each classification (Table 1).

The analytical pipeline summarized in Fig. 1A shows the steps through which informative gene networks and predictive gene pairs were extracted for each classification. In brief, an extensive network of curated protein-protein and gene-gene interactions was built by merging KEGG pathways with micro-RNA and transcription factor databases [5]. The network was replicated for each subgroup and average gene expressions were used to infer weights on network nodes. Networks were then filtered by weight removing genes with mean expression under the 75th percentile. Adjacent gene-gene pairs within the network overlapping across subgroups were also removed, thus revealing subgroup-specific networks where statistical significance of the gene-gene links was then assessed using robust linear regression on RNA-Seq gene expression data. When running this pipeline for DAS28-ESR EULAR response categorization, pathway linked gene-gene pairs which demonstrated a statistically significant interaction in a linear model were subsequently selected for incorporation into a logistic

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

Table 1 Baseline demographics of treatment-naïve RA patients recruited into the Pathobiology of Early Arthritis Cohort (PEAC)

	Lymphoid ($N=49$)	Myeloid ($N=18$)	Fibroid ($N=16$)	Total ($N=83$)	*p*-value
Age (years)	52.3 (16.2)	50.4 (16.5)	53.2 (15.0)	52.1 (15.9)	0.865
Gender					0.608
F	37 (75.5%)	12 (66.7%)	13 (81.2%)	62 (74.7%)	
M	12 (24.5%)	6 (33.3%)	3 (18.8%)	21 (25.3%)	
Disease duration (months)	5.9 (3.3)	4.8 (2.6)	7.0 (3.5)	5.8 (3.3)	0.152
RF					0.203
pos	32 (69.6%)	9 (52.9%)	7 (46.7%)	48 (61.5%)	
neg	14 (30.4%)	8 (47.1%)	8 (53.3%)	30 (38.5%)	
CCP					0.025
pos	39 (84.8%)	10 (55.6%)	9 (60.0%)	58 (73.4%)	
neg	7 (15.2%)	8 (44.4%)	6 (40.0%)	21 (26.6%)	
DAS28	6.2 (1.2)	5.6 (1.2)	5.2 (1.6)	5.8 (1.3)	0.029
ESR (mm/hr)	50.9 (28.1)	37.6 (25.4)	30.8 (27.7)	44.1 (28.4)	0.025
CRP (µg/mL)	25.0 (26.5)	17.5 (26.0)	14.4 (41.7)	21.3 (29.9)	0.395
TJC	12.6 (7.2)	9.9 (6.3)	10.9 (8.9)	11.7 (7.4)	0.396
SJC	8.4 (5.7)	7.1 (4.3)	5.2 (5.0)	7.5 (5.4)	0.116
VAS	67.9 (24.0)	61.1 (21.7)	57.8 (27.1)	64.5 (24.2)	0.287
HAQ	1.6 (0.8)	1.4 (0.6)	1.6 (0.8)	1.5 (0.7)	0.499
DAS28 EULAR					0.603
Good	15 (36.6%)	4 (30.8%)	7 (53.8%)	26 (38.8%)	
Moderate	20 (48.8%)	8 (61.5%)	4 (30.8%)	32 (47.8%)	
None	6 (14.6%)	1 (7.7%)	2 (15.4%)	9 (13.4%)	

regression model to predict EULAR response. See supplementary methods for full details of methods at each stage of the analysis pipeline. For comparison, predictive models on synovial RNA-Seq were compared with blood RNA-Seq in matched patients ($n=67$), of which 59 patients had matched histology. Genotyping was performed as previously described [14]. HLA imputation is described in the Supplementary methods.

Results

Lympho-myeloid pathotype gene network is associated with leukocyte chemokines, chemoattractants and antigen processing

Synovial biopsies from patients with RA demonstrate distinctive histological pathotypes associated with corresponding gene signatures [6, 7]. In the present analysis gene-gene interaction networks specific for the

(See figure on next page.)

Fig. 1 Network analysis of synovial RNA sequencing in early RA reveals gene-gene interactions uniquely linked to the lympho-myeloid pathotype. **A** Analytical pipeline using network approach to extract informative networks and predictive gene pairs from RNA-seq profiles. Having defined subgroups of patients, an extensive network of interactions is built using merged KEGG pathways enriched with micro-RNAs and transcription factors. The network is replicated for each subgroup and the average expression level of each gene in a subgroup is used to infer a weight on each network node. A first filtering step removes, from each network, nodes (genes) whose weight (subgroup average expression level) is below an optimal threshold obtained via percolation analysis. The second filtering step pull out links (gene-gene interactions) overlapping two or more networks. Robust linear regression with interaction term is used to extract significant gene-gene links. A logistic regression model is built for each significant gene-gene pair to predict response. Ability to predict response is tested by receiver operating characteristic (ROC) curve analysis. **B** Network of unique active interactions in the lympho-myeloid pathotype. Clusters LM1-LM4. Selected clusters of interest. Labels are determined by gene ontology (GO)/pathway enrichment analysis. Percentages indicate the number of cluster genes included in the associated GO/pathway term. Cluster LM1. Cluster of chemokines needed for leukocyte recruitment (93.5% enrichment). Cluster LM2. Antigen processing and presentation with T cell activation genes (100% enrichment). Cluster LM3. Group of focal adhesion genes comprising collagens, integrins and laminins (93.9% enrichment). Cluster LM4. TNF signaling through mTOR (48.8% enrichment). Cluster LM5. Interferon regulation signaling (87.5% enrichment). Cluster LM6. Genes of the intrinsic and extrinsic apoptotic pathways (50.8% enrichment). **C** Correlation plots showing differential gene-gene correlations with interactions associated with pathotype. Statistical analysis by robust linear regression model. *p*-value of the gene to pathotype interacting term is shown. Correlation plots of gene pairs *CD28* and *PIK3R1*, *CD79A* and *LYN*, and *TNC* and *ITGB7* across different pathotypes

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

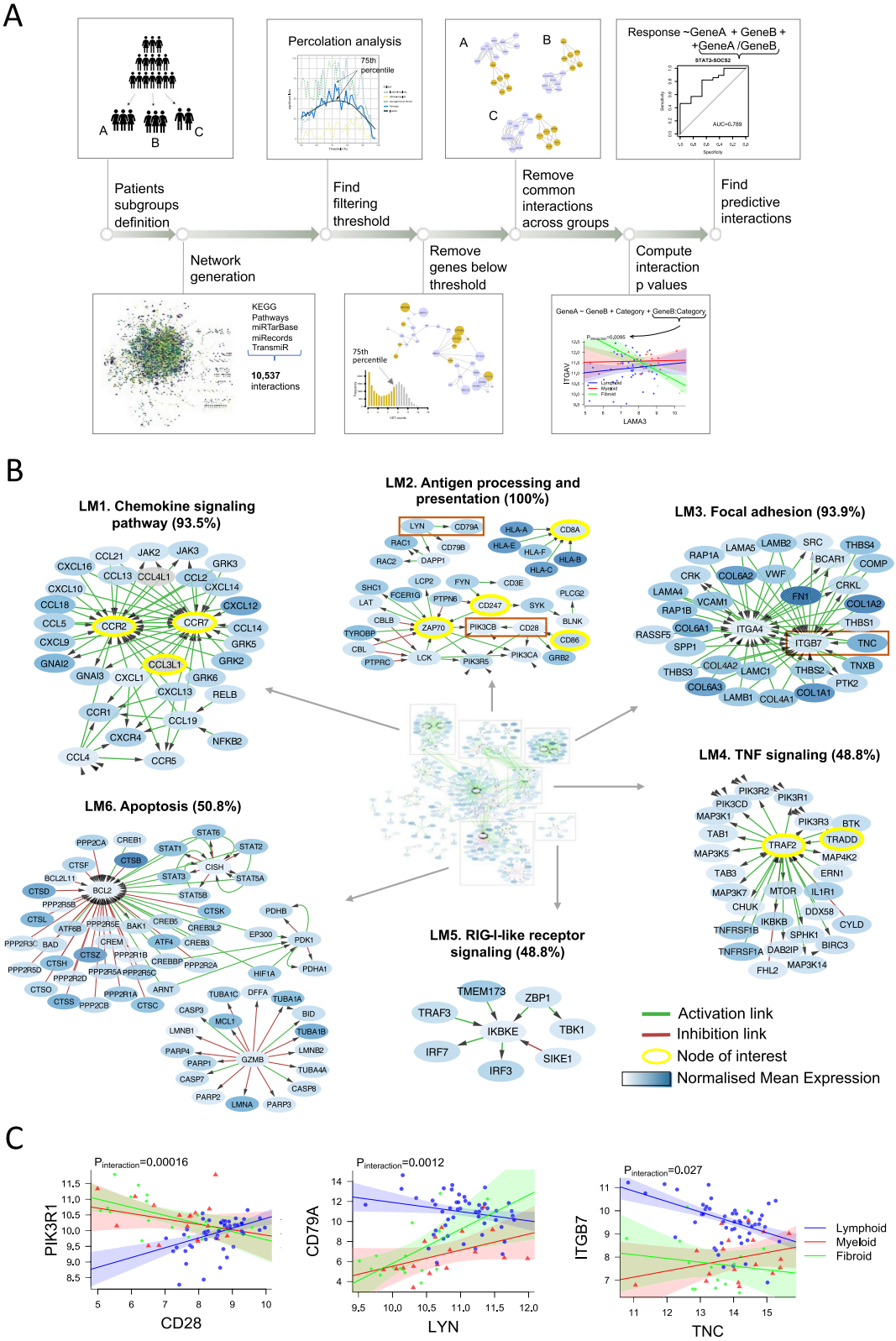

Fig. 1 (See legend on previous page.)

Journal : **BMCTwo 13075** | Dispatch : **16-5-2022** | Pages : **14**

Article No : **2803** | ☐ LE | ☐ TYPESET

MS Code : | ☑ CP | ☑ DISK

lympho-myeloid, diffuse myeloid and pauci-immune fibroid pathotypes were generated (Fig. 1A, see Supplementary Methods for more detail). In Fig. 1B, the six most prominent clusters of the lympho-myeloid specific network are shown (full network Fig. S1). Gene ontology (GO) enrichment analysis was employed to label gene clusters associated with chemokine signaling (LM1), antigen processing/presentation (LM2), focal adhesion (LM3), tumor necrosis factor (TNF) signaling (LM4), retinoic acid-inducible gene (RIG)-I-like receptor signaling (LM5), and apoptosis (LM6).

The presence of multiple chemokine signaling elements including *CCR1* and its ligands *CCL5* and *CCL14*, *CCR2* together with its ligand *CCL2* and *CCR7* and its ligands *CCL19* and *CCL21*, suggests that local activation and recruitment of lymphocytes to inflamed joints is a central feature of the lympho-myeloid pathotype (Fig. 1B, cluster LM1) [15, 16].

The lympho-myeloid network contained LM2, an antigen processing and presentation cluster (Fig. 1B) with prominence of T cell genes including *CD8A* with different major histocompatibility complex (MHC) class I genes and T cell activation genes (*CD247*, *ZAP70*, *CD28*, *CD86*). Using a robust linear model, the lympho-myeloid pathotype showed a significant difference ($p = 0.00016$) in correlation between *CD28* and the class I phosphoinositide 3-kinase (PI3K) signaling regulator *PIK3R1*, which is important for T cell function downstream of CD28, when compared to the other two pathotypes. B cell recruitment and stimulation was also implied by the presence of *CXCL13* and the *CD79A-LYN* link (Fig. 1B, cluster LM1). Phosphorylation of the B cell receptor binding CD79a by Lyn kinase is an initial event in B cell receptor engagement. Differential correlation between *LYN* and *CD79A* was observed in the lympho-myeloid subgroup ($p = 0.0012$) compared to the diffuse-myeloid and pauci-immune fibroid subgroups (Fig. 1C, Table S1).

Invasion and migration of cells requires interaction with the extracellular matrix through macromolecular assemblies known as focal adhesions. Cluster LM3 (Fig. 1B) included multiple focal adhesion genes including collagens, laminins, integrins, and Tenascin C (*TNC*), which plays an important role in the development and regeneration of articular cartilage [17] and whose interaction with integrins has been widely studied in cancer [18]. Correlation between *ITGB7* and *TNC* was poor in the diffuse-myeloid and pauci-immune fibroid subgroups but significantly stronger in the lympho-myeloid subgroup ($p = 0.027$).

Other active pathways in the lympho-myeloid pathotype included NF-kB and mammalian target of rapamycin (mTOR) signaling as part of chemokine and TNF signaling (Fig. 1B, clusters LM1 and LM4), and RIG-I-like receptor signaling centered around inhibitor of nuclear factor kappa B kinase subunit epsilon (*IKBKE)* (cluster LM5). Increased cell turnover in the lympho-myeloid pathotype is suggested by cluster LM6 which contained intrinsic and extrinsic apoptosis-related genes *TRAF2*, *BAD*, *BAK*, *BCL2*, and cytotoxic T cell marker *GZMB* (granzyme B).

Macrophage activation and T cell activation underlie the diffuse-myeloid pathotype gene network

The diffuse-myeloid specific network was of much smaller size (Fig. 2A, full network Fig. S2) after common links were removed, which may reflect the fact that this subgroup has overlapping characteristics with both of the other two pathotypes. On one hand, this category is characterized by the infiltration of macrophages, which are also present in the lympho-myeloid subgroup; on the other hand, the absence of B and plasma cell aggregates is a feature in common with the pauci-immune fibroid pathotype.

One cluster with the same associated GO term as observed in the lympho-myeloid subgroup was for focal adhesion (Fig. 2A, cluster DM1), consistent with the role of integrins in macrophage infiltration into tissues. Uniquely for the diffuse-myeloid pathotype, genes from the PPAR (peroxisome proliferator-activated receptor) signaling pathway, which is involved in fatty acid storage and has been linked to pathological synovial inflammation in RA [19–21], were observed in this network (Fig. 2A, cluster DM2). PPAR-γ (*PPARG*), which is critical for macrophage reprogramming [22], and surrounding network genes including adiponectin (*ADIPOQ*) and its receptor (*ADIPOR2*) are significantly upregulated in the diffuse-myeloid pathotype specifically.

(See figure on next page.)

Fig. 2 PPAR-γ signaling is key driver of the diffuse-myeloid pathotype while Wnt/Notch signaling pathways characterize the pauci-immune fibroid pathotype. **A** Network of unique active interactions in the diffuse-myeloid pathotype. Cluster DM1. Extracellular matrix genes for focal adhesion (75.6% enrichment). Cluster DM2. Cluster of PPAR signaling pathway (78.6% enrichment). **B** Network of unique active interactions in the pauci-immune fibroid pathotype. Cluster PF1. Group of focal adhesion genes comprising collagens, integrins and laminins (93.3% enrichment). Cluster PF2. Cluster of genes of the Ras signaling pathway (76.6% enrichment). Cluster PF3. Clusters of Notch-, Wnt- and TGF-beta signaling (95.8% enrichment). Cluster PF4. Cytokine-cytokine interaction of pro-inflammatory genes (100% enrichment). **C** Correlation plots showing differential gene-gene correlations with interactions associated with pathotype. Statistical analysis by robust linear regression model. *p*-value of the gene to pathotype interacting term is shown. Regression plot of *ITGAV* and *LAMA3*, *WNT11* and *SFRP2* in different pathotypes

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

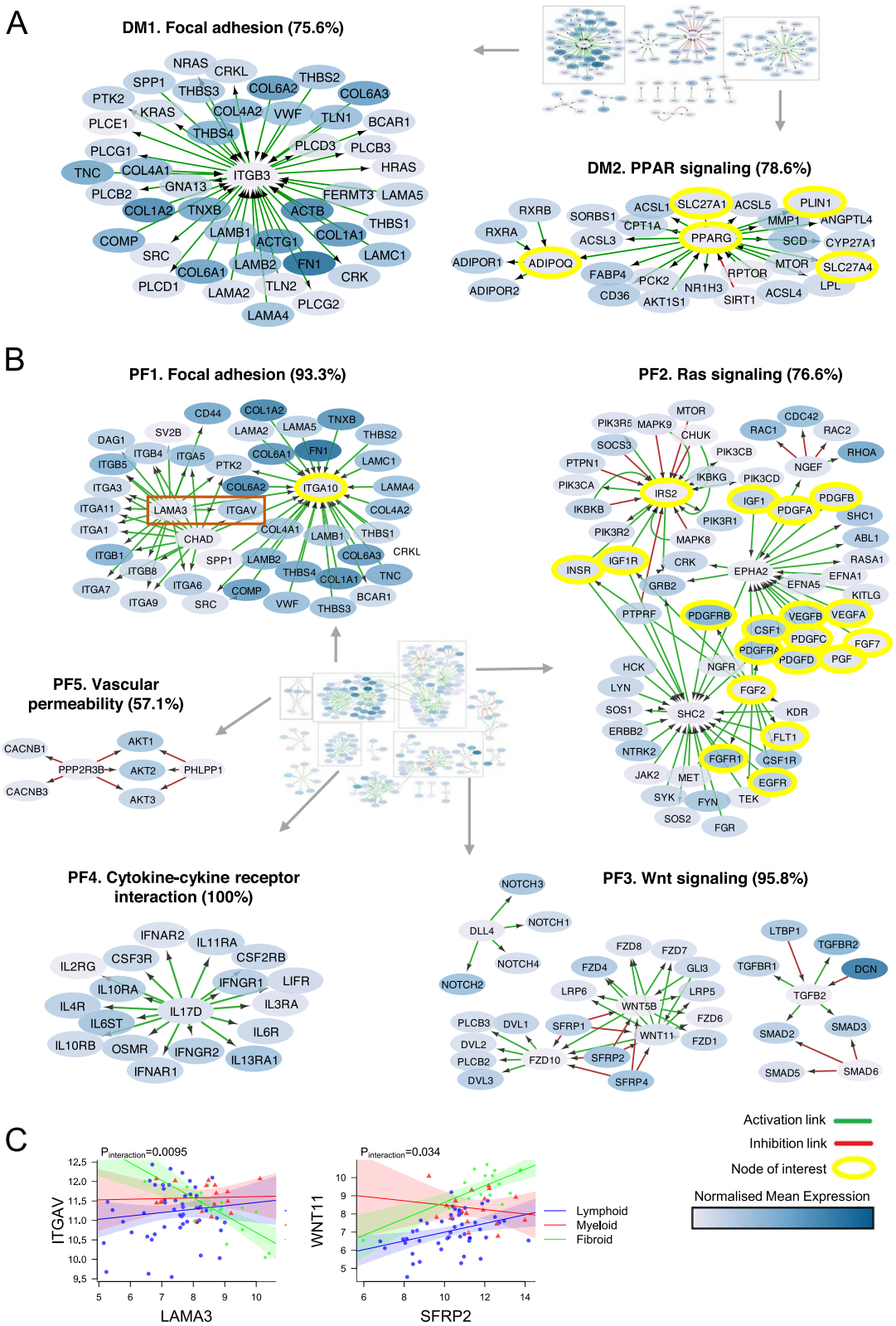

Fig. 2 (See legend on previous page.)

Journal : **BMCTwo 13075** | Dispatch : **16-5-2022** | Pages : **14**
Article No : **2803** | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

The Wnt signaling pathway characterizes the pauci-immune fibroid pathotype network

Compared to the diffuse-myeloid pathotype, the pauci-immune fibroid pathotype had a more extensive network (Fig. 2B, full network Fig. S3). Extracellular matrix genes including collagens (*COL1A1,* etc.) and laminins (*LAMB1/2*, etc.) were present as an overlapping theme across all three pathotypes in fibroid cluster PF1 (Fig. 2B), lympho-myeloid cluster LM3 (Fig. 1B) and diffuse-myeloid cluster DM1 (Fig. 2A), with different integrins (*ITGA4*, *ITGB7*, *ITGB3*, *ITGA10*) as hubs. Of these, the fibroid hub *ITGA10* is highly expressed by chondrocytes [23] and selectively binds collagen [24]. Significant statistical interactions across pathotypes were observed for correlations of *ITGA10* and several neighboring gene nodes (Fig. S4). Another hub node, chondroadherin (*CHAD*), is a cartilage matrix protein that promotes attachment of chondrocytes, fibroblasts, and osteoblasts [25]. *CHAD* was most strongly correlated with *ITGA10*, *ITGB4*, and *ITGA3* in the lympho-myeloid pathotype (Fig. S5). Among the genes linked to laminin alpha-3 (*LAMA3*), the integrin subunit alpha V (*ITGAV*) is of particular interest since polymorphisms of this gene have been associated with both angiogenesis [26] and susceptibility to RA [27]. In our data, its interaction with *LAMA3* showed negative correlation in the pauci-immune fibroid subgroup in contrast to the other two pathotypes (Fig. 2C). These results suggest that *ITGA10*, *ITGAV*, *CHAD*, and *LAMA3* play central roles in differentiating the pathotypes.

In a separate cluster PF2 we observed several nodes related to the Ras signaling pathway (Fig. 2B), comprising epidermal, fibroblast, nerve, vascular endothelial, insulin-like and platelet-derived growth factors, associated with signaling molecules through Src, *MAPK*, and *PI3K*. Fibroblast related pathways were found in cluster PF6 linked to transforming growth factor (TGF)-beta and Wnt signaling pathway (cluster PF3), which included the secreted frizzled related proteins *SFRP1* and *SFRP2*. *SFRP1* and *2* are Wnt inhibitors which show reduced expression in RA [28, 29] synovium. Hence, it is notable that we found a stronger positive correlation between *SFRP2* and *WNT11* in the fibroid pathotype (Fig. 2C).

Another cluster of major interest was found around the pro-inflammatory cytokine Interleukin (IL)-17D in cytokine-cytokine receptor interaction cluster PF4 (Fig. 2B). IL-17 family members are involved in RA pathogenesis and *IL17D* is expressed in rheumatoid nodules [30]. Cluster PF4 comprised cytokine receptors playing key roles in RA, implicating pro-inflammatory activation of fibroblasts and stromal cells in the fibroid pathotype given the absence of immune effector cells. Key transcription factors identified in the fibroid specific network included *RUNX1*, *AKT*, *FOXO1A*, and mTOR (Table 2). In summary, these analyses revealed functional links between genes characterizing core biological differences which shape each of the three pathotypes.

Apoptosis genes characterize the good-response network

We performed a separate analysis to identify gene networks related to treatment outcome. To allow a cleaner definition of response signatures we excluded samples classified as *moderate responders* in this phase of the analysis. Synovial gene expression of patients who responded well to csDMARD was associated with a relatively small gene network (Fig. 3A, full network Fig. S6). The most prominent cluster was centered around B cell lymphoma 2 (*BCL2*) with genes linked to PI3K-Akt signaling (Fig. 3A, cluster R1). Edges to this node included other cell death regulating genes (*BAX*, *BAD*, *BAK1*), multiple cathepsins (*CTSB*, *CTSS*, *CTSK*) needed for caspase activation, as well as STAT and mitogen-activated protein (MAP) kinase signaling genes. Additional clusters of genes linked to the good-responder group consisted of alpha and gamma chain laminin genes (cluster R2) and key chemokines and chemokine receptors including *CCL19* and *CXCL13* (cluster R3).

Activation of the SOCS2-STAT2 negative feedback loop is predictive of good response to csDMARD

Among links characterizing the good-responder group, the *SOCS2-STAT2* link is of particular interest. Suppressor of cytokine signaling 2 (*SOCS2*) is one of a family of negative regulators of cytokine receptor signaling that acts on the *JAK/STAT* pathway. Regression analysis revealed differential correlation of *SOCS2* and *STAT2* between responders and non-responders ($p = 0.015$, Fig. 3B, Table S4). Following on from this observation, we fitted a logistic regression model to predict response as a function of the two genes. After evaluation of possible confounding factors, we added *age* as an additive covariate to the linear model (Table S6). The interaction term (the gene ratio between *STAT2* and *SOCS2*) was significant ($p = 0.010$) as observed in a plot of the regression model dichotomizing *STAT2* expression at ± 1 standard deviation (Fig. 3C). ROC curve analysis of the ability of the model to predict response found that the combined model incorporating *SOCS2*, *STAT2*, and *STAT2*/*SOCS2* ratio showed an area under the curve (AUC) value of 0.87 (Fig. 3D). Removal of the *STAT2*/*SOCS2* ratio term from the linear model resulted in a substantial drop in AUC to 0.71, confirming that the gene ratio interaction term strongly improved the predictive ability of the model.

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

Table 2 GO/pathway enrichment analysis on network clusters

Network	Cluster	Enrichment term	Nr. of associated genes	% of associated genes	Adj *p*-value
Lympho-myeloid	LM1	Chemokine signaling pathway[a]	29/31	93.5	6.49e−55
	LM2	Antigen processing and presentation[a]	6/6	100.0	5.78e−14
	LM3	Focal adhesion	31/33	93.9	2.54e−58
	LM4	TNF signaling pathway[a]	20/41	48.8	3.72e−32
	LM5	RIG-I-like receptor signaling pathway[a]	7/8	87.5	1.19e−15
	LM6	Apoptosis[a]	33/65	50.8	2.74e−52
	LM7	MAPK signaling pathway[a]	44/71	62.0	1.4e−59
	LM8	Wnt signaling pathway	13/16	81.2	7.56e−24
	LM9	positive regulation of interleukin-8 production[a]	6/12	50.0	5.92e−09
	LM10	Interleukin-2 family signaling[a]	3/3	100.0	2.46e−07
	LM11	disulfide oxidoreductase activity[a]	4/4	100.0	1.27e−09
Diffuse-myeloid	DM1	Focal adhesion	34/45	75.6	7.1e−57
	DM2	PPAR signaling pathway[a]	22/28	78.6	1.11e−47
	DM3	Dopaminergic synapse[a]	23/30	76.7	8.94e−43
	DM4	EPHA-mediated growth cone collapse[a]	3/3	100.0	3.78e−08
	DM5	Cam-PDE 1 activation[a]	4/4	100.0	1.41e−13
	DM6	Adherens junction[a]	5/6	83.3	1.35e−08
Pauci-immune Fibroid	PF1	Focal adhesion	42/45	93.3	2.06e−79
	PF2	Ras signaling pathway[a]	49/64	76.6	1.09e−79
	PF3	Wnt signaling pathway	23/24	95.8	1.2e−12
	PF4	Cytokine-cytokine receptor interaction[a]	18/18	100.0	4.41e−37
	PF5	VEGFR2 mediated vascular permeability[a]	4/7	57.1	1.07e−06
	PF6	Regulation of RUNX1 Expression and Activity[a]	3/3	100.0	1.59e−07
	PF7	TGF-beta signaling pathway[a]	9/14	64.3	6.25e−17
	PF8	mTOR signaling[a]	6/6	100.0	7.35e−16
	PF9	Adrenergic signaling in cardiomyocytes[a]	9/11	81.8	5.71e−16
	PF10	Vascular smooth muscle contraction[a]	12/19	63.2	2.75e−20
Good Responders	R1	PI3K-Akt signaling pathway[b]	34/64	53.1	3.76e−39
	R2	ECM-receptor interaction[b]	7/7	100.0	2.74e−16
	R3	Chemokine receptors bind chemokines[b]	6/6	100.0	1.07e−15
	R4	MAP2K and MAPK activation[b]	5/7	71.4	1.91e−10
	R5	disulfide oxidoreductase activity[b]	4/4	100.0	1.27e−09
	R6	Triglyceride catabolism[b]	3/3	100.0	3.77e−08
	R7	Cam-PDE 1 activation[b]	4/4	100.0	1.41e−13
Non Responders	NR1	Antigen processing and presentation[b]	6/6	100.0	5.78e−14
	NR2	Chemokine signaling pathway[b]	25/25	100.0	1.6e−49
	NR3	Wnt signaling pathway[b]	23/31	74.2	1.02e−16
	NR4	VEGFR2 mediated vascular permeability[b]	7/18	38.8	3.12e−14
	NR5	Olfactory transduction[b]	32/50	64.0	6.9e−10
	NR6	RIG-I-like receptor signaling pathway[b]	5/6	83.3	4.88e−31
	NR7	Platelet activation, signaling and aggregation[b]	24/39	61.5	1.34e−17

Enrichment terms marked with an asterisk ([a]) are unique across pathotypes, those marked with a dagger ([b]) are unique among good/poor responders

Endothelial activation genes link to differential responsiveness to DMARD therapy

The gene expression network specific to the non-responder group showed similarities to the lympho-myeloid gene network as we obtained cluster NR1 of class I human leukocyte antigen (HLA) genes linked by pathway analysis to antigen presentation (Fig. 4A, full network Fig. S7) and cluster NR2 of leucocyte attracting chemokine genes around nodes *CCR2* and *CXCR5*. The B cell mediator activity of *CXCR5* is known to be

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | LE | TYPESET
MS Code : | CP | DISK

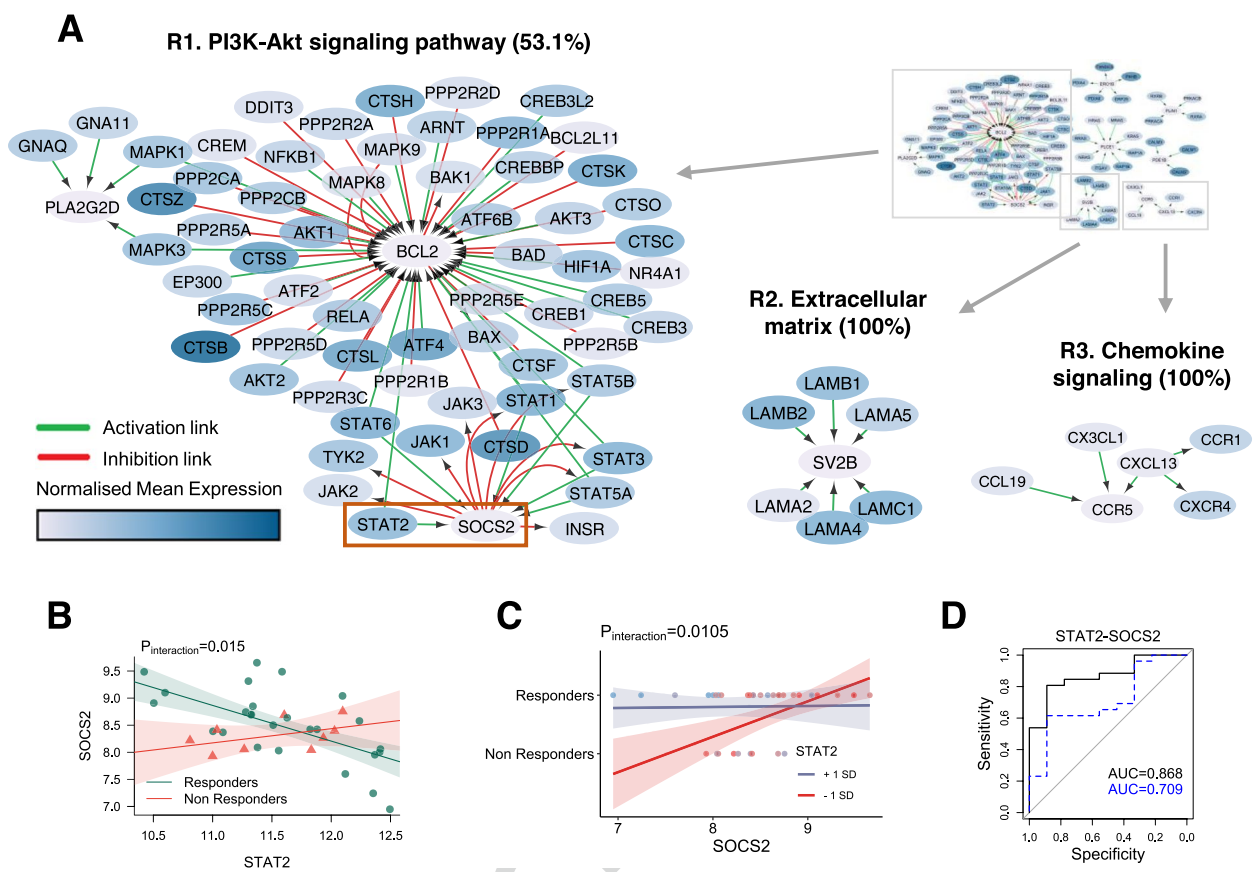

Fig. 3 Apoptosis and SOCS/STAT signaling differentiate responders to methotrexate-based therapy from non-responders. **A** Network of unique active interactions in conventional synthetic DMARD responders. Cluster R1. Cell survival genes part of the PI3K-Akt signaling pathway (53.1% enrichment). Cluster R2. Extracellular matrix receptor genes (100% enrichment). Cluster R3. Chemokines receptors binding chemokines (100% enrichment). **B** Robust linear regression of SOCS2 and STAT2 with interaction term associated with response. *p*-value of the gene to response interacting term is shown. **C** Logistic regression of response as a function of SOCS2 and STAT2. *p*-value of the response to gene interacting term is shown. Expression of *STAT2* is dichotomized at ± 1 standard deviation. **D** Receiver operating characteristic curve analysis of the response prediction ability of the robust linear model incorporating the ratio term (*black line*) or not (*dotted blue line*)

initiated by G-protein family genes and particularly depends upon the availability of $G\alpha_{i2}$ and $G\alpha_{i3}$. *GNAI3*, which encodes $G\alpha_{i3}$, showed differential correlation with *CXCR5* when comparing non-responders to responders (Fig. 4B). The non-responder network also showed a Wnt signaling cluster (NR3) analogous to cluster PF3 in the pauci-immune fibroid network, and an angiogenesis cluster, NR4, centered around *NOS3* (nitric oxide synthase 3, eNOS) with surrounding genes linked by pathway analysis to *VEGFR2*-mediated vascular permeability. Vascular endothelial growth factor (VEGF) can activate eNOS either through Ca^{2+}/calmodulin or by kinase-mediated

(See figure on next page.)

Fig. 4 Gene pair interactions linked to endothelial activation and Akt signaling enhance prediction of response to methotrexate-based therapy. **A** Network of unique active interactions in conventional synthetic DMARD poor responders. Cluster NR1. Antigen processing and presentation cluster (100% enrichment). Cluster NR2. Genes of the chemokine signaling pathway (100% enrichment). Cluster NR3. Cluster associated to Wnt signaling pathway (74.2% enrichment). Cluster NR4. Cluster linked to VEGFR2 mediated vascular permeability (38.8% enrichment). Red boxes highlight predictive gene pairs. **B**–**E**, **H**, **L** Robust linear regression with interaction term associated with response for **B** *GNAI3* and *CXCR5* **C** *NOS3* and *CAMK1*, **D** *NOS3* and *AKT3* **E** *AKT1* and *PPP2R3B*, **H** *NOS3* and *CAMK2D*, **L** *ATP1B1* and *PIK3CD*. *p*-values of the interacting terms are shown. **F**, **I**, **M** Logistic regression of response as a function of **F** *AKT1* and *PPP2R3B*, **I** *NOS3* and *CAMK2D*, **M** *ATP1B1* and *PIK3CD*. *p*-values of the response to gene interacting term are shown. Expression of the second gene is dichotomized at ± 1 standard deviation. **G**, **J**, **N** Receiver operating characteristic (ROC) curve analysis of the of robust linear model ability to predict response using **G** *AKT1* and *PPP2R3B* **J** *NOS3* and *CAMK2D* **N** *ATP1B1* and *PIK3CD*. All plots show a ROC curve for both the model including the gene-gene ratio interaction term (*in black*) and the equivalent model excluding the ratio

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

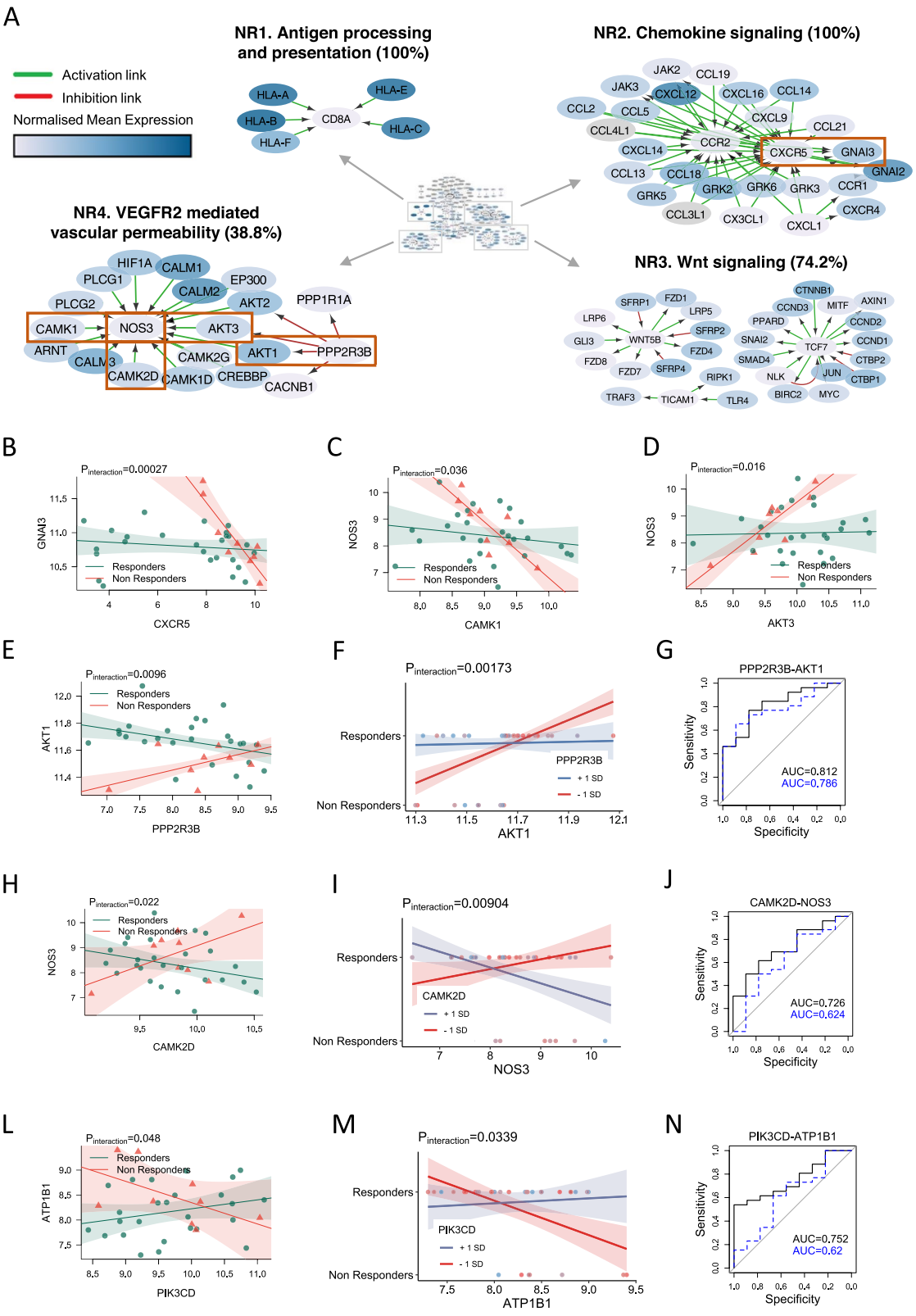

Fig. 4 (See legend on previous page.)

Journal : **BMCTwo 13075** | Dispatch : **16-5-2022** | Pages : **14**

Article No : **2803** | ☐ LE | ☐ TYPESET

MS Code : | ☑ CP | ☑ DISK

phosphorylation [31]. We observed a molecular signature for both processes, with differential correlation between *NOS3* and *CAMK1* (calcium/calmodulin-dependent protein kinase I) or *AKT3* (AKT serine/threonine kinase 3) with evidence of statistical interaction between responders and non-responders ($p=0.036$ and $p=0.016$ respectively, Fig. 4C, D). *AKT1*, a member of the same AKT family that activates *NOS3*, was also found to be correlated with its regulator *PPP2R3B* (protein phosphatase 2 regulatory subunit B"beta). Statistical analysis by robust linear regression showed a significant interaction term between response and gene expression when comparing response categories ($p=0.0096$, Fig. 4E). A logistic model incorporating *AKT1*, *PPP2R3B*, and the ratio between the two genes was highly predictive of response (Fig. 4F), reaching an AUC of 0.81 (Fig. 4G). Another Ca^{2+}/calmodulin-dependent kinase gene, *CAMK2D*, demonstrated differential interaction with *NOS3* between responders and non-responders ($p=0.022$, Fig. 4H). The *NOS3/CAMK2D* ratio was found to be a significant term ($p=0.00904$) in a logistic model for prediction of response to csDMARD (Fig. 4I) with an AUC of 0.73, which fell to 0.62 if the *NOS3*/*CAMK2D* ratio term was excluded (Fig. 4J).

Another noteworthy process involved the class I phosphoinositide 3-kinase (PI3K) gene *PIK3CD* primarily found in leukocytes. Correlation between *PIK3CD* and *ATP1B1* differed significantly between responders and non-responders ($p=0.048$, Fig. 4L). A logistic regression model fitted to predict response outcome showed a significant interaction term ($p=0.0339$) for *PIK3CD*/*ATP1B1* ratio (Fig. 4M). The ability of this model in predicting response outcome was good with an AUC of 0.75 (Fig. 4N), which dropped to 0.62 if the interaction term was excluded.

HLA-DRB1 alleles did not show link to differential responsiveness to DMARD therapy

It is well known that specific HLA alleles are the most important genetic risk factors to develop anti-citrullinated autoantibody-positive RA [32]. As mentioned above, HLA genes were observed in both the lympho-myeloid and the non-responder specific networks. Furthermore, anti-CCP positive patients dominate the lympho-myeloid group (Table 1). On the basis of this observations, we identified the top five most frequent *HLA-DRB1* alleles in our cohort and assessed whether differential recurrence could be observed across pathotypes, CCP status and EULAR response groups. Table S10 shows the distribution of the *HLA-DRB1* alleles across pathotypes indicating no statistical difference across them. Similarly, Tables S11-S12 show no significant association of *HLA-DRB1* alleles either in anti-CCP positive patients or in EULAR non-responders by linear regression analysis (see supplementary methods). To further investigate the possible predictive role of *HLA-DRB1* alleles, we also systematically added additional *HLA-DRB1* allele terms to the predictive models described in the previous paragraphs. Results shown in Tables S13-S16 indicate that the *HLA-DRB1* alleles as additive terms never reached significant *p*-value levels and typically worsened statistical significance of the remaining terms in the linear models.

Predictive models derived from synovial RNA-seq show indication of prediction ability in a small subset of matched blood samples

For comparison, predictive models on synovial RNA-Seq were tested in blood RNA-Seq samples from 54 matched patients. Of the four predictive gene-gene interactions discussed in the previous paragraphs, only *PPP2R3B-AKT1* showed significant *p*-value in the interaction term of the model formula, with subsequent AUC reaching 0.89 (Fig. S8A, B).

PIK3CD-ATP1B1, *STAT2-SOCS2*, and *CAMK2D-NOS3* did not reach significance on *p*-value levels for their interaction term (Fig. S8C, E, G), although *PIK3CD-ATP1B1* and *STAT2-SOCS2* showed reasonable AUC (0.81 and 0.75 respectively, Fig. S8D, F). However, the number of individuals with blood RNA-Seq were small, so the AUC estimate is noisier than for synovial RNA-Seq.

Discussion

Investigating differential gene-gene interactions in specific groups of patients can enhance understanding of functional pathogenic mechanisms that cannot be captured from single gene level analyses such as differential gene expression studies. Using an extensive network of experimentally validated interactions, we characterized gene networks in histological subgroups (pathotypes) as well as responder/non-responder subgroups of patients from the PEAC cohort [6, 7]. We confirmed our previous finding that the pathotypes are delineated by the type of cells infiltrating the tissue, namely macrophages for the diffuse-myeloid pathotype, B cells for the lympho-myeloid pathotype and fibroblasts for the pauci-immune pathotype. However, beyond this, we uncovered critical gene interactions driving processes which clearly differentiate the three pathotypes and may explain underlying mechanisms that differ between pathotypes. Network analysis showed that alterations in the interplay between collagens, laminins and integrins play a central role in differentiating the three pathotypes. This may reflect tissue destructive processes within the joint and extracellular matrix (ECM) remodeling processes leading to

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

differential effects on immune cell tissue infiltration during RA pathogenesis distinguishing the pathotypes.

In lympho-myeloid and the diffuse-myeloid subgroups, the specific gene networks were dominated by TNF and chemokine signaling consistent with macrophage infiltration characterizing the diffuse-myeloid subgroup and B/T cell infiltration characterizing the lympho-myeloid subgroup. In addition to these well-known pathways, we observed cytotoxic T cell genes in the lympho-myeloid network where we observed HLA class I genes around CD8 (Fig. 1B cluster LM2) and pro-inflammatory genes around granzyme B (*GZMB*, cluster LM6), which is a marker for two distinct synovial $CD8^+$ T cell subtypes (SC-T5, SC-T6) recently identified in single-cell RNA-Seq studies [33]. The diffuse-myeloid gene network (Fig. 2A) demonstrated subnetworks centered around PPAR-γ and its control over fatty acid metabolism which fits with the importance of these pathways in regulating M1/M2 tissue macrophage differentiation. Adiponectin (*ADIPOQ*) has received attention for its role in RA pathogenesis [19] and its expression is elevated in early RA patients [34, 35].

In the pauci-immune fibroid pathotype, we found gene networks involving (i) multiple integrin genes which may represent the interaction between fibroblasts and the ECM, (ii) TGF-beta together with SMAD signaling molecules involved in fibroblast differentiation, and (iii) an array of growth factor genes (*FGF*, *PDGF*, *IGF*, *VEGF*) and specific cytokines (IL-17D) (Fig. 2B). Thus, the pauci-immune fibroid pathotype consists of an environment driving fibroblast chemotaxis, proliferation and differentiation [36].

In the second phase of our analysis, we examined networks specific for good-responders to methotrexate-based DMARD therapy in comparison to poor-responders. Multiple chemokines and chemokine receptors were observed in the good-response network, consistent with their importance in immune cell infiltration into inflamed tissues. Humby et al. [7] showed that good-responders had significant reduction in synovial expression of genes associated with lymphoid aggregation, as measured by Nanostring panel, including *CXCL13* and *CCL19* which overlap with the present study's good-response network.

Along with leukocyte recruitment and T cell activation, the lympho-myeloid subgroup also expressed cluster LM6 which contained apoptosis related genes that showed some degree of overlap with a similar cluster (R1) in the good-responders which was centered around *BCL2*. The role of apoptosis in RA is highly debated [37]. One previous study has shown increased caspase activation in inflamed synovial tissue which normalized alongside downmodulation of apoptosis regulators following successful DMARD therapy [38].

Multiple STAT and JAK genes were also observed in the good-responder network (Fig. 3A) consistent with their importance in promoting synovial tissue inflammation and the development and mainstream usage of JAK inhibitors as key therapeutics in RA. When analyzing gene-gene pairs, we observed a statistically significant interaction between *STAT2* and *SOCS2* expression which differentiated responders and non-responders (Fig. 3B, C). Accordingly, the ratio of *STAT2-SOCS2* significantly improved a prediction model of treatment response (Fig. 3D). A previous study looking at SOCS1-3 found increased expression of *SOCS2* in RA peripheral blood T cells and synovial fluid macrophages [39]. SOCS genes are typically suppressors of STAT-mediated cytokine signaling, so it is highly plausible that the ratio between specific STAT and SOCS genes could regulate resolution of inflammation and thus influence response to therapy.

This theme of interactions between pairs of genes known to regulate each other was also observed for other gene pairs including *AKT1* and its regulator *PPP2R3B*. Statistical interaction was found between *PPP2R3B* and *AKT1* and response, and the ratio of *PPP2R3B* to *AKT1* improved the AUC of a predictive model. Similarly, *CAMK2D* and *NOS3* showed statistical interaction with response, and the *NOS3/CAMK2D* ratio improved prediction of response. We found similar statistical interactions between *CAMK1* and *NOS3* and *AKT3* and *NOS3*. These results suggest that altered biological interactions involving these gene pairs differentiates responders from non-responders. The strong involvement of Ca^{2+}/calmodulin-dependent kinases and eNOS in inflammation-induced vascular permeability suggests that vascular permeability may be a novel mechanism which potentially explains therapeutic response vs failure of methotrexate-based DMARD therapy.

In addition, we also observed notable interactions in other parts of the PI3K/AKT/mTOR pathway, including an improved predictive model for the ratio of PI3 kinase *PIK3CD* and the sodium-potassium ATPase *ATP1B1*. Interestingly, reduction in synovial expression of *PIK3CD* has been linked with response to anti-TNF therapy in RA patients [40, 41]. *ATP1B1* has been identified as a biomarker of prognosis and treatment response in different cancer settings [42], which suggests that it may be a globally important predictive biomarker.

Our study has limitations (i) in the number of patients for which synovial and blood RNA-Seq data was available and (ii) in the lack of similar validation cohorts in early RA. However, we aim to validate the observed gene-gene interactions and predictive models, in future cohorts

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

including the R4RA trial [43] and forthcoming STRAP trial [44].

Conclusions

In summary, we identified gene-gene networks specific to histologically defined pathotypes in early RA, revealing new biological mechanisms which underlie the development of each pathotype. We identified specific gene networks which differentiate responders to DMARDs from poor-responders. Further analysis of these networks identified gene pairs whose ratios enhanced models predicting response at 6 months. This approach has significant clinical potential to identify interacting pairs of genes which can be used to stratify patients into responders and non-responders.

Abbreviations

AKT3: AKT serine/threonine kinase 3; AUC: Area under curve; BCL2: B cell lymphoma 2; CAMK1: Calcium/calmodulin-dependent protein kinase I; csDMARD: Conventional synthetic disease-modifying anti-rheumatic drugs; ECM: Extracellular matrix; GO: Gene ontology; HLA: Human leukocyte antigen; IL: Interleukin; ITGAV: Integrin subunit alpha V; LAMA3: Laminin alpha-3; MAP: Mitogen-activated protein; mTOR: Mammalian target of rapamycin; NOS3: Nitric oxide synthase; PI3K: Class I phosphoinositide 3-kinase; PEAC: Pathobiology of early arthritis cohort; PPAR: Peroxisome proliferator-activated receptor; PPP2R3B: Protein phosphatase 2 regulatory subunit B"beta; RA: Rheumatoid arthritis; RIG: Retinoic acid-inducible gene; ROC: Receiver operating characteristic; RNA-Seq: RNA sequencing; SOCS2: Suppressor of cytokine signaling 2; TGF: Transforming growth factor; TNC: Tenascin C; TNF: Tumor necrosis factor; VEGF: Vascular endothelial growth factor.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13075-022-02803-z.

Additional file 1. File including supplementary methods figures, and tables S6-16.

Additional file 2. Supplementary tables S1-5.

Acknowledgements

Not applicable.

Authors' contributions

ES performed bioinformatic analysis. ES and ML wrote the paper receiving input from all authors. SA, AP and AF developed the underlying network methodology called MITHRiL. VL advised on the network theory. CP provided samples and ML supervised the study. The authors read and approved the final manuscript.

Funding

The Pathobiology of Early Arthritis Cohort (PEAC) was supported by funding from the UK Medical Research Council (MRC) (grant number G0800648).

Availability of data and materials

The dataset supporting the conclusions of this article is available in the ArrayExpress repository, https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6141/ . The code used for analysis is publicly available on the online github hosting service at https://github.com/elisabettasciacca/DEGGs . The code is currently being developed into an R package and will be submitted to the Bioconductor repository shortly.

Declarations

Ethics approval and consent to participate

The study received ethical approval from the UK Health Research Authority (REC 05/Q0703/198, National Research Ethics Service Committee London – Dulwich). All patients gave written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [2]Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [3]Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy. [4]School of Mathematical Sciences, Queen Mary University of London, London, UK. [5]Dipartimento di Fisica ed Astronomia, Università di Catania and INFN, I-95123 Catania, Italy. [6]Digital Environment Research Institute, Queen Mary University of London, London, UK.

Received: 19 January 2022 Accepted: 4 May 2022

References

1. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
2. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, et al. MiRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 2011;39(suppl_1):D163–9.
3. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009;37(suppl_1):D105–10.
4. Tong Z, Cui Q, Wang J, Zhou Y. TransmiR v2.0: an updated transcription factor-microRNA regulation database. Nucleic Acids Res. 2019;47(D1):D253–8.
5. Alaimo S, Giugno R, Acunzo M, Veneziano D, Ferro A, Pulvirenti A. Post-transcriptional knowledge in pathway analysis increases the accuracy of phenotypes classification. Oncotarget. 2016;7(34):54572–82.
6. Lewis MJ, Barnes MR, Blighe K, Goldmann K, Rana S, Hackney JA, et al. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. Cell Rep. 2019;28(9):2455–2470.e5.
7. Humby F, Lewis M, Ramamoorthi N, Hackney JA, Barnes MR, Bombardieri M, et al. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. Ann Rheum Dis. 2019;78(6):761–72.
8. van der Heijden JW, Dijkmans BAC, Scheper RJ, Jansen G. Drug Insight: Resistance to methotrexate and other disease-modifying antirheumatic drugs - From bench to bedside. Nat Clin Pract Rheumatol. 2007;3(1):26–37.
9. Mittal N, Mittal R, Sharma A, Jose V, Wanchu A, Singh S. Treatment failure with disease-modifying antirheumatic drugs in rheumatoid arthritis patients. Singap Med J. 2012;53(8):532.
10. Badot V, Galant C, Toukap AN, Theate I, Maudoux AL, Van den Eynde BJ, et al. Gene expression profiling in the synovium identifies a predictive signature of absence of response to adalimumab therapy in rheumatoid arthritis. Arthritis Res Ther. 2009;11(2):1–13.
11. Lindberg J, Wijbrandts CA, van Baarsen LG, Nader G, Klareskog L, Catrina A, et al. The gene expression profile in the synovium as a predictor of the clinical response to infliximab treatment in rheumatoid arthritis. PLoS One. 2010;5(6):e11310.
12. De Groof A, Ducreux J, Humby F, Nzeusseu Toukap A, Badot V, Pitzalis C, et al. Higher expression of TNFαα-induced genes in the synovium of patients with early rheumatoid arthritis correlates with disease activity,

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

and predicts absence of response to first line therapy. Arthritis Res Ther. 2016;18(1):1–12.
13. Mandelin AM, Homan PJ, Shaffer AM, Cuda CM, Dominguez ST, Bacalao E, et al. Transcriptional profiling of synovial macrophages using minimally invasive ultrasound-guided synovial biopsies in rheumatoid arthritis. Arthritis Rheum. 2018;70(6):841–54.
14. Cherlin S, Lewis MJ, Plant D, Nair N, Goldmann K, Tzanis E, et al. Investigation of genetically regulated gene expression and response to treatment in rheumatoid arthritis highlights an association between IL18RAP expression and treatment response. Ann Rheum Dis. 2020.
15. Nerviani A, Pitzalis C. Role of chemokines in ectopic lymphoid structures formation in autoimmunity and cancer. J Leukoc Biol. 2018;104(2):333–41.
16. Firestein GS, McInnes IB. Immunopathogenesis of rheumatoid arthritis. Immunity. 2017;46(2):183–96.
17. Hasegawa M, Yoshida T, Sudo A. Role of tenascin-C in articular cartilage. Mod Rheumatol. 2018;28(2):215–20.
18. Yoshida T, Akatsuka T, Imanaka-Yoshida K. Tenascin-C and integrins in cancer. Cell Adhes Migr. 2015;9(1–2):96–104.
19. Fatel EC d S, Rosa FT, ANC S, Dichi I. Adipokines in rheumatoid arthritis. Adv Rheumatol (London, England). 2018;58.
20. Arias de la Rosa I, Escudero-Contreras A, Rodríguez-Cuenca S, Ruiz-Ponce M, Jiménez-Gómez Y, Ruiz-Limón P, et al. Defective glucose and lipid metabolism in rheumatoid arthritis is determined by chronic inflammation in metabolic tissues. J Intern Med. 2018;248(1):61–77.
21. Yarwood A, Martin P, Bowes J, Lunt M, Worthington J, Barton A, et al. Enrichment of vitamin D response elements in RA-associated loci supports a role for vitamin D in the pathogenesis of RA. Genes Immun. 2013;14(5):325–9.
22. Chawla A. Control of macrophage activation and function by PPARs. Circ Res. 2010;106(10):1559–69.
23. Bengtsson T, Aszodi A, Nicolae C, Hunziker EB, Lundgren-Åkerlund E, Fässler R. Loss of α10β1 integrin expression leads to moderate dysfunction of growth plate chondrocytes. J Cell Sci. 2005;118(2):929–36.
24. Tulla M, Pentikainen OT, Viitasalo T, Kapyla J, Impola U, Nykvist P, et al. Selective binding of collagen subtypes by integrin alpha 1I, alpha 2I, and alpha 10I domains. J Biol Chem. 2001;276(51):48206–12.
25. Sommarin Y, Larsson T, Heinegård D. Chondrocyte-matrix interactions. Attachment to proteins isolated from cartilage. Exp Cell Res. 1989;184(1):181–92.
26. Ahnert P, Kirsten H. Association of ITGAV supports a role of angiogenesis in rheumatoid arthritis. Arthritis Res Ther. 2007;1–2.
27. Jacq L, Garnier S, Dieudé P, Michou L, Pierlot C, Migliorini P, et al. The ITGAV rs3738919-C allele is associated with rheumatoid arthritis in the European Caucasian population: A family-based study. Arthritis Res Ther. 2007;9(4):1–7.
28. Trenkmann M, Brock M, Gay RE, Kolling C, Speich R, Michel BA, et al. Expression and function of EZH2 in synovial fibroblasts: epigenetic repression of the Wnt inhibitor SFRP1 in rheumatoid arthritis. Ann Rheum Dis. 2011;70(8):1482–8.
29. Imai K, Morikawa M, D'Armiento J, Matsumoto H, Komiya K, Okada Y. Differential expression of WNTs and FRPs in the synovium of rheumatoid arthritis and osteoarthritis. Biochem Biophys Res Commun. 2006;345(4):1615–20.
30. Stamp LK, Easson A, Lehnigk U, Highton J, Hessian PA. Different T cell subsets in the nodule and synovial membrane: absence of interleukin-17A in rheumatoid nodules. Arthritis Rheum. 2008;58(6):1601–8.
31. Spiller F, Oliveira Formiga R, da Silva F, Coimbra J, Alves-Filho JC, Cunha TM, et al. Targeting nitric oxide as a key modulator of sepsis, arthritis and pain. Nitric Oxide Biol Chem. 2019;89:32–40.
32. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012.
33. Zhang F, Wei K, Slowikowski K, Fonseka CY, Rao DA, Kelly S, et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. Nat Immunol. 2019;20(7):928–42.
34. Yoshino T, Kusunoki N, Tanaka N, Kaneko K, Kusunoki Y, Endo H, et al. Elevated serum levels of resistin, leptin, and adiponectin are associated with c-reactive protein and also other clinical conditions in rheumatoid arthritis. Intern Med. 2011;50(4):269–75.
35. Popa C, Netea MG, De Graaf J, Van Den Hoogen FHJ, Radstake TRDJ, Toenhake-Dijkstra H, et al. Circulating leptin and adiponectin concentrations during tumor necrosis factor blockade in patients with active rheumatoid arthritis. J Rheumatol. 2009;36(4):724–30.
36. Chaudhary NI, Roth GJ, Hilberg F, Müller-Quernheim J, Prasse A, Zissel G, et al. Inhibition of PDGF, VEGF and FGF signalling attenuates fibrosis. Eur Respir J. 2007;29(5):976–85.
37. Liu H, Pope RM. The role of apoptosis in rheumatoid arthritis. Curr Opin Pharmacol. 2003;3(3):317–22.
38. Smith MD, Weedon H, Papangelis V, Walker J, Roberts-Thomson PJ, Ahern MJ. Apoptosis in the rheumatoid arthritis synovial membrane: modulation by disease-modifying anti-rheumatic drug treatment. Rheumatology (Oxford). 2010;49(5):862–75.
39. Isomäki P, Alanärä T, Isohanni P, Lagerstedt A, Korpela M, Moilanen T, et al. The expression of SOCS is altered in rheumatoid arthritis. Rheumatology. 2007;46(10):1538–46.
40. Marsal S, Julià A, Ávila G, Celis R, Sanmarti R, Ramirez J, et al. 240. Pik3Cd Overexpression in the synovial membrane of rheumatoid arthritis patients is associated with response to anti-TNF therapy. Rheumatology. 2014;53(suppl_1):i149–50.
41. Whitehead MA, Bombardieri M, Pitzalis C, Vanhaesebroeck B. Isoform-selective induction of human p110δ PI3K expression by TNFα: identification of a new and inducible PIK3CD promoter. Biochem J. 2012;443(2):857–67.
42. Shi JL, Fu L, Ang Q, Wang GJ, Zhu J, Wang WD. Overexpression of ATP1B1 predicts an adverse prognosis in cytogenetically normal acute myeloid leukemia. Oncotarget. 2016;7(3):2585.
43. Humby F, Durez P, Buch MH, Lewis MJ, Rizvi H, Rivellese F, et al. Rituximab versus tocilizumab in anti-TNF inadequate responder patients with rheumatoid arthritis (R4RA): 16-week outcomes of a stratified, biopsy-driven, multicentre, open-label, phase 4 randomised controlled trial. Lancet. 2021;397(10271):305–17.
44. Experimental Medicine & Rheumatology Department QMUL. STRAP (Stratification of Biologic Therapies for RA by Pathobiology). 2021. Available from: http://www.matura-mrc.whri.qmul.ac.uk/

Publisher's Note

Journal : BMCTwo 13075 | Dispatch : 16-5-2022 | Pages : 14
Article No : 2803 | ☐ LE | ☐ TYPESET
MS Code : | ☑ CP | ☑ DISK

Journal:	**13075**
Article:	**2803**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Please check if the affiliation/s is/are presented correctly.	
AQ2	Please check if the section headings are assigned to appropriate levels.	
AQ3	Figures 1,2,3 and 4 contain poor quality and small text inside the artwork. Please do not re-use the file that we have rejected or attempt to increase its resolution and re-save. It is originally poor, therefore, increasing the resolution will not solve the quality problem. We suggest that you provide us the original format. We prefer replacement figures containing vector/editable objects rather than embedded images. Preferred file formats are eps, ai, tiff and pdf.	
AQ4	Please check if the figure captions are captured and presented correctly.	
AQ5	As per standard instruction, the statement "The authors read and approved the final manuscript." is required in the "Authors' contributions" section. Please note that this was inserted in the said section. Please check if correct.	
AQ6	Citation details for reference/s [14, 32] is/are incomplete. Please supply the "volume ID and page number" of this/these reference/s. Otherwise, kindly advise us on how to proceed.	
AQ7	Citation details for reference/s [19] is/are incomplete. Please supply the "page number" of this/these reference/s. Otherwise, kindly advise us on how to proceed.	
AQ8	Citation details for reference/s [26] is/are incomplete. Please supply the "volume ID" of this/these reference/s. Otherwise, kindly advise us on how to proceed.	