TRANSCRIPTIONAL AND HISTOLOGICAL CHARACTERISATION OF THE SYNOVIUM IN EARLY INFLAMMATORY ARTHRITIS

by

HAYLEY LOUISE CARR

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Rheumatology Research Group
Institute of Inflammation and Ageing
School of Medical and Dental Sciences
University of Birmingham
November 2021

ABSTRACT

There is a large amount of heterogeneity in early inflammatory arthritis, across disease presentation, clinical manifestations, and response to treatment. This study aimed to capture the heterogeneity seen in tissue morphology and gene expression across BEACON, the Birmingham Early Arthritis Cohort, which includes patients with short duration rheumatoid arthritis (RA), longer duration RA, other persistent inflammatory arthritis, and resolving disease.

A novel H&E scoring system was developed that allows for stratification based on the density and aggregation of the lymphoid infiltrate, resulting in a summary pathotype. This correlated with measures of local and systemic inflammation and may associate with treatment response in RA.

Whole tissue RNA sequencing was used to assess gene expression across the early inflammatory arthritis cohort and highlighted major sources of variation, including the presence of prominent immune cell, adipocyte, and fibroblast signatures. Signatures associated with resolution and response to treatment in RA were identified, with response to treatment being predominantly associated with immune and inflammatory processes. Resolution was associated with reduced activation of the adaptive immune response and increased lipid synthesis when compared to short duration RA, although the implications of these signatures requires further validation.

Overall, this study extended the knowledge of heterogeneity in early inflammatory arthritis and highlights the potential utility of patient stratification for predicting response to treatment. Furthermore, some novel mechanisms of

resolution were identified that could lead to the identification of new treatment targets for RA that drive pro-resolution pathways and inhibit pathways associated with persistence.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr Andrew Filer and Prof Dagmar Scheel-Toellner for their support and guidance throughout my PhD, and in particular for the many hours spent looking at histology. I would also like to thank Prof Jean-Baptiste Cazier and Prof Karim Raza for their knowledge and insight, supervision meetings with them helped give perspective and drive the project forward.

I am very grateful to Dr Csilla Varnai for all her bioinformatics support, whose patience and encouragement made the computational analyses enjoyable. I am also thankful to Dr Jason Turner for his support with bioinformatics, and for answering my many questions in this area. I would also like to thank the current and past members of the RRG, there are too many people that have contributed to list but I give particular thanks to Dr Jennifer Marshall, Dr Triin Major, Miss Holly Adams, and Dr Elizabeth Clay for their histological and technical support and to Dr Valentina Pucino for helping with the metabolomics.

I would like to thank the collaborators on this project, including Prof Ellen Gravallese, Prof Brendan Boyce, Prof Edward DiCarlo, and Prof Jen Anolik from the AMP consortium and Prof Clare Verrill and Dr Nasullah Khalid Alham from the University of Oxford.

I would also like to thank my family. My parents and sister gave me love and support throughout my PhD and my partner, Isaac, helped me stay (somewhat) sane during the pandemic.

Thank you to the Medical Research Council for funding this project, to the IMPACT DTP for providing support and training, and to the patients who contributed samples to this project.

STATEMENT OF CONTRIBUTIONS

This project would not have been possible to carry out alone and was only made possible by collaboration, therefore the contributions made by others are detailed below, as well as throughout the document.

Sample collection and preparation: biopsies were carried out by Dr Andrew Filer, with initial sample preparation for storage being predominantly carried out by Miss Holly Adams before the start of this project. Cutting and H&E staining of slides was predominantly carried out in the University Hospitals Birmingham pathology department, however some was also done by Dr Jennifer Marshall, Dr Triin Major, or myself. Scanning of H&E stained slides was predominantly done by Dr Jennifer Marshall or Dr Triin Major, with some also being done by Dr Emily Taylor and myself.

Histology scoring: the scoring system was devised by Dr Andrew Filer, Prof Dagmar Scheel-Toellner, and myself, with the internal reliability testing also being done by the same people. External reliability testing was undertaken by Prof Dagmar Scheel-Toellner, Prof Ellen Gravallese, Prof Brendan Boyce, and Prof Edward DiCarlo. All analyses were carried out by myself.

AI-assisted scoring: The initial proof-of-concept APP development for lining layer and aggregate detection was done by myself, with these being taken over by Dr Nasullah Khalid Alham for further development and validation. The lymphocyte and plasma cell APPs were developed entirely by Dr Nasullah Khalid Alham. Annotations for training after proof-of-concept and assessment for validation was

undertaken by Dr Andrew Filer, Prof Dagmar Scheel-Toellner, and Prof Clare Verrill.

RNA sequencing and analysis: RNA extraction was carried out by myself and QC testing, library preparation, and sequencing were done by Genomics Birmingham. Bioinformatic analysis of the RNA sequencing dataset was carried out by myself, under the supervision of Dr Csilla Varnai and with help from Dr Jason Turner. Some scripts were adapted from those developed by Dr Csilla Varnai, particularly in early analyses (initial data preparation and alignment), and the script for GO biological process analysis was adapted from one developed by Dr Dina Abdelmottaleb. Gene lists used for exploring metabolism in chapter 6 were curated by Dr Valentina Pucino.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACPA	Anti-citrullinated protein antibodies
ACR	American College of Rheumatology
AIDA	Annotation of Image Data by Assignments
AMP	Accelerating medicines partnership
APP	Analysis Protocol Packages
AS	Ankylosing spondylitis
bDMARD	Biological disease modifying anti-rheumatic drug
BEACON	Birmingham Early Arthritis Cohort
CaStLeS	Compute and Storage for Life Sciences
cDMARD	Conventional disease modifying anti-rheumatic drug
CIA	Collagen-induced arthritis
CRP	C-Reactive Protein
DAS28	Disease Activity Score 28
DC	Dendritic cell
DEG	Differentially expressed gene
DMARD	Disease modifying anti-rheumatic drug
ESR	Erythrocyte Sedimentation Rate
EULAR	European League against Rheumatism
FFPE	Formalin fixed, paraffin embedded
GC	Glucocorticoid
GO	Gene ontology
GWAS	Genome-wide association study
H&E	Haematoxylin and Eosin
HLA	Human leukocyte antigen
IAN	Immuno-associated nucleotide
ICC	Intra-class correlation coefficient
iNOS	Inducible nitric oxide synthase
LDA	Low disease activity
ldRA	Long duration arthritis
LFC	Log-fold change
LL	Lining layer
lncRNA	Long non-coding RNA
MCTD	Mixed connective tissue disease
MHC	Major histocompatibility complex
MIBI-TOF	Multiplexed ion beam imaging by time of flight
misc RNA	Miscellaneous RNA
MMP	Matrix metalloproteinase
mRNA	Messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
NonRA	Non-rheumatoid arthritis inflammatory arthritis
Norm	Normal

NSAID	Non-steroidal anti-inflammatory drug
OA	Osteoarthritis
OCT	Optimal cutting temperature
p	p-value
padj	Adjusted p-value
PBMC	Peripheral blood mononuclear cells
PC	Principal component
PCA	Principal component analysis
PsA	Psoriatic arthritis
QC	Quality control
RA	Rheumatoid Arthritis
ReA	Reactive arthritis
Res	Resolving arthritis
RF	Rheumatoid factor
RIN	Ribonucleic acid integrity number
RNA	Ribonucleic Acid
ROS	Reactive oxygen species
scMEP	Single-cell metabolic regulome profiling
sdRA	Short duration rheumatoid arthritis
SLE	Systemic lupus erythematosus
SpA	Spondyloarthritis
TAG	Triacylglycerol
Treg	Regulatory T cell
TS	Thick section
UA	Undifferentiated arthritis
UMAP	Uniform manifold approximation and projection
US GS	Ultrasound greyscale
US PD	Ultrasound power Doppler
VAS	Visual analogue scale
VST	Variance stabilisation transformation
WGCNA	Weighted gene correlation network analysis
WT	Whole tissue

1 INTRODUCTION

1.1 INFLAMMATORY ARTHRITIS

Inflammatory arthritis is a collection of conditions that cause inflammation of the joint. Rheumatoid arthritis (RA) is the most common inflammatory arthropathy, affecting 1% of the population and is around three times more common in women (Gorman and Cope, 2008). Other inflammatory arthritides include spondyloarthritis (SpA), which is a group of disorders that includes psoriatic arthritis (PsA), reactive arthritis (ReA), and ankylosing spondylitis (AS). Inflammatory arthritis may also be a presenting feature of connective tissue diseases, such as mixed connective tissue disease (MCTD) and systemic lupus erythematosus (SLE), and synovitis may be caused by infections such as parvovirus. Additionally, some patients present with undifferentiated arthritis (UA), meaning they do not meet classification criteria for any specific disease (Hazes and Luime, 2011, Stockman et al., 2006, van Aken et al., 2003).

RA is a chronic inflammatory disease primarily affecting the smaller joints of the hands and feet. Risk for RA is built up of both genetic and environmental factors, which will be discussed in more detail later in the chapter. If the disease is not controlled with appropriate treatment, damage to the joints occurs in most patients. This leads to loss of function, pain, and disability, as well as increased socioeconomic cost (Smolen et al., 2016a). Autoantibodies are present in 50-70% of patients, with anti-citrullinated protein antibodies (ACPA) and rheumatoid factor (RF) being the most common (Smolen et al., 2016a).

RA is associated with a number of comorbidities which may have a substantial impact on patients' quality of life and reduce life expectancy, as well as increasing economic burden and making treatment more challenging (Gabriel and Michaud, 2009). The most significant causes of mortality in this respect are cardiovascular disease, infection, and malignancy (Dougados et al., 2014). However, depression is also a significant problem among RA patients, as with most chronic diseases (Gabriel and Michaud, 2009). Some comorbid conditions are associated with RA treatment, as well as the disease itself. For example, the susceptibility to infection may be due to immunosuppressive treatments or modulation of the immune system in RA (Listing et al., 2013).

As SpA makes up the second largest proportion of patients after RA within the cohort for this study, this will generally be discussed in this introduction in preference to other inflammatory arthropathies. AS predominantly affects the spine, rather than peripheral joints, so is typically less challenging to differentiate from RA and other peripheral inflammatory arthropathies in early arthritis cohorts (Ebrahimiadib et al., 2021). PsA is the most common SpA in the cohort and ReA makes up a proportion of the patients in the cohort with resolving disease, so these will be discussed in more detail. PsA has a number of manifestation alongside synovitis, including skin and nail disease, and cardiovascular disease is a notable comorbidity (Veale and Fearon, 2018). Unlike RA, PsA does not have a higher prevalence in women, instead affecting both sexes equally (Ocampo and Gladman, 2019).

ReA typically affects younger populations, around 20-40 years old, and similarly to PsA affects both sexes equally (Selmi and Gershwin, 2014). It is initially caused by infection, typically of the genitourinary or gastrointestinal tract, but

arthritis often does not present until weeks after, meaning tests for infectious agents can be negative at the point of arthritis development. Furthermore, not all individuals are susceptible to ReA, with less than 15% of infections resulting in arthritis, and the genetic allele HLA-B27 increasing risk (Selmi and Gershwin, 2014).

Another group of patients falls into a UA group, typically making up around 30% of early inflammatory arthritis cohorts (Hazes and Luime, 2011). Patients in this group do not have any known cause of their inflammatory arthritis, failing to meet classification criteria for any particular condition. An estimated 20-60% of these patients spontaneously resolve, however others later fulfil classification criteria for another inflammatory arthropathy (Hazes and Luime, 2011). One study found that 24% of UA patients are classified as having RA within the first year, although this figure can vary greatly between studies, with estimates between 13% and 54%, depending on the cohort and classification criteria used (Krabben et al., 2012, Hazes and Luime, 2011). This group presents a challenge clinically, as uncertainty about future prognosis makes optimal treatment unclear. It would therefore be beneficial to be able to classify those who will later fulfil criteria for RA, or other persistent inflammatory arthritides, earlier in the disease course.

1.1.1 Classification criteria

1.1.1.1 Rheumatoid arthritis

The most commonly used classification criteria for RA currently are the 2010 ACR/EULAR criteria (Table 1.1), which were designed to attempt to identify early RA. The 1987 ACR criteria (Table 1.2) that they superseded, prevented RA from

being classified prior to 6 weeks of symptoms, and heavily favoured classification of later stage, established RA, with one study suggesting that only 50% of RA patients satisfy the criteria one year after symptom onset (Cader et al., 2011, Heidari, 2011). The 2010 ACR/EULAR criteria improved upon this by shifting the predominant weighting from the duration of symptoms to joint involvement and seropositivity, however over half of RA patients may still fail to be classified correctly at first visit (Cader et al., 2011). In addition, the 2010 ACR/EULAR criteria result in more false positive classifications and, if used alone for treatment decisions, would result in increased unnecessary treatment of patients with self-limiting disease (Cader et al., 2011).

	Score
Target population: patients who 1. Have at least one joint with definite clinical synovitis (swelling) 2. With the synovitis not better explained by another disease	
Classification criteria for RA (score-based algorithm: add score of categories A–D a score of ≥6/10 is needed for classification of a patient as having definite RA)	
A. Joint involvement	
1 large joint	0
2−10 large joints	1
1−3 small joints (with or without involvement of large joints)	2
4−10 small joints (with or without involvement of large joints)	3
>10 joints (at least one small joint)††	5
B. Serology (at least 1 test result is needed for classification)	
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high-positive ACPA	3
C. Acute-phase reactants (at least one test result is needed for classification)	
Normal CRP and normal ESR 0	0
Abnormal CRP or normal ESR 1	1
D. Duration of symptoms	
<6 weeks	0
≥6 weeks	1
†† At least one of the involved joints must be a small joint; the other joints can include any combination of large and additional small joints.	

Table 1.1: 2010 ACR/EULAR classification criteria (Aletaha et al., 2010).

Criterion	Definition
1. Morning stiffness	Morning stiffness in and around the joints, lasting at least 1 hour before maximal improvement
2. Arthritis of 3 or more joint areas	At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee, ankle, and MTP joints
3. Arthritis of hand joints	At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint
4. Symmetric arthritis	Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPS, MCPs, or MTPs is acceptable without absolute symmetry)
5. Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxtaarticular regions, observed by a physician
6. Serum rheumatoid factor	Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive in 4% of normal control subjects
7. Radiographic changes	Radiographic changes typical of rheumatoid arthritis on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify)

For classification purposes, a patient shall be said to have rheumatoid arthritis if he/she has satisfied at least 4 of these 7 criteria. Criteria 1 through 4 must have been present for at least 6 weeks. Patients with 2 clinical diagnoses are not excluded.

Table 1.2: 1987 ACR classification criteria (Arnett et al., 1988b).

Prediction rules have also been developed, designed to predict RA in patients who do not yet fulfil criteria, with one such example being the Leiden Score (Table 1.3). This is based on 9 variables: age, sex, localisation of symptoms, morning stiffness, tender joint count, CRP, RF, and ACPA and was developed in an UA cohort. These variables were weighted and together used to produce a score out of 14 (van der Helm-van Mil et al., 2007). In the test cohort, 84% of the patients with high scores (≥8) progressed to RA and 91% of patients with scores ≤6 did not develop RA. However, scores between these values, accounting for 25% of patients, are unable to make an accurate prediction (van

der Helm-van Mil et al., 2007). For this reason it is unlikely that the Leiden score will ever be widely utilised in clinical settings.

Variable	Points
Sex	1
Age	0.02/year
Localization in small joints hand/feet	0.5
Symmetric localization	0.5
Localization in upper extremities	1
Localization in both upper and lower extremities	1.5
Morning stiffness score on 100-mm VAS	
0-25	-
26-50	1
51-90	1
>90	2
Number of tender joints	
0-3	-
4-10	0.5
>10	1
Number of swollen joints	
0-3	-
4-10	0.5
>10	1
CRP level, mg/litre	
0-4	-
5-50	0.5
>50	1.5
RF positivity	1
Anti-CCP positivity	2

Table 1.3: Leiden prediction score (van der Helm-van Mil et al., 2007).

Even once a classification has been made, there is still a need for new stratification methods, as RA is a heterogeneous disease. Patients show varying symptoms, rates of disease progression, and responses to treatment. A number of studies have used synovial tissue to stratify patients into subtypes, or pathotypes, with each using slightly different methods and classification systems (Pitzalis et al., 2013, Orange et al., 2018, van der Pouw Kraan et al., 2003, Humby et al., 2019a). Some of these studies have shown some initial success, in

particular in predicting prognosis and response to treatment (Dennis et al., 2014, van Baarsen et al., 2010, Humby et al., 2019a). This will be discussed in more detail later in this chapter.

1.1.1.2 Other inflammatory arthritis

Early diagnosis has also been associated with better outcomes in PsA and there are similar challenges around diagnosis as well, with an estimated 50% of cases being missed in some cohorts (Coates and Helliwell, 2017). There are a number of clinical features that differ between RA and PsA that can aid in distinguishing between them, with skin and nail disease being associated with PsA (present in 80% and 60% of patients, respectively) but being absent in RA and PsA predominantly being seronegative (Coates and Helliwell, 2017). Additionally, the number and pattern of joint involvement tends to differ between these two diseases (Coates and Helliwell, 2017).

To meet the CASPAR (ClASsification criteria for Psoriatic ARthritis) criteria, a patient must have inflammatory articular disease (joint, spine or entheseal) with ≥3 points from the following 5 categories:

1. Evidence of current psoriasis, a personal history of psoriasis, or a family history of psoriasis.
    a. Current psoriasis is defined as psoriatic skin or scalp disease present today as judged by a rheumatologist or dermatologist.†
    b. A personal history of psoriasis is defined as a history of psoriasis that may be obtained from a patient, family physician, dermatologist, rheumatologist, or other qualified health care provider.
    c. A family history of psoriasis is defined as a history of psoriasis in a first- or second-degree relative according to patient report.
2. Typical psoriatic nail dystrophy including onycholysis, pitting, and hyperkeratosis observed on current physical examination.
3. A negative test result for the presence of rheumatoid factor by any method except latex but preferably by enzyme-linked immunosorbent assay or nephelometry, according to the local laboratory reference range.
4. Either current dactylitis, defined as swelling of an entire digit, or a history of dactylitis recorded by a rheumatologist.
5. Radiographic evidence of juxtaarticular new bone formation, appearing as ill-defined ossification near joint margins (but excluding osteophyte formation) on plain radiographs of the hand or foot.

† Current psoriasis is assigned a score of 2; all other features are assigned a score of 1.

Table 1.4: CASPAR criteria for the classification of PsA. (Taylor et al., 2006)

There are also classification criteria for PsA, with a commonly used one being the Classification Criteria for Psoriatic Arthritis (CASPAR) (Table 1.4). The CASPAR criteria were developed using a number of other inflammatory arthritides as controls, including RA, AS, UA, and connective tissue disorders, and allow for the classification of PsA with high specificity by looking at the presence of psoriasis, nail disease, dactylitis, and new bone formation in the absence of RF (Taylor et al., 2006).

Although there are no formal classification criteria for ReA, there are ACR guidelines that aim to define ReA, which are described in Table 1.5. This results

in a 'definite' or 'probable' diagnosis of ReA, with both definitions requiring evidence of a prior infection (Selmi and Gershwin, 2014).

Major criteria	1. Arthritis with 2 of 3 of the following findings: I. Asymmetric II. Mono or oligoarthritis III. Lower limb involvement 2. Preceding symptomatic infection with 1 or 2 of the following findings: I. Enteritis (defined as diarrhoea for at least 1 day, and 3 days to 6 weeks before the onset of arthritis) II. Urethritis (dysuria or discharge for at least 1 day, 3 days to 6 weeks before the onset of arthritis)
Minor criteria	At least one of the following: 1. Evidence of triggering infection: I. Positive urine ligase reaction or urethral/cervical swab for *Chlamydia trachomatis* II. Positive stool culture for enteric pathogens associated with reactive arthritis 2. Evidence of persistent synovial infection (positive immunohistology or PCR for *Chlamydia*)

A "definite" diagnosis of reactive arthritis is based on the fulfilment of both major criteria and a relevant minor criterion, while a "probable" diagnosis is characterized by both major criteria but no relevant minor criterion or one major criterion and one or more of the minor criteria. The identification of the trigger infection is also required.

Table 1.5: Guidelines for the definition of reactive arthritis. (Selmi and Gershwin, 2014)

1.1.2 Treatment

1.1.2.1 Rheumatoid arthritis

Earlier treatment of RA results in better long-term outcomes. It has been shown that treatment within 3 months of symptom onset results in reduced erosions and increased remission rates (Gremese et al., 2013, Bosello et al., 2011). This highlights the need for early diagnosis but also raises concerns about the lack of specific diagnostic tests.

Methotrexate, a conventional disease modifying anti-rheumatic drug (cDMARD), with initial glucocorticoids (GC) is often considered the first treatment option for patients diagnosed with RA, however this approach is unsuccessful in around 25% of patients (Schneider and Kruger, 2013, Bluett et al., 2018). Treatment can also be initiated with, or escalated to, a combination of cDMARDs before a biological DMARD (bDMARD or biologic) is considered (Schneider and Kruger, 2013). However, this approach is currently losing support as biologic and novel targeted synthetic DMARDs, such as Jak/Stat inhibitors, become more accessible. Treatment options are discussed by, and decided between, patients themselves and their rheumatologist after consideration of the individual patient's risk profile, although currently there are challenges around the lack of stratification using biomarkers for treatment response (Schneider and Kruger, 2013, Smolen et al., 2016b).

Biologic treatments for RA target specific inflammatory processes, for example anti-TNF treatments (including infliximab, etanercept, and adalimumab) target the pro-inflammatory cytokine TNF-α, which is elevated in RA and can increase the expression of other pro-inflammatory cytokines (Moelants et al., 2013, Radner and Aletaha, 2015). Abatacept inhibits T cell activation via the blockade of costimulatory signalling, while rituximab reduces the number of B cells by targeting CD20 (Abbasi et al., 2019). However, despite these differing mechanisms of action currently there are no biomarkers used clinically that predict response to a given bDMARD and an estimated 20% of patients fail to respond to two dDMARDs (Buch, 2018).

Treating to target is a concept that has been widely discussed in the literature, which is the concept that treatment should be altered regularly until a disease

state target has been achieved (Schneider and Kruger, 2013, Bluett et al., 2018, Versteeg et al., 2018, Bergstra and Allaart, 2018, Smolen et al., 2016b). What that target should be has also been widely discussed, with low disease activity (LDA), clinical remission, and radiological remission being among those suggested (Bergstra and Allaart, 2018). Smolen et al. (2016b) suggested that clinical remission is the best target to aim for, with LDA being a secondary choice when this is not possible, however Bergstra and Allaart (2018) favour LDA as a target, as they suggested that aiming for remission causes increased treatment burden, lower adherence, and no functional improvement compared to LDA (Bergstra and Allaart, 2018). Both of these studies highlight the need for direct experimental comparisons to find the best treatment strategy, as well as the need for personalised decisions for the individual patient.

Another topic of debate regarding treatment strategies in RA is the decision whether or not to taper or discontinue treatment in those patients that achieve clinical remission. There are multiple reasons for the interest in tapering as it can reduce costs, treatment burden, and side effects (Kuijper et al., 2017, Fautrel, 2018). Completely ceasing DMARDs tends to result in more flares, which can lead to increased joint damage meaning the risk-to-benefit ratio is likely not favourable for this strategy (Fautrel, 2018). However, tapering of bDMARD dose appears to be successful in a proportion of patients (Fautrel, 2018, Ibrahim et al., 2017). More studies are needed to assess the level of treatment reduction possible for different treatment types as Ibrahim et al. (2017) showed that too large a reduction in anti-TNF treatment can increase the frequency of flares when compared to a small reduction or maintained dose (Ibrahim et al., 2017). Rheumatologists currently have the decision of whether to taper or not but it has

been shown that there is a lack of consistency in this regard (Kuijper et al., 2017).

1.1.2.2 Other inflammatory arthritis

Treatment for PsA typically follows the principal of escalating treatment sequentially, starting with non-steroidal anti-inflammatory drugs (NSAIDs) and topical therapies before building up to biologics following failure of all other treatment options (Coates and Helliwell, 2017, Gossec et al., 2016). Treatments overlap significantly with those used for RA, with methotrexate being the first DMARD used in PsA and many biologics, including TNF inhibitors, being used in both diseases (Coates and Helliwell, 2017). However, the effectiveness of methotrexate for the treatment of PsA is not guaranteed, with one study finding it was actually ineffective for the treatment of synovitis (Kingsley et al., 2012).

Similarly to RA, using a treat-to-target approach improves outcomes in PsA patients, although there is a question over the cost-benefit ratio due to the relatively small improvements seen for a much greater cost and increased incidence of adverse events when compared to standard of care (Coates et al., 2015).

ReA treatment is different to that of PsA and RA, with treatment for the causal infection with antimicrobials sometimes being required. Around 75% of ReA spontaneously resolves over time. Treatment of the synovitis typically involves the use of NSAIDs and glucocorticoids, with DMARDs, such as sulfasalazine, only being used in persistent cases (Selmi and Gershwin, 2014). This difference in treatment again highlights the need for early diagnosis in inflammatory arthritis,

to ensure the correct treatments are given to the right patients and to avoid overtreatment.

1.1.3 Pathophysiology

Rheumatoid arthritis is a chronic inflammatory autoimmune disease, with features characteristic of this, including the presence of autoantibodies such as RF and ACPA. Risk for RA is built up of both genetic and environmental factors, with genetic factors estimated to make up 40-50% of risk in seropositive individuals (Deane et al., 2017). The largest proportion of genetic risk has been associated with the 'shared epitope', which is a group of alleles in the major histocompatibility complex (MHC) human leukocyte antigen (HLA) region. It is thought the shared epitope may increase presentation of citrullinated proteins to T cells, thus increasing the risk of developing ACPA and seropositive RA (Deane et al., 2017). Another of the strongest genetic risk factors lies in the PTPN22 gene, with the risk allele thought to lower thresholds for activation of an immune response, in particular in T cells (Deane et al., 2017). PTPN22 encodes a protein tyrosine phosphatase that modulates signalling in both innate and adaptive immune cells. It inhibits T cell activation by inhibiting signalling downstream of the T cell receptor and induces type I interferon in myeloid cells by increasing downstream pattern recognition receptor signalling (Stanford and Bottini, 2014). In addition to genetic polymorphisms, there is now emerging evidence that epigenetic changes have influence over RA disease course, with changes in synovial fibroblasts being linked to increased disease activity in RA, although it is currently unknown at what stage these changes may occur (Deane et al., 2017). The largest environmental risk factor for RA is smoking, which makes up an estimated 20-30% of the environmental risk (Deane et al., 2017).

Even though there are known genetic and environmental risk factors, it is unknown exactly how RA is initiated. It is thought it may initiate in the mucosa, potentially in the lungs, before spreading to the joints via unknown mechanisms. It is hypothesised that the lungs may be involved as there are multiple environmental risk factors for RA that predominantly affect lung tissue, such as smoking, noxious agents, and silica dust (Guo et al., 2018). Smoking in particular has been linked to pathogenesis in the lungs, with high levels of citrullinated proteins, which act as autoantigens in RA, being found in bronchiolar lavage from smokers (Tracy et al., 2017).

Another potential site of initiation of RA is the gastrointestinal tract. A number of studies have found an altered microbiome in RA patients and those at risk of RA (Horta-Baas et al., 2017, Zhang et al., 2015b, Chen et al., 2016, Alpizar-Rodriguez et al., 2019, Wells et al., 2020, Zaiss et al., 2021). Treatment has been shown to have an effect on this, resulting in a transition towards a more normal microbiome (Zhang et al., 2015b). In mouse models of RA, actively altering the microbiome has been shown to have an impact on disease severity and susceptibility, with absence or removal of the microbiome reducing severity and colonisation with microbiota from RA patients resulting in increased disease incidence in susceptible mice (Jubair et al., 2018, Maeda et al., 2016, Wu et al., 2010, Zaiss et al., 2021). Furthermore, secretory immunoglobulins that are produced at mucosal sites have been found in RA patients, including those reactive to autoantigens associated with RA (Zaiss et al., 2021). Diet is thought to affect RA risk, with high levels of sugar, salt, and red meat increasing risk and vitamin D and antioxidants reducing risk, thus further implicating the gut (Horta-

Baas et al., 2017, Benito-Garcia et al., 2007, Hu et al., 2017, Pattison et al., 2004a, Pattison et al., 2004b, Deane et al., 2017).

The production of autoantibodies alone is not sufficient to cause RA as their presence can be detected years before symptom onset in some patients (Tracy et al., 2017, Horta-Baas et al., 2017, Guo et al., 2018). However, there is evidence to suggest that autoantibodies still play a pathogenic role in the disease. Seropositive disease is associated with a more aggressive disease course and rituximab, a B cell depleting therapy, is more effective in these patients (Derksen et al., 2017). The exact mechanism of pathogenesis is unknown, however it may involve Fc receptor binding, complement activation, neutrophil extracellular trap formation, or osteoclast activation (Derksen et al., 2017, Harre et al., 2012).

Normal synovium is comprised of a thin lining layer that is 1-3 cells thick and a sublining layer. The lining layer is composed of fibroblasts and macrophages, while the sublining is mostly comprised of adipocytes, blood vessels, fibroblasts and resident macrophages (Figure 1.1) (Smith, 2011). Both the lining and sublining is greatly expanded in RA and other inflammatory arthropathies. This is due to the expansion of both resident fibroblasts and macrophages, and the infiltration of immune cells (Figure 1.1) (Smolen et al., 2016a, Bartok and Firestein, 2010, Raza et al., 2005b). The immune cell infiltrate predominantly consists of T cells, B cells, dendritic cells (DCs), and macrophages (Gorman and Cope, 2008). In addition to increased cellularity, which results in thickening, the synovium also becomes more vascularised and invasive. The invasive area of the synovium is termed the pannus, which mostly consists of fibroblasts and macrophages (Gorman and Cope, 2008, Firestein, 2003, Bartok and Firestein,

2010). The pannus is sometimes compared to an invasive tumour, as it shares many cellular and molecular properties with cancerous tissue, including the presence of overlapping somatic mutations (Firestein, 2003, Bartok and Firestein, 2010, Guo et al., 2018). The pannus secretes destructive enzymes, such as matrix metalloproteinases (MMPs), which degrade surrounding cartilage and bone (Gorman and Cope, 2008, Guo et al., 2018).

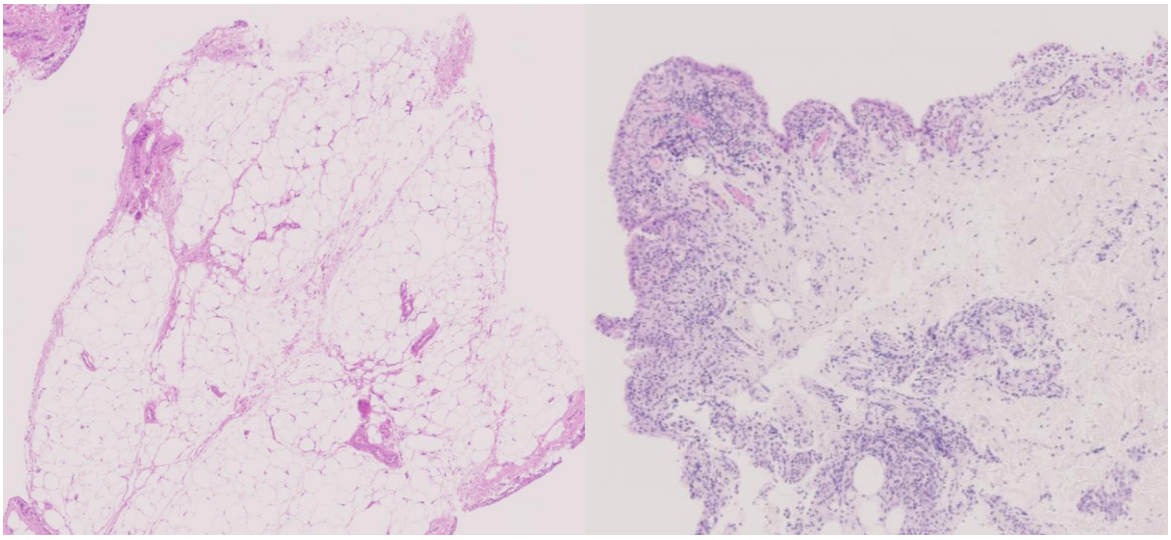


Figure 1.1: Normal vs RA synovium example images. H&E stained synovial tissue from normal (left) and early RA (right) joints. Normal tissue predominantly contains adipocytes. RA tissue shows an enlarged lining layer and infiltrating immune cells. H&E images are from the BEACON cohort and were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

Synovial fibroblasts, along with resident macrophages, play a key role in orchestrating immune responses, with aberrant production of cytokines and chemokines leading to chronic inflammation via recruitment and activation of both innate and adaptive immune cells (Buckley et al., 2001, Bartok and Firestein, 2010, Firestein, 2003). They increase greatly in number during RA pathogenesis and it is thought that this is mediated primarily via increased cell

survival, as cell division in these cell types is relatively rare (Bartok and Firestein, 2010, Guo et al., 2018). In addition to recruitment, synovial fibroblasts can inhibit apoptosis and induce proliferation in a number of immune cell types, further increasing the inflammatory infiltrate (Bartok and Firestein, 2010). Synovial fibroblasts also produce proangiogenic factors, such as VEGF, which helps supports the growing synovium (Bartok and Firestein, 2010). Interestingly, proangiogenic factors, including VEGF and Ang2, have been found to be increased in PsA compared to RA and increased vascularity and sprouting of blood vessels can be seen in SpA compared to RA (Fearon et al., 2003, Baeten et al., 2000, Reece et al., 1999).

T cells interact heavily with synovial fibroblasts and macrophages. This interaction inhibits T cell apoptosis, as well as resulting in their sustained retention, both of which are resolution mechanisms which can reduce T cell numbers back to normal levels (Raza et al., 2005b, Buckley et al., 2001, Bartok and Firestein, 2010). It also induces the production of pro-inflammatory cytokines (Raza et al., 2005b). T cells comprise around 40% of the inflammatory infiltrate, with $CD4^+$ T cells being the predominant subtype, and play a key role in disease pathogenesis (Gorman and Cope, 2008, Raza et al., 2005b, Bartok and Firestein, 2010). This is supported by MHC class II alleles being strong risk factors for RA (Gorman and Cope, 2008, Raza et al., 2005b). It was previously thought that Th1 cells were the primary inflammatory T cell mediator, however more recently it has been demonstrated that Th17 cells, which produce IL-17, may have a more significant contribution to inflammation in RA (Gorman and Cope, 2008, Bartok and Firestein, 2010).

Immature and mature DCs can both be found within the RA synovium, where they help drive a strong T cell response in the presence of inflammation (Gorman and Cope, 2008, Firestein, 2003). However, they may also have a role during disease initiation as they have been shown to travel to the lymph nodes to activate naïve T cells in murine models (Raza et al., 2005b, Gorman and Cope, 2008).

The presence of autoantibodies in RA implicates B cells in its pathogenesis (Gorman and Cope, 2008). They can also activate T cells and produce cytokines, albeit with a less documented significance than other cell types (Gorman and Cope, 2008). B cell apoptosis is inhibited in RA synovium and is primarily mediated by synovial fibroblasts, resulting in increased B cell survival (Bartok and Firestein, 2010). B cells and plasma cells can sometimes be seen in aggregates in the RA synovium, although it is unknown whether plasma cells are formed within the synovium or if they have migrated from elsewhere (Dörner and Lipsky, 2009). Plasma cells and B cells, alongside macrophages, were identified in one study as potentially being able to discriminate RA from other arthritides, including SpA and OA (Kraan et al., 1999).

The use of single cell sequencing in recent years has furthered our understanding of the cellular composition of the synovium, enabling the characterisation of new subtypes of many cell types, including fibroblasts, macrophages, T cells, and B cells, with some subtypes being associated with specific disease processes (Schonfeldova et al., 2022). For example, two subsets of FAPα^{+} fibroblasts have been found to have separate roles in inflammation (FAPα^{+} THY1^{+}) and bone and cartilage damage (FAPα^{+} THY1^{-}), showing a role for fibroblasts in driving both inflammation and joint destruction in RA (Croft et al., 2019). Four macrophage

subtypes were identified in RA synovium, with the MERTK$^+$ subsets being identified as upregulated in remission compared to active disease (Alivernini et al., 2020). Similarly, four B cell subsets have been identified in the synovium, consisting of naïve B cells, memory B cells, autoimmune-associated B cells, and plasmablasts, with the autoimmune-associated B cell subtype being more prevalent in inflamed than OA synovium (Zhang et al., 2019). Single cell investigation of T cells has also uncovered interesting subtypes, with three CD4$^+$ and three CD8$^+$ T cell subsets being identified (Zhang et al., 2019). T peripheral helper cells were found to be associated with inflammation in RA, increasing the production of antibodies by supporting differentiation of B cells (Rao et al., 2017). Increased exploration of the different cellular subtypes present in inflammatory arthritis may help to elucidate further disease mechanisms, potentially enabling the specific targeting of responsible subpopulations in future therapeutics.

1.2 BIOMARKERS

Biomarkers are important tools in the diagnosis and monitoring of disease. They can be used to aid diagnosis, prognosis, and inform treatment decisions (Gavrilă et al., 2016). Diagnostic biomarkers can aid in the early identification of RA patients, allowing for earlier treatment and reduced uncertainty, while prognostic markers can aid in decisions on the aggressiveness of initial treatment. There are multiple different disease measures that may be considered, including disease activity, inflammation, and bone erosion, with the potential for different biomarkers predicting prognosis for different disease measures (Kang et al., 2014, Seror et al., 2016).

Biomarkers are also useful for predicting and monitoring response to therapy. They may help inform, and potentially speed up, treatment decisions, including about when the optimum time is to start a new therapy if the initial one is not successful. This may prove particularly useful when using a treat to target approach, where regular modification of treatment is often required. However, the ideal biomarker for treatment decisions would be one which could predict which treatment will be most successful in a particular patient, allowing for patient stratification and more informed treatment decisions.

1.3 DIAGNOSTIC BIOMARKERS

1.3.1 Imaging biomarkers

The presence of bone erosion is highly predictive of RA so may be a suitable diagnostic biomarker, however is only rarely seen in early disease, particularly when using conventional radiography. Ultrasound (US) and magnetic resonance imaging (MRI) increase both sensitivity and specificity for RA detection (Rahmani et al., 2010). Although MRI is the most sensitive technique of the three it is the most expensive and least readily available. Therefore, US may be an option that increases the sensitivity of radiography without greatly increasing the cost or reducing availability as, although it has lower sensitivity than MRI, there is high agreement between them (Rahmani et al., 2010). The use of US would be a great improvement as, in the same study, US was able to pick up 28 erosions versus the 5 that radiography was able to detect (Rahmani et al., 2010). Although US is an improvement upon radiography in this regard, there are still only very few patients that present with bone erosion at very early stages, meaning that even if all patients with erosions are detected, there would still be

a large number of RA patients missed at baseline if this were the sole criterion used (Filer et al., 2011).

US and MRI can also detect swelling and inflammation in joints. Both modalities are able to detect subclinical disease, therefore potentially being able to diagnose RA earlier than clinical observation allows. Filer et al. (2011) demonstrated that scanning of selected joints using US may improve upon the sensitivity and specificity of the 2010 ACR/EULAR criteria and a study by Sidhu et al. (2021) found that synovitis in small joints detected by MRI could aid in the prediction of progression to RA in undifferentiated arthritis affecting large joints.

The detection of tenosynovitis in specific joints of the hand has also been found to aid diagnosis of early RA (Rogier et al., 2020). This is detectable by both MRI and US, with a study by Sahbudin et al. (2018) finding that addition of tenosynovitis to a model including ACPA positivity and presence of joint synovitis, as measured by US, could improve the prediction of RA development (Eshed et al., 2009, Sahbudin et al., 2018).

1.3.2 Peripheral blood biomarkers

The 2010 ACR/EULAR criteria currently include use of multiple peripheral blood markers, namely RF, ACPA, CRP and ESR (Aletaha et al., 2010). Of these, RF was the only marker previously used in the 1987 criteria (Arnett et al., 1988b, Gavrilă et al., 2016). RF and ACPA have specificities of 85% and 96% respectively, with both having sensitivities of less than 70%, meaning they miss a significant number of RA patients (Shi et al., 2015). However, using RF and ACPA together increases overall sensitivity as positivity is not completely overlapping. RF, and to a lesser extent ACPA, can be detected in other conditions

including PsA, SLE, and Sjogren's syndrome, with RF also being associated with age and chronic infections, such as tuberculosis (Tseng et al., 2009, Renaudineau et al., 2005). CRP and ESR are both systemic markers of inflammation, so lack specificity for synovitis but allow for the detection of inflammation (Tseng et al., 2009, Mc Ardle et al., 2015).

Following those included in the 2010 ACR/EULAR criteria, perhaps the next most utilised diagnostic biomarker is the 14-3-3η protein, as it is included in tests that have been approved for use in the US and Canada (Gavrilă et al., 2016). 14-3-3η appears to be a promising candidate for diagnosis of early RA as, when used alongside ACPA and RF, 78% of RA patients could be identified (Maksymowych et al., 2014).

Other autoantibodies have been investigated for potential use as diagnostic biomarkers. Anti-mutated citrullinated vimentin antibodies (anti-MCV) have been found to have similar sensitivity and specificity as ACPA but the two biomarkers are independent of each other (Gavrilă et al., 2016, Luime et al., 2010). This means anti-MCV may be a useful additional tool in currently seronegative patients, who are typically harder to diagnose than their seropositive counterparts are.

Anti-carbamylated protein antibodies and anti-acetylated vimentin antibodies have also been investigated. However, Anti-carbamylated protein antibodies have lower sensitivity and specificity than RF and ACPA and can also be present in a range of other inflammatory joint diseases and anti-acetylated vimentin antibodies have high specificity (86%) but low sensitivity for RA (37%) (Shi et

al., 2015, Juarez et al., 2016). Therefore, neither of these antibodies add diagnostic value compared to currently used markers.

Low IL-7 levels have been associated with RA and, although this has limited standard use due to having lower sensitivity and specificity than ACPA and RF, it has limited overlap with ACPA so may aid in the identification of ACPA-negative RA patients (Goeb et al., 2013).

1.3.3 Joint biomarkers

1.3.3.1 Synovial fluid

There have been many studies looking at peripheral blood biomarkers in RA as they can be tested with minimally invasive techniques. However, there is likely an increased chance of identifying specific and sensitive biomarkers in the joints of RA patients, as this is where predominant disease pathology occurs. This is highlighted by a recent study, which showed that connective tissue growth factor (CTGF) has relatively high sensitivity and specificity when measured in the serum (86% and 92%, respectively) (Yang et al., 2017). However, this is increased further in the synovial fluid, with 96% sensitivity and 91% specificity being recorded (Yang et al., 2017).

Calprotectin has also been studied as both a serum and synovial fluid biomarker. In serum, levels correlate with disease activity and are reduced following treatment response in both RA and PsA. It is increased in serum of RA patients compared to SpA but finding a diagnostic cut-off has proved challenging (Ometto et al., 2017). However, calprotectin is more clearly increased in RA above other inflammatory arthritis in the synovial fluid, making its use as a diagnostic biomarker more feasible in synovial fluid than in serum (Ometto et al., 2017).

The synovial fluid cytokine profile has been shown to be discrete in early RA compared to other inflammatory arthritis and established RA. In particular IL-13, IL-2, IL-4, IL-15, bFGF, and EGF have been shown to be the most important for this distinction (Raza et al., 2005b). This may be useful for diagnosis of early RA, however due to their transient nature may only be useful within a short window of disease duration. It is possible this signature may be associated with the 'window of opportunity' for commencement of treatment, in which remission is most successfully induced, which means it could have utility for informing treatment decisions.

PAD2, one of the enzymes responsible for citrullination of proteins, is another potential biomarker for RA, being found in 3-fold higher levels in RA compared to OA (Damgaard et al., 2016). However, PAD2 is found in higher levels in the relatively easier to diagnose ACPA-positive patients compared to ACPA-negative patients (Damgaard et al., 2016). In addition, this study compares RA and OA, meaning it may not be able to distinguish RA from other inflammatory arthritis but rather identify inflammatory disease.

A number of miRNAs have been shown to distinguish RA from OA when measured in synovial fluid. Namely miR-16, miR-146a miR-155 and miR-223 have all been shown to be significantly increased in RA (Murata et al., 2010). However, this study again only included RA, OA, and healthy controls, meaning that their utility in distinguishing RA from other inflammatory arthritis is currently unknown.

1.3.3.2 Synovial tissue

Synovial tissue is likely to be even more representative than synovial fluid, as it is where RA predominantly manifests. It is therefore likely to have the most differences between patient groups and hold the most information around mechanisms of disease. Synovial tissue is also the most challenging to obtain as it requires a relatively more invasive biopsy procedure. However, synovial biopsy techniques are generally well tolerated and have even been associated with reduced joint pain, swelling, and stiffness after the procedure (Just et al., 2018).

It has been suggested that there is an activated fibroblast phenotype present in early RA (Choi et al., 2017). This may mean that activation markers in synovial tissue are able to distinguish early RA from other clinical outcomes. One of these such markers is FAP, which was shown to be increased in both the lining and the sublining layer, with the sublining expression discriminating between RA and other inflammatory arthritis (Choi et al., 2017).

A study by Yeo et al. (2016) showed that CXCL4 and CXCL7 can differentiate early RA from resolving and established disease. They were shown to be produced by synovial macrophages, suggesting that this difference may only been seen in the synovium and may be missed in peripheral blood. In addition, the majority of CXCL4 and CXCL7 was seen outside the vasculature, further supporting this hypothesis (Yeo et al., 2016). These markers could potentially be useful for differentiating patients with RA from those who will resolve, aiding in both diagnosis and treatment decisions.

Many studies to date have looked at later stage RA, meaning that the biomarkers identified are not guaranteed to be applicable in early RA, when it is the most

challenging to diagnose. There is a need for further study of RA-specific biomarkers in clinically relevant cohorts including patients with undifferentiated arthritis and other inflammatory arthritis. Biomarkers identified in these cohorts would be able to aid early diagnosis and allow for early disease-specific treatment without over treatment of patients who are destined to resolve.

1.4 SYNOVIAL TISSUE HETEROGENEITY AND PATIENT STRATIFICATION

Histology has hugely helped develop our understanding of arthritis by enabling visualisation of the tissue pathology. This has allowed for the documentation of the level of heterogeneity that exists in the synovial tissue of patients with inflammatory arthritis, with some studies trying to capture this heterogeneity for the stratification of patients, particularly in RA. Heterogeneity exists between patients but also within patients, with different areas of a single joint often displaying differing phenotypes (Kennedy et al., 1988, Dolhain et al., 1998, Boyle et al., 2003). For this reason, investigation of the synovial tissue requires multiple different areas to be sampled in order to get an overall picture of the whole joint.

There have been a number of studies looking to score tissue in order to capture heterogeneity between patients. Some scoring systems have been developed using hematoxylin and eosin (H&E) staining as it is a relatively simple technique that can be used to look at the morphology of tissue. Rooney et al. (1988) developed such a scoring system that captures the level of synoviocyte hypertrophy, fibrosis, proliferating blood vessels, perivascular infiltrates of

lymphocytes, focal aggregates of lymphocytes, and diffuse infiltrates of lymphocytes, although it was never widely adopted due to its complexity.

Krenn et al. (2006) later developed another semi-quantitative scoring method, also based on H&E staining, that scored the hyperplasia of the lining layer, level of inflammatory infiltrate, and density of resident cells and that was able to distinguish degenerative and inflammatory arthritis. There was no difference between the scores for RA, PsA, and ReA so it is unable to distinguish between different inflammatory arthropathies and was developed on later stage disease, which generally has more severe inflammation than early disease. However, this scoring system was simpler than the score developed by Rooney et al. (1988) so has been used more widely.

Orange et al. (2018) looked at 20 features of synovium, also on later stage disease and using H&E, and found that there was a significant amount of interrater variability in a number of the features and many were not present in any of the samples tested. Therefore, they cut this down to scoring 10 features that had fair reproducibility across scorers and that were present in at least 5% of samples, resulting in the scoring of plasma cells, binucleate plasma cells, Russell bodies, giant cells, neutrophils, fibrin, mucin, detritus, lining hyperplasia, and lymphocytes. Alongside histology, they used RNA sequencing to identify three patient subgroups based on their transcriptomic profile, comprising low, mixed, and high inflammatory subtypes. Those classified as high inflammatory had higher clinical inflammatory markers, such as ESR and CRP, although there was no correlation with pain, tender or swollen joint counts, or disease duration. They then looked to replicate these patient subgroups using their H&E scoring and a machine-learning model to allow for cheaper and quicker classification of

patients. This showed that both high and low inflammatory subtypes could be identified reasonably well but that patients in the mixed inflammatory group were harder to identify using H&E alone (Orange et al., 2018).

A study by Lewis et al. (2019) also integrated histology and transcriptomics but on treatment-naïve early RA and identified subgroups, or pathotypes, in histology first before exploration in transcriptomics. In this study they used immunohistochemistry staining for T cells (CD3), B cells (CD20), macrophages (CD68), and plasma cells (CD138) to identify three pathotypes: pauci-immune fibroid, which lacks immune infiltrate; diffuse-myeloid, which has sublining macrophage infiltration; and lympho-myeloid, which has B cell aggregates. Overall, they found that the histology broadly matched the gene expression, with scoring for the presence of certain cell types correlating with corresponding gene modules for the same cell types. Furthermore, they found that patients with a higher inflammatory phenotype at baseline had better response to DMARD treatment after 6 months, suggesting potential clinical utility for the stratification of patients in this way. Interestingly, analysis of the histological pathotype after 6 months of DMARD treatment showed that patients who changed to a less inflammatory pathotype had larger reductions in disease activity when compared to those who changed to a more inflammatory pathotype (Lewis et al., 2019). Altogether, this suggests that analysis of histological pathotype when using IHC staining has utility for patient stratification and predicting response to treatment and that following successful treatment patients can convert to less inflammatory pathotypes.

Other studies have also identified phenotypes that are associated with response versus non-response to treatment, with myeloid pathotypes being found to have

better response to anti-TNF therapies, while lymphoid pathotypes did not (Dennis et al., 2014, Canete et al., 2009). Exploration of gene expression profiles prior to treatment identified the presence of T cell and macrophage-related gene signatures as being correlated with future response to rituximab, while genes associated with remodelling and IFNα were indicative of poor response (Hogan et al., 2012). These studies provide a good case for the future use of biomarkers for the stratification of patients based on predicted response to specific treatments, thereby highlighting potential for a future of personalised medicine in rheumatoid arthritis.

1.5 HYPOTHESES/AIMS

This study was designed to investigate in detail the histology and gene expression of an early inflammatory arthritis cohort. This was intended to build upon work carried out by Yeo et al. (2016), in which they used PCR to investigate cytokines and identified CXCL4 and CXCL7 as being able to distinguish early RA from established RA and resolving disease. It was hypothesised that by looking at both histology and whole tissue transcriptomics, further biomarkers and mechanisms of disease could be identified that would aid in the diagnosis and stratification of early inflammatory arthritis patients and that might identify new genes or pathways that could be targeted to drive resolution of inflammatory arthritis.

The overarching aim of this project is to investigate biomarkers of outcome and understand mechanisms of disease in early inflammatory arthritis by integrating novel methods for classifying H&E stained tissue, including pathotype derivation and AI-assisted analysis, with whole tissue RNA sequencing data.

I set out to address the following main hypotheses:

1. Pathotypes, identified from H&E staining alone, will identify early inflammatory arthritis patient subtypes and correlate with clinical variables, including disease activity, markers of local and systemic inflammation, and response to treatment.
2. Differentially expressed genes or signatures in short duration RA patients will distinguish them from other patient groups.
3. Mechanisms that drive resolution versus persistence can be identified using whole tissue RNA sequencing.

A number of secondary hypotheses were also identified, aiming to answer additional questions in the early inflammatory arthritis cohort:

1. Pathotypes, identified histologically, will correlate with gene expression data.
2. There are genes or pathways that differ between short duration and long duration RA that can be identified using whole tissue RNA sequencing.
3. Biomarkers for good or poor outcomes in RA can be identified in gene expression data.

2 METHODS

2.1 PATIENT COHORT AND SAMPLE COLLECTION

Patient samples were obtained from the Birmingham Early Arthritis Cohort (BEACON). Treatment-naïve patients were recruited from either the rapid access clinic at Sandwell and West Birmingham Hospitals NHS Trust, the Early Arthritis clinic at University Hospitals Birmingham NHS Foundation Trust or the Modality primary care consortium rheumatology unit. Patients were categorised into one of four different outcomes after 18-month follow up, namely short duration RA (sdRA), long duration RA (ldRA), persistent non-RA disease (nonRA) and resolving disease (Res) (Figure 2.1). Patients were classified as sdRA if they had symptoms for less than 3 months at baseline, and ldRA if they had symptoms for 6 months or longer at baseline, with both RA groups meeting 2010 ACR/EULAR and/or 1987 ACR criteria at 18-month follow up (Arnett et al., 1988a, Aletaha et al., 2010). Resolving disease had symptom duration of less than 3 months at baseline and was classified by an absence of clinical joint swelling at 18-month follow up without the use of DMARD or glucocorticoid treatment in the previous 3 months. Persistent nonRA also had symptom duration of less than 3 months at baseline and included patients with chronic joint inflammation but with a diagnosis other than RA at 18-month follow up.

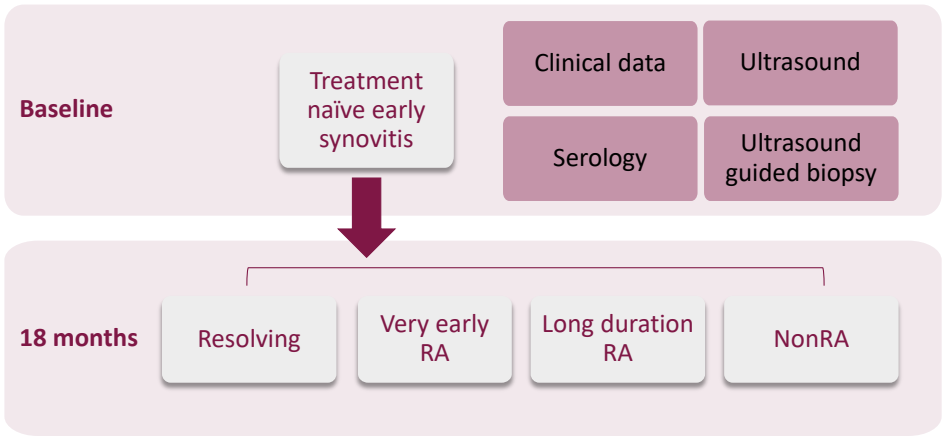


Figure 2.1: Diagram showing the structure of the BEACON cohort.

Synovial tissue was taken using ultrasound-guided biopsies, as previously described (Kelly et al., 2015), and subsequently processed. Tissue was fixed, embedded in OCT for frozen sections, and snap frozen as whole tissue in liquid nitrogen. Tissue was also placed in formalin and underwent Paraffin embedding, sectioning and H&E staining in the University Hospitals Birmingham pathology department. For each tissue preservation format, 6-8 representative biopsy fragments were processed in order to account for heterogeneity within the joint (Kennedy et al., 1988, Dolhain et al., 1998, Boyle et al., 2003).

Normal control samples were obtained at exploratory arthroscopy from patients with joint pain but no imaging, macroscopic or histological synovial abnormalities. All patients gave written informed consent and the study had ethical approval granted by West Midlands Black Country research ethics committee: "Outcomes in patients with inflammatory arthritis (BEACON)", ethics reference: 12/WM/0258; and "Ultrasound guided synovial biopsy in patients with arthritis (BEACON biopsy)", ethics reference: 07/H1203/57. Summaries of

patient characteristics can be found in Chapters 3 and 4 and full characteristics can be found in in Appendix 9.1.

2.1.1 AMP RA Phase II cohort

The Accelerating Medicines Partnership (AMP) RA phase II cohort was used as an external control for the BEACON scoring system. The cohort is a mixed RA cohort consisting of treatment naïve or minimally exposed patients, methotrexate inadequate responders (≥12 weeks of treatment with stable dose for 4 weeks), and patients with inadequate response to one or two TNF-alpha inhibitors (≥12 weeks of treatment with TNF inhibitors), with a range of disease durations. Patients were recruited and samples collected across 15 sites (13 in the US and 2 in the UK), with clinical data being collected at baseline and after 3 and 6 months. Synovial tissue was acquired using either ultrasound-guided biopsies or surgical procedures and 6-8 fragments were formalin fixed and paraffin embedded prior to H&E staining. All patients gave written informed consent and the study had ethical approval under the same approvals as the BEACON cohort (see above).

2.2 HISTOLOGY STAINING AND SCORING

Following paraffin embedding, sectioning and H&E staining, H&E stained slides were scanned at x20 magnification using a Zeiss Axio Scan Z1 by Dr Jennifer Marshall, Dr Triin Major, Dr Emily Taylor or myself and images stored centrally to facilitate reporting by multiple observers using the Zeiss software Zen 2012 (blue edition).

H&E stained images from synovial biopsies were independently scored for four characteristics – lining layer thickness, Krenn inflammatory infiltrate score,

BEACON density, and BEACON aggregates. BEACON density and aggregates scores are described in Chapter 3.2. The lining layer thickness was recorded as an absolute cell number, with the number of cells being manually counted out in a straight line from the sublining to the edge of the lining layer. As there were multiple fragments per H&E image, fragments were scored individually for each section, before a mean or worst-case score was calculated. If a fragment was not considered to be synovial tissue, as determined by the absence of both a lining layer and a characteristic synovial tissue structure, then it was excluded from scoring.

The Krenn inflammatory infiltrate score was scored as described previously (Krenn et al., 2006), and is as follows:

- 0 points – no inflammatory infiltrate
- 1 point – few mostly perivascular situated lymphocytes or plasma cells
- 2 points – numerous lymphocytes or plasma cells, sometimes forming follicle-like aggregates
- 3 points – dense band-like inflammatory infiltrate or numerous large follicle-like aggregates

Graphs and statistics were produced using GraphPad Prism v9.0.0, with the exception of calculation of intra-class coefficients and weighted Kappa coefficients, which were calculated using the irr package in R.

2.2.1 AMP RA Phase II validation

Scoring of the AMP RA Phase II cohort was undertaken independently by Prof Ellen Gravallese, Prof Brendan Boyce, and Prof Edward DiCarlo for BEACON density, aggregate, and pathotype, as described above. A consensus was then

determined and used for analysis, using a mean score for density and median for aggregates and pathotype. Graphs and statistics were produced by myself using GraphPad Prism v9.0.0.

2.2.2 AI-assisted Scoring

AI-assisted scoring was carried out using the Visiopharm Integrator System (VIS). Initial proof-of-concept trial for the detection of tissue, lining layer, and aggregates, with some further Analysis Protocol Packages (APP) development was carried out by myself. These APPs where then further developed and validated alongside new APPs for the detection of lymphocytes and plasma cells by Dr Nasullah Khalid Alham using VIS version 2020.09.0.8195. Identification of features for training and validation was carried out using AIDA (Annotation of Image Data by Assignments) by Prof Clare Verrill, Dr Andrew Filer, and Prof Dagmar Scheel-Toellner. AIDA is an annotation platform developed at the University of Oxford by Alan Aberdeen, Dr Nasullah Khalid Alham, Prof Clare Verrill, and Prof Jens Rittscher.

Tissue and aggregate detection APPs were based on threshold classification, lining layer was developed using DeepLabv3 in the AI module, and lymphocyte and plasma cell detection APPs were developed using the U-Net model also within the AI module. Lining layer, plasma cell and lymphocyte APPs were developed through two, three, and five rounds of training, respectively, with training annotations being accepted where there was agreement between at least two of the three annotators. Validation was undertaken using AIDA, with reviewers marking annotations as ‘correct’, ‘partially correct’, ‘not sure’, or ‘incorrect’ for the lining layer and aggregate APPS and with comparison of cell

counts within small areas used for the lymphocyte and plasma cell APPs. Validation of the plasma cell and lymphocyte APPs is still ongoing.

2.3 RNA Extraction and Library Preparation

Prior to conducting RNA extraction on samples from the BEACON cohort, 6 trial runs were undertaken in order to optimise technique and ensure high concentrations of good quality RNA could be obtained from tissue. Tonsil samples obtained from patients undergoing tonsillectomy were used for initial trial runs (n=4), before moving on to synovial tissue from late stage RA patients undergoing joint replacement (n=2). These samples were obtained from the Royal Orthopaedic Hospital NHS Trust from patients seen by Mr. Andrew Thomas under existing ethical permissions. Trial runs (n=4) were also undertaken on thick sections (TS) of synovial biopsies from the BEACON cohort to ensure high quality RNA could be extracted from these.

Snap frozen whole tissue (WT) or 40 µm thick sections (TS) (sectioned on a cryostat from frozen tissue embedded in OCT) were transferred to a PowerBead Tube (Qiagen) containing 2.38 mm metal beads and 300 µl buffer RLT prior to homogenisation in a TissueLyser II (Qiagen) for 2 minutes at 30 Hz. RNA was extracted from the resulting sample using an RNeasy Fibrous Tissue Mini Kit (Qiagen) according to manufacturer's instructions. Aliquots of RNA samples were then sent to Genomics Birmingham for quality control (QC) analysis on an Agilent 2200 TapeStation system and Qubit, which was carried out according to manufacturer's instructions. An RNA integrity number (RIN) and concentration was obtained (Figure 2.2). The RIN is an algorithm that captures the level of RNA degradation in a sample by assessing the signal intensity of ribosomal RNA (28S

and 18S) compared to the signal from shorter RNA fragments, as visualised on an electropherogram (Schroeder et al., 2006). Samples were accepted if they had a RIN score over 5, or if the electropherogram showed minimal RNA degradation when assessed in collaboration with Genomics Birmingham, despite a low RIN score.

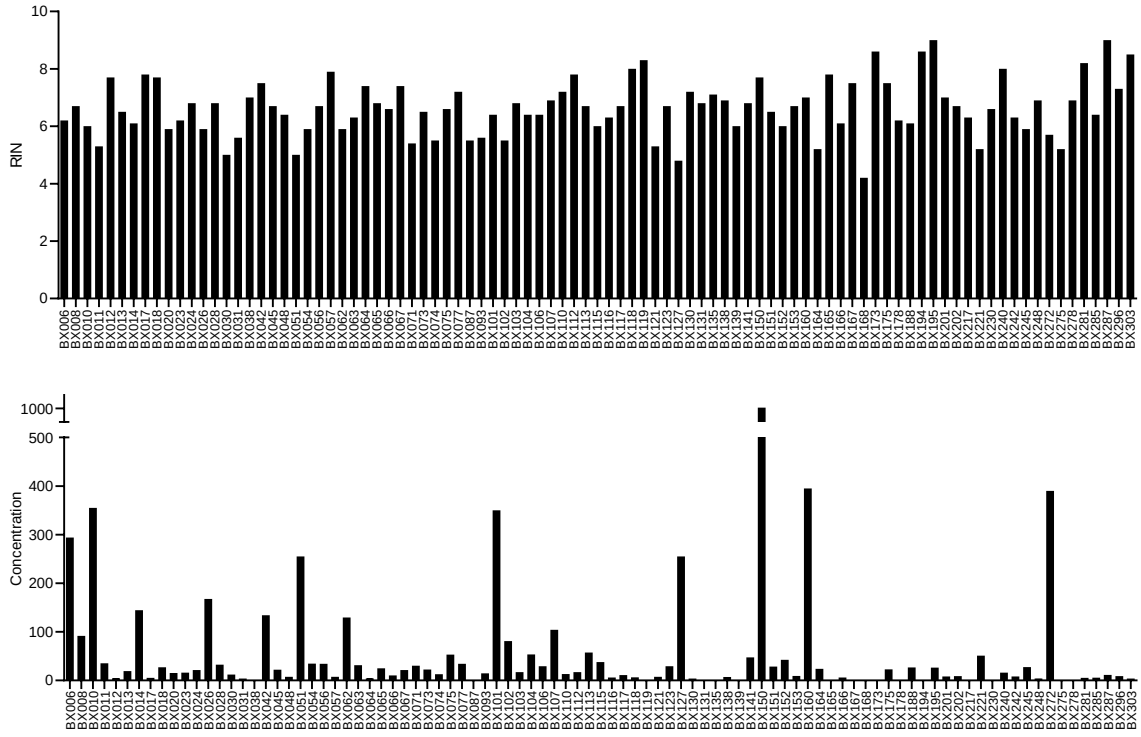


Figure 2.2: RIN scores and concentrations of samples for sequencing. RIN scores (above) and concentration (below) were measured by Genomics Birmingham using a Tapestation and a Qubit, respectively.

Samples were normalised to contain 60ng RNA in 50 µl volume and sent to Genomics Birmingham for library preparation and sequencing. Library preparation was conducted using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and the NEBNext Poly(A) mRNA Magnetic Isolation Module according to manufacturer's instructions. A sample with high concentration and RIN score was split across all library preparation batches to allow for

quantification of batch effect during analyses. Figure 2.3 shows the resulting library sizes and concentrations. Twelve samples had lower library concentrations than expected (<3 ng/μl). Ten of these were repeated with higher input RNA concentration but two had very low initial RNA concentrations so were excluded at this stage. Samples were sequenced to a read depth of 25 million reads per sample using NextSeq 500/550 v2.5 flow cells on a NextSeq system (Illumina).

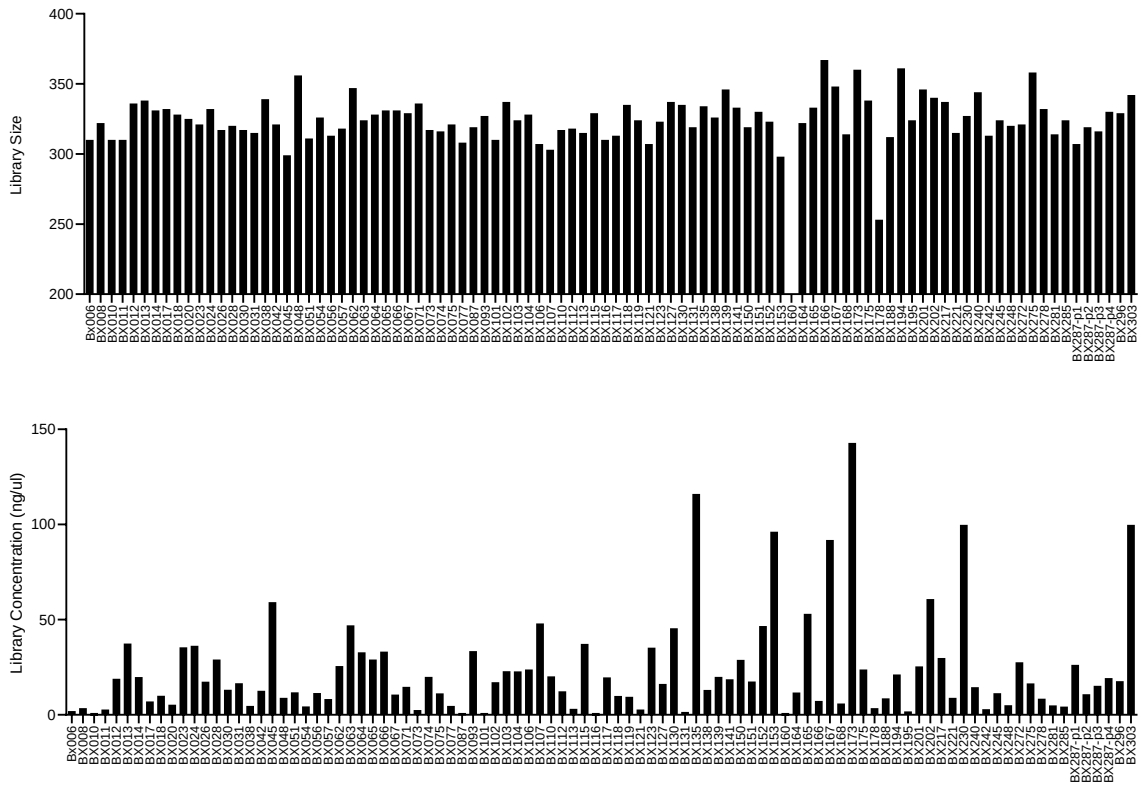


Figure 2.3: Library preparation quality control. Library preparation was conducted by Genomics Birmingham. Library size (above) and concentration (below) are shown.

2.4 RNA SEQUENCING DATA ANALYSIS

The sequencing data were prepared for analysis under the supervision of Dr Csilla Varnai and Dr Jason Turner using CaStLeS (Compute and Storage for Life

Sciences), which is a compute and storage resource for life sciences at the University of Birmingham (Thompson et al., 2019). Lanes were merged (script in Appendix 9.3), contaminating adapter sequences were trimmed using cutadapt (Martin, 2011), and QC was undertaken using FastQC (Andrews, 2010). BAM files were merged (script in Appendix 9.3), reads were aligned using STAR (Dobin et al., 2013) to the NCBI human reference genome GRCh38 (Schneider et al., 2017) and summarised using Subread featurecounts (Liao et al., 2014). 2 samples were excluded following FastQC, due to failure of sequence quality and GC content checks.

Subsequent analysis was carried out in R studio version 1.4.1106 using R version 4.1.0 under the supervision of Dr Csilla Varnai, with help from Dr Jason Turner (RStudio Team, 2021, R Core Team, 2021). Packages used for analysis were from Comprehensive R Archive Network (https://cran.r-project.org/) or Bioconductor (https://www.bioconductor.org/).

Normalisation and differential expression analysis was undertaken using DESeq2 (Love et al., 2014), which is based on a negative binomial model and uses the Wald test for hypothesis testing. Features with fewer than 50 raw counts across all samples were excluded to reduce the size of the dataset. This is an arbitrary cut-off to reduce the size of the dataset and decrease processing times, rather than to increase statistical power, as DESeq2 performs independent filtering based on normalised counts to address this. Stepwise regression analysis that was used to test for factors that influence variation in the dataset was done using the olsrr package (Hebbali, 2020). Heatmaps of the most variable and differentially expressed genes were produced using ComplexHeatmap (Gu et al., 2016).

DESeq2 was also used for variance stabilisation transformation (VST), for normalisation prior to visualisation of the data, and for Principal Component Analysis (PCA). PCA gene contribution plots and scree plot were produced using the factoextra package (Kassambara and Mundt, 2020). UMAP was carried out using the umap package (Konopka, 2020) and k-means clustering using the R stats package. Tests for optimal number of clusters were conducted using factoextra (Kassambara and Mundt, 2020). K-means is a simple algorithm that forms clusters around centroids, grouping similar samples together in an unsupervised way, given the number of clusters. K-means clustering was selected for this analysis as it is simple and gives rapid clustering solutions, and because the nature of the dataset meant the limitations of the approach would have minimal impact on results. This was due to the lack of outliers in the dataset, which can have large affects on the clustering results with this algorithm, the clear number of optimal clusters given by clustering tests, and the presence of relatively uniform circular clusters. Furthermore, the use of k-means clustering on UMAP data has previously been shown to have good performance (Hozumi et al., 2021).

clusterProfiler was used for Gene ontology (GO) pathway analysis (Yu et al., 2012). GO analysis was only carried out on genes with associated Entrez IDs. Scripts for this analysis were adapted from a script developed by Dr Dina Abdelmottaleb. The R scripts used for analysis of the RNA sequencing dataset can be found in Appendix 9.4.

Metabolic gene lists used for exploration of metabolic pathways in Chapter 6 were manually curated using KEGG and relevant literature by Dr Valentina Pucino

(Pucino et al., 2019). Genes included in each process can be found in Appendix 9.7.

3 HISTOLOGY SCORING

3.1 INTRODUCTION

The histology of the synovium in RA has been an area of study for a long time, however there is still a lot that is unknown about the causes and consequences of synovial heterogeneity. Furthermore, until recently synovium was predominantly obtained during arthroplasty, which is only undertaken in very late-stage disease. With the increasing availability of synovium from patients with early disease due to the development of less invasive biopsy techniques, there is a need to understand how early stage RA differs from both the late-stage disease that has been studied previously and from other inflammatory arthritides at an early stage. Furthermore, stratification of patients with early inflammatory arthritis may enable targeted treatment, making this an important area of research.

There have been a number of scoring systems developed to try to capture some of the heterogeneity of the synovium in RA, varying in complexity of the staining and expertise required for scoring (Rooney et al., 1988, Krenn et al., 2006, Orange et al., 2018, Lewis et al., 2019). An early scoring system by Rooney et al. (1988) was developed on H&E stained tissue to capture variation in the synovial tissue from late-stage RA patients. In this scoring system six features of the synovium are scored on an 11-point scale (0-10), including synoviocyte hyperplasia, fibrosis, blood vessels, and three different lymphocyte measures (Rooney et al., 1988). However, this scoring system was not widely adopted due to its complexity.

A scoring system that is commonly used was developed by Krenn et al. (2006) to distinguish rheumatic disease from degenerative disease, such as osteoarthritis. This system was adopted in favour of the score by Rooney et al. (1988) predominantly due to its simplicity. The score is based on readily available H&E staining but, having been developed predominantly on late-stage disease, may lack the sensitivity required to stratify early stage inflammatory arthritis. In addition, the scoring lacks clarity and has minimal example images to support reliable, reproducible scoring (Figure 3.1). This system also does not provide a way of stratifying into pathotypes and combines the complexity and density of inflammatory infiltrate into a single score.

A more recent approach, developed by Orange et al. (2018), undertook a comprehensive exploration of the RA synovium, scoring 20 histological features for presence and reliability, again from H&E stained tissue, and relating these to subtypes identified from gene expression data, comprising low, mixed, and high inflammatory phenotypes. Plasma cell features were found to define the high inflammatory subtype, although histological features were less able to separate the low and mixed inflammatory subtypes. This scoring approach required extensive expertise in synovial histology, making it inaccessible in most research settings, and also beyond the reach of most NHS departments in the UK, and was developed on tissue from late-stage disease.

Lewis et al. (2019) also used both histology and gene expression data to identify pathotypes in RA but using biopsies from treatment-naïve early RA patients with symptom duration of less than 12 months. They identified three pathotypes; lympho-myeloid, diffuse-myeloid, and pauci-immune fibroid, but required

immunohistochemistry staining for B cells, T cells, macrophages, and plasma cells.

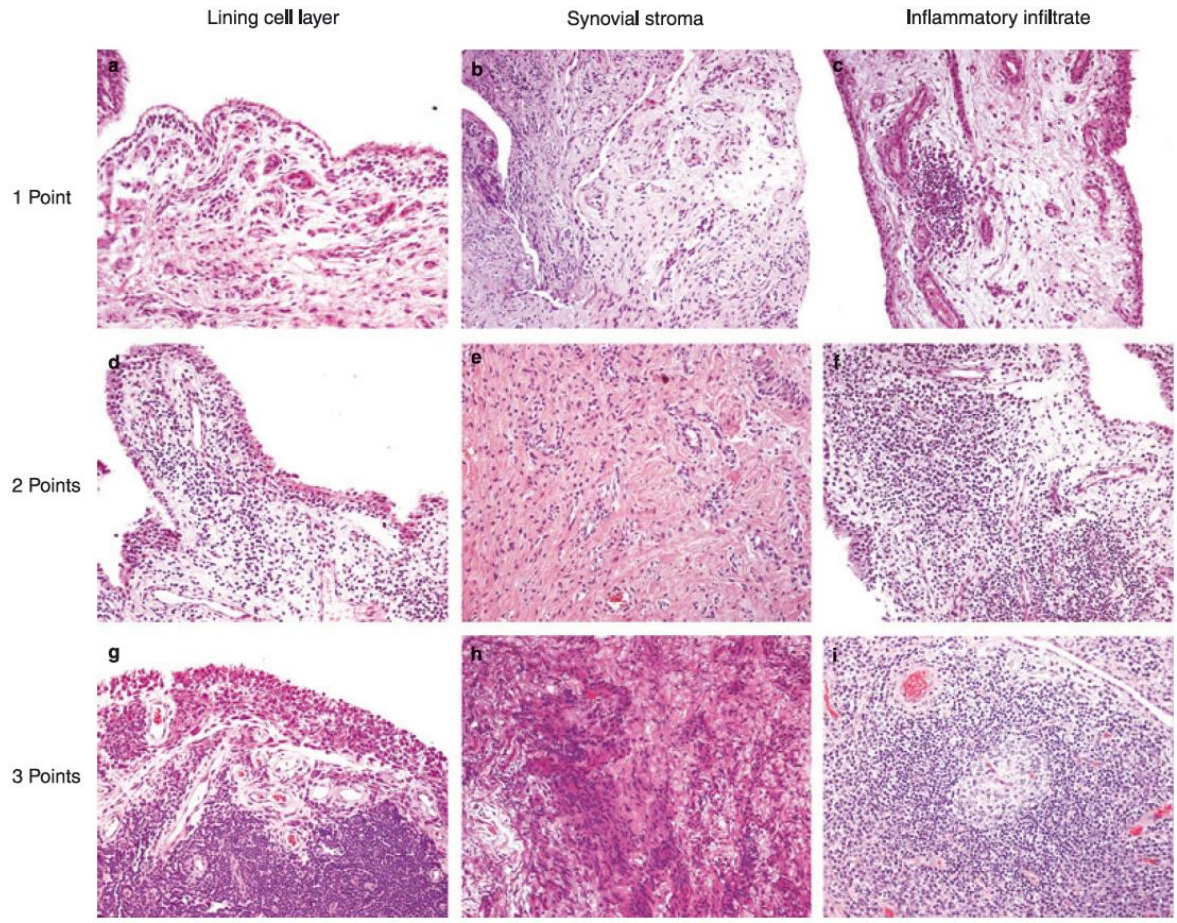


Figure 3.1: Example images provided for the scoring system developed by Krenn et al. (2006). This figure is Figure 1 in the Krenn et al. paper and shows example images and scoring for lining cell layer, synovial stroma, and inflammatory infiltrate.

This study therefore aimed to improve upon these scoring systems by creating a new system that is simple to use, enables pathotype derivation, and is based on H&E staining alone. This new scoring system was developed on a mixed inflammatory arthritis cohort, rather than solely on RA, making it applicable across all inflammatory arthritides. Furthermore, the possibility of developing an AI-assisted scoring system was explored, which would allow for automatic

detection of features of the synovium with more quantitative associated measures without the need for relatively time-intensive manual scoring.

3.2 Scoring Development

The BEACON scoring system was formulated to improve upon currently available scoring systems by providing a simple system that allows for pathotype derivation from H&E stained synovial tissue. It was developed on a cohort of treatment-naïve early inflammatory arthritis patients from the BEACON cohort. This scoring system involves grading the density of the immune infiltrate and the degree of complexity of aggregates, and uses these to calculate a summary pathotype for the overall tissue. The aggregate grading system was broadly based on aggregate scoring from previous studies developed using staining, predominantly the system described by Manzo et al. (2005) and used by Lewis et al. (2019) in their pathotype derivation. The density and aggregate grades are scored as follows:

BEACON density grading:

- Grade 0 – no infiltrate
- Grade 1 – low density (scarce) infiltrate
- Grade 2 – medium density infiltrate
- Grade 3 – high density or band like infiltrate

BEACON aggregate grading:

- Grade 0 – no aggregates
- Grade 1 – low 6-9 radial cell count
- Grade 2 – medium 10-19 radial count

- Grade 3 – high ≥ 20 radial count

Aggregates are graded on maximum longitudinal diameter of the structure, with both lymphocyte and plasma cell aggregates allowed. For each biopsy 6-8 fragments are obtained to ensure any tissue heterogeneity is captured, so fragments are graded independently before summary grades for the whole tissue are calculated. Density is reported as a mean of fragment grades and aggregates as a worst case of any fragment. An atlas describing the grading was created, including representative images whose grades were agreed upon by three independent scorers (Appendix 9.2). Representative images for the grading can be seen in Figure 3.2.

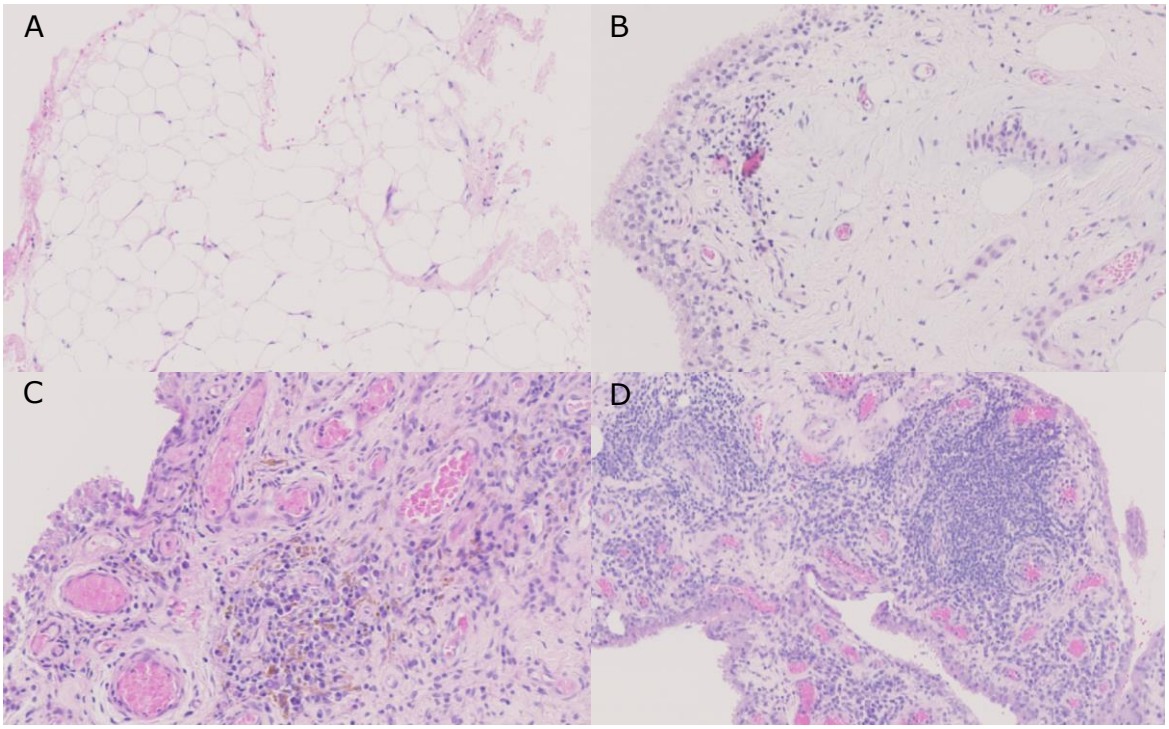


Figure 3.2: Representative images showing varying levels of immune infiltrate. (A) An uninflamed section, grade of 0 for both density and aggregates. (B) Grade 1 density and no aggregates. (C) Grade 2 density and no aggregates. (D) Large aggregate with a grade of 3, grade 3 density. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

Pathotypes are then assigned based on the density and aggregate grading as follows:

- Lymphoid: presence of significant aggregates
    - Presence of ≥1 grade 1 aggregate *in at least two fragments*, or any grade 2 aggregate, or any grade 3 aggregate.
- Diffuse: presence of immune infiltrate but absence of significant aggregates
    - Does not meet lymphoid criteria, mean fragment density grade ≥1.
- Pauci-immune: Absence of immune infiltrate

- Does not meet lymphoid criteria, mean density grade <1.

Representative images of tissues classified as pauci-immune, diffuse, and lymphoid can be seen in Figure 3.3.

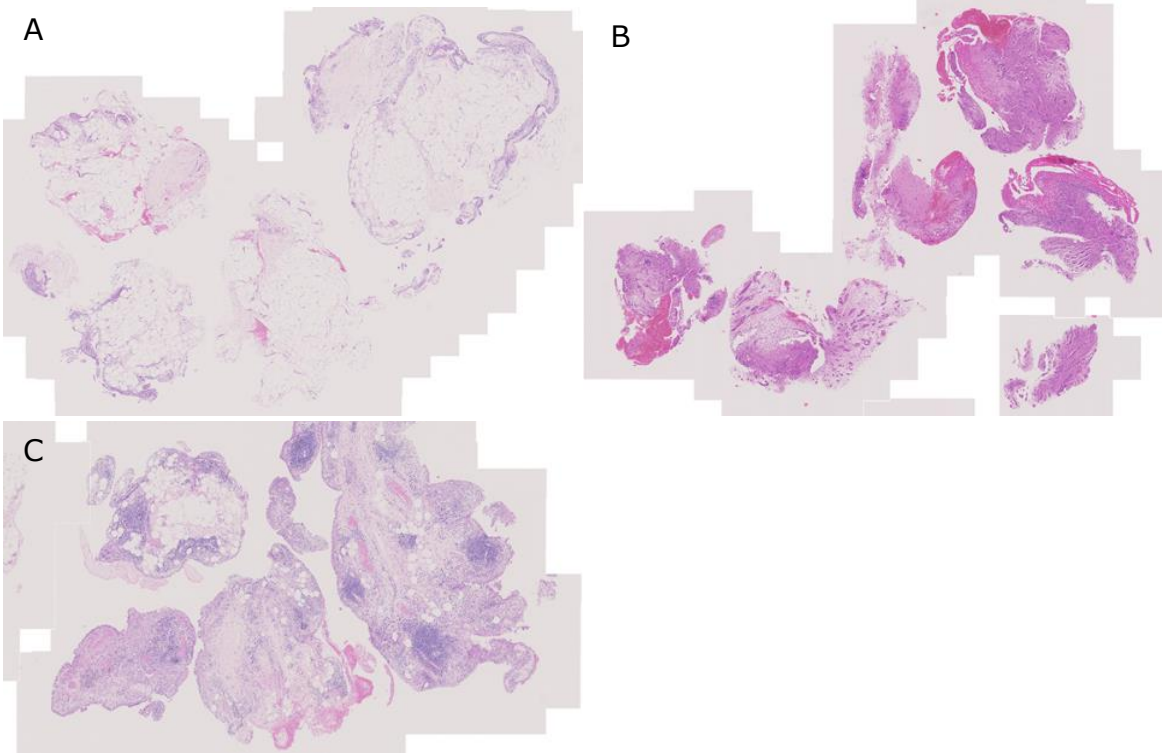


Figure 3.3: Representative images of pauci-immune, diffuse, and lymphoid tissues. Pathotypes were classified based on density and aggregate scores from the BEACON scoring system. (A) The pauci-immune tissue lacks immune infiltrate. (B) The diffuse tissue has immune infiltrate but it is not organised into aggregates. (C) The lymphoid tissue contains large aggregates. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

3.3 Reliability Testing

Two rounds of inter-reader reliability testing were undertaken sequentially. Firstly, an internal reliability test, conducted by those developing the scoring system at the University of Birmingham who had varying levels of experience

with synovial histology but who were not trained histopathologists, to assess whether those involved in developing the scoring system had good agreement. Secondly, an external reliability test, conducted by trained histopathologists from outside the University of Birmingham to assess whether external histopathologists had good agreement when using the BEACON scoring system.

Ten representative H&E images were independently scored for three characteristics, namely Krenn infiltrate, BEACON infiltrate density, and BEACON aggregate size, as shown in Figure 3.4. There was good internal agreement when using the BEACON scoring system, with mean density having an ICC of 0.938 and the aggregate grading having a slightly lower but still good Kappa coefficient of 0.769. BEACON density had a very similar ICC to the Krenn infiltrate score (0.938 and 0.937, respectively). The summary BEACON pathotype also showed good reliability, with a Kappa score of 0.722. This shows that internally the BEACON scoring system was reproducible across scorers, therefore external validation was sought to test reproducibility across other sites.

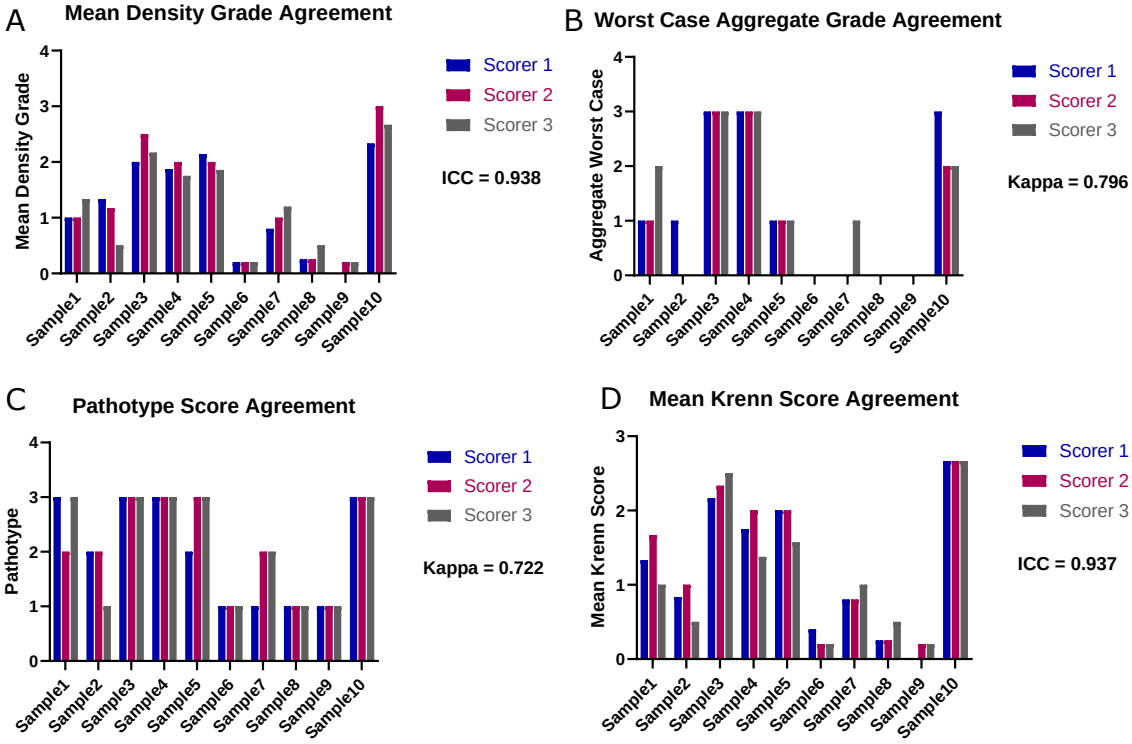


Figure 3.4: Internal reliability testing. 10 representative H&E samples were scored independently by 3 scorers (myself, Dr Andrew Filer and Prof Dagmar Scheel-Toellner). Each fragment within a sample was scored for Krenn infiltrate (D), BEACON density (A), and BEACON aggregates (B) and a mean or worst case score calculated before pathotype derivation (C). Pathotypes are shown as numbers, with one being pauci-immune, two being diffuse, and three being lymphoid. The Krenn score refers to the Krenn infiltrate score alone. Weighted Kappa or Intraclass Correlation Coefficients (ICC) are shown.

For the external reliability testing ten representative H&E images were independently scored for density and aggregates, this time by four histopathologists and members of the Accelerating Medicines Partnership (AMP) (https://www.nih.gov/research-training/accelerating-medicines-partnership-amp) consortium before a pathotype was calculated. The density score showed high reliability between scorers, with an ICC of 0.896 (Figure 3.5). However, there was lower agreement when grading aggregates, predominantly due to the

presence of outliers (Kappa = 0.735), resulting in lower agreement on the final pathotype score (Kappa = 0.571).

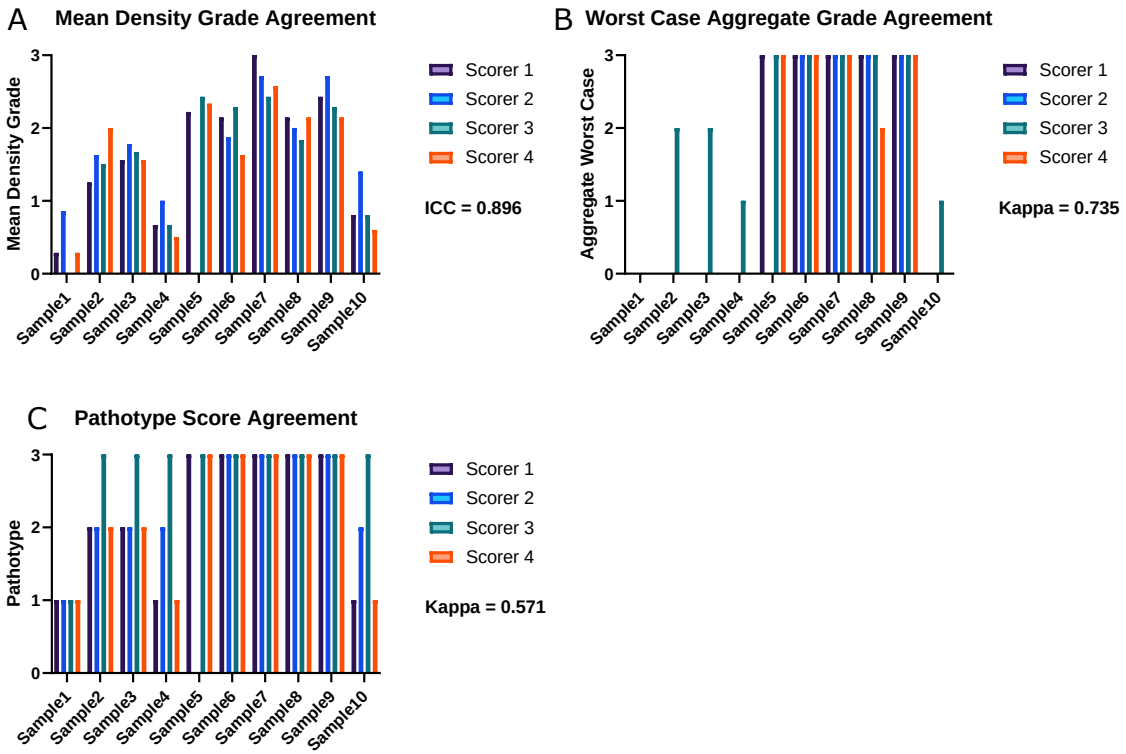


Figure 3.5: External reliability testing. 10 representative images were independently scored by four individuals in AMP (Prof Brendan Boyce, Prof Ellen Gravallese, Prof Edward DiCarlo, and Prof Dagmar Scheel-Toellner) for density (A) and aggregates (B) and a pathotype (C) was derived from this. Pathotypes are shown as numbers, with one being pauci-immune, two being diffuse, and three being lymphoid. Weighted Kappa or ICC are shown. Sample 5 was excluded from ICC and Kappa calculations due to missing data.

As there was disagreement in the aggregate grading between scorers, the exercise was repeated following a consensus exercise and refinement of the atlas. Figure 3.6 shows the resulting agreement of the aggregate and pathotype scores. The aggregate regrading showed good agreement, with a Kappa score of 0.862, which resulted in improved agreement for the pathotype score as well, with a Kappa of 0.754.

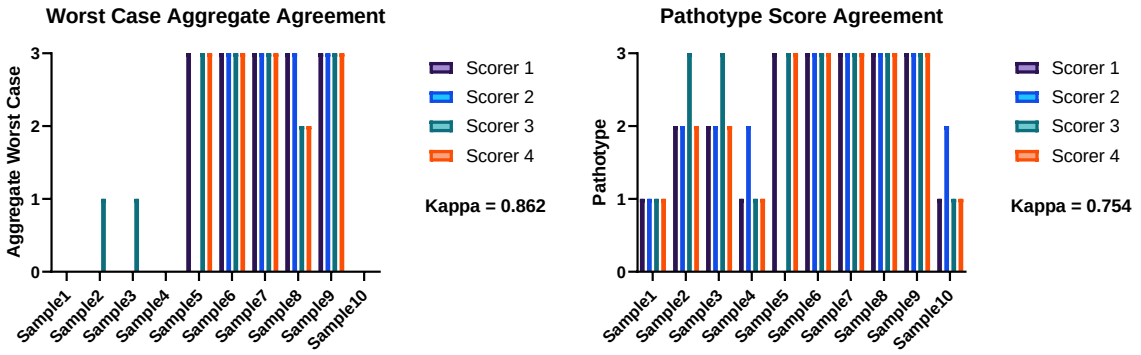


Figure 3.6: External reliability testing: aggregate regrade. Aggregates (left) were regraded by the same four scorers and a pathotype score (right) was calculated. Weighted Kappa coefficients are shown. Sample 5 was excluded from ICC and Kappa calculations due to missing data.

Degree of correlation between the density, aggregate, and Krenn infiltrate score was assessed. The absence of aggregates (worst-case aggregate grade of 0) was associated with a lower mean density grade and Krenn infiltrate score when compared to any size of aggregates ($p<0.0001$). However, there was no significant difference between the mean density grades across different aggregate sizes (Figure 3.7). There was a strong correlation between BEACON density and Krenn infiltrate scores ($r=0.92$, $p<0.0001$).

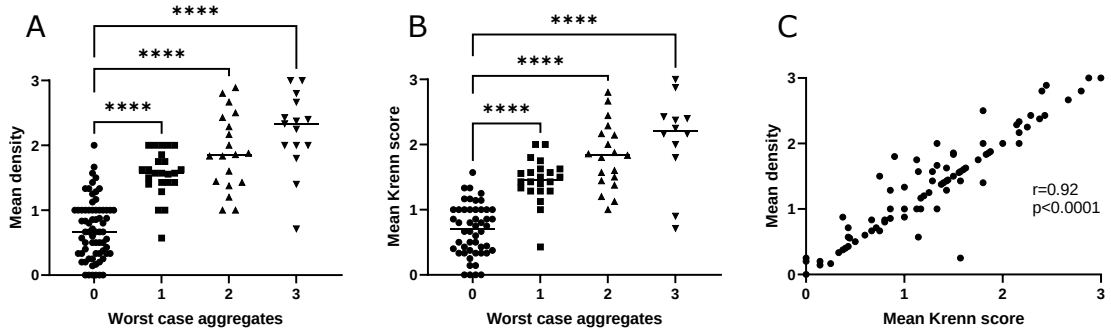


Figure 3.7: Relationship between BEACON density and aggregate grades, and the Krenn infiltrate score. (A) BEACON density vs aggregates, (B) Krenn infiltrate vs BEACON aggregates, (C) BEACON density vs Krenn infiltrate. Krenn score refers to the Krenn inflammatory infiltrate score alone. Bars show median values. Kruskal-Wallis with post-hoc Dunn's multiple comparison test; ****p<0.0001. Spearman rank used to assess correlation.

3.4 BEACON HISTOLOGY SCORING AND CLINICAL VARIABLES

Following development and validation, a larger cohort of samples (n=127) were then scored for density, aggregates and pathotype and compared to clinical variables, including global inflammatory markers, disease activity, and ultrasound variables, to test whether the BEACON scoring system may have clinical utility. Table 3.1 shows a summary of patient characteristics for this cohort and full details can be found in Appendix 9.1.

	Resolving	NonRA	sdRA	ldRA	Total
Number	21	15	28	63	127
Female (n)	9 (43%)	8 (53%)	15 (54%)	37 (59%)	69 (54%)
Age	47	45	58	56	54
DAS28-ESR	3.9	4.6	5.3	5.6	5.1
US GS	1.4	1.6	2.0	2.1	1.9
US PD	0.7	1.1	1.0	1.2	1.1

Table 3.1: Summary of patient characteristics. Values show mean unless otherwise stated. DAS28-ESR = disease activity score, ESR = erythrocyte sedimentation rate, US GS = ultrasound greyscale, US PD = ultrasound power Doppler.

Although there might be a tendency towards fewer pauci-immune pathotypes in ldRA, there was no significant association between pathotype and clinical group (Figure 3.8). There was also no association for the density and aggregate grades (Figure 3.9 and Figure 3.10). However, pathotype was significantly associated with ultrasound grey scale (US GS) ($p=0.001$), a measure of synovial hypertrophy, and the global inflammatory marker CRP ($p<0.001$). Interestingly, although there was no difference between either DAS28-CRP or DAS28-ESR scores across the pathotypes, there was higher patient reported joint pain, assessed using a visual analogue scale (VAS), associated with the diffuse pathotype than pauci-immune ($p=0.003$), and higher physician global assessment in both the diffuse and lymphoid pathotypes when compared with pauci-immune ($p=0.01$ and 0.04, respectively, Figure 3.8). There were insufficient numbers to assess DAS28-ESR response at 12 months statistically, due to data only being available for a subset of RA patients, however there was a trend towards increased lymphoid pathotype in responders, with the lymphoid pathotype making up 69% of responders compared to 23% diffuse and 8% pauci-immune, and increased pauci-immune in non-responders, with 50% being pauci-immune, 33% diffuse, and 17% lymphoid.

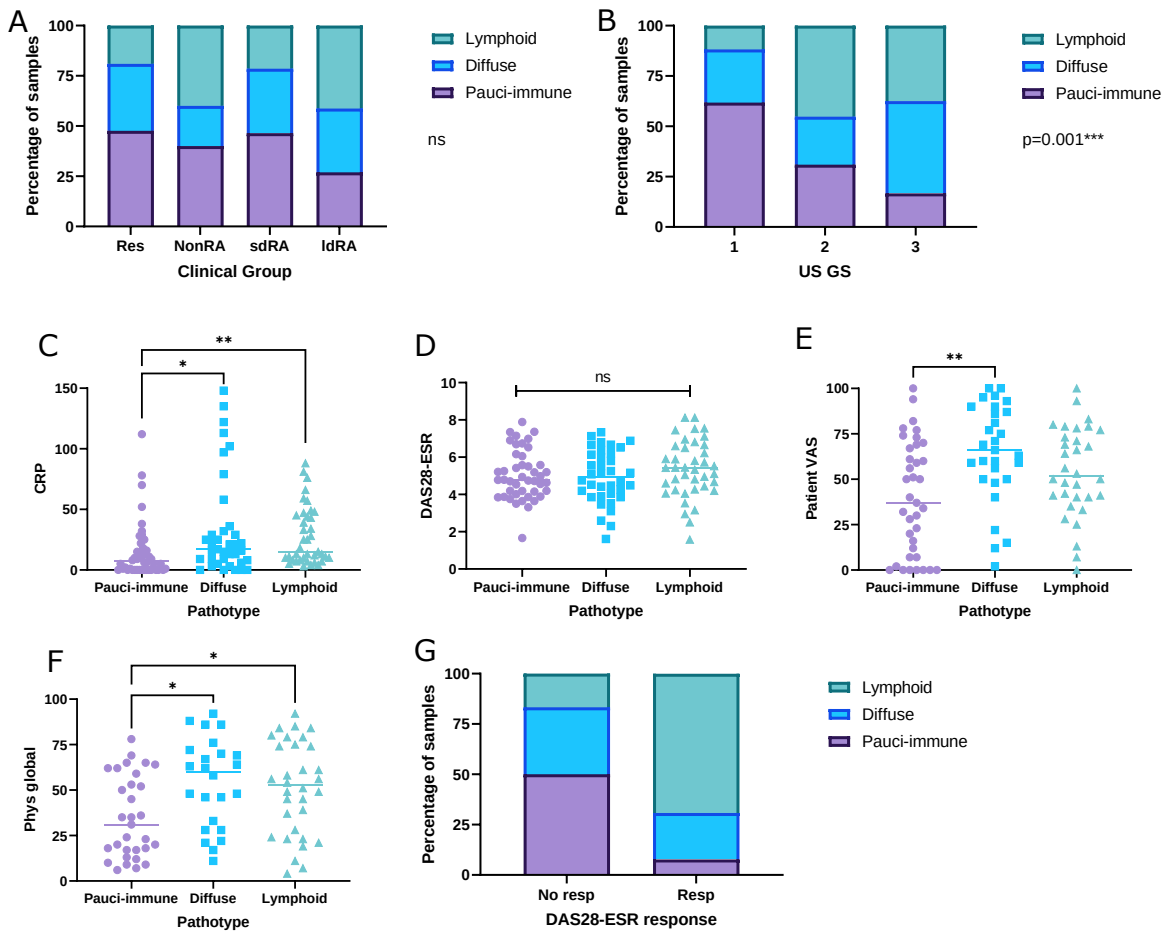


Figure 3.8: Relationship between pathotype and clinical variables. Clinical group (A, n=127), US GS (B, n=100), and DAS28-ESR response (G, n=19) show percentages. For CRP (C), DAS28-ESR (D), patient VAS (E), and physician global (F) bars show median values. Chi squared or Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; *p<0.05, **p<0.01, ***p≤0.001. Patient VAS = patient joint pain visual analogue scale, phys global = physician global assessment, DAS28-ESR response = EULAR response assessed at 12 months, response includes good and moderate.

Examining the individual grades, density correlated with a number of clinical variables, including US GS, CRP, patient VAS, and physician global assessment (p=0.0004; r=0.41, p<0.0001; r=0.33, p=0.0009; r=0.31, p=0.004, Figure 3.9). Significance associated with US PD was lost following Benjamini-Hochberg correction for multiple hypothesis testing.

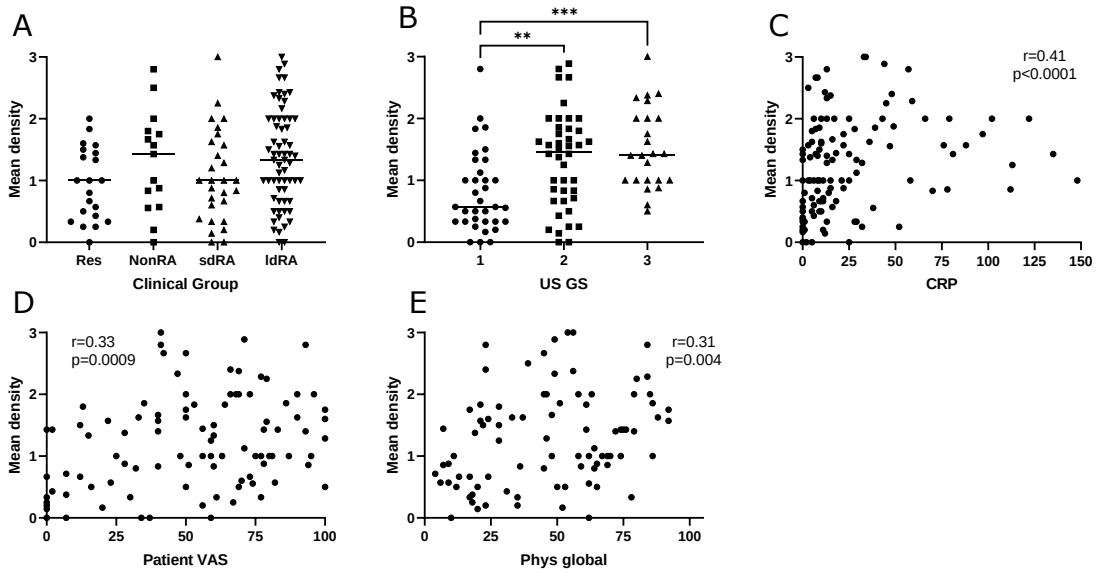


Figure 3.9: Relationship between BEACON density and clinical variables. (A) clinical group, (B) US GS, (C) CRP, (D) patient VAS, (E) physicial global. Bars show median values. Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; **p<0.01, ***p≤0.001. Spearman rank used to assess correlation. Patient VAS = patient joint pain visual analogue scale, phys global = physician global assessment.

Aggregate worst case, as used in the calculation of pathotype, had insufficient numbers for contingency statistics so mean aggregate grade has been used to assess association with clinical variables. Mean aggregate grade had a significant but weak correlation with CRP (r=0.28, p=0.002), although from the scatter plot this does not seem to be a simple linear relationship, and was associated with US GS (p=0.002), with lower US GS being associated with lower aggregate grade (Figure 3.10). Higher aggregate grade was associated with DAS28-ESR response at 12 months in the subset of RA patients with available data (p=0.03), however this significance was lost following Benjamini-Hochberg correction.

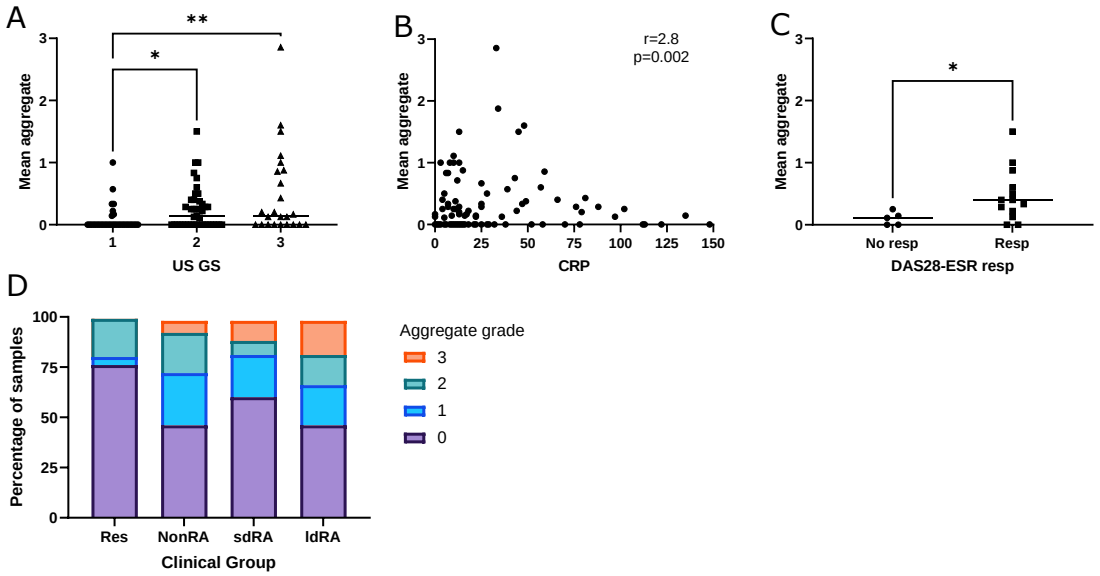


Figure 3.10: Relationship between BEACON aggregate and clinical variables. (A) US GS, (B) CRP (C) DAS28-ESR response, (D) clinical group. Bars show median values. Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; *p<0.05, **p<0.01. Spearman rank used to assess correlation. DAS28-ESR response was not significant following Benjamini-Hochberg correction. DAS28-ESR response = EULAR response assessed at 12 months, response includes good and moderate responders.

3.5 External validation of the BEACON scoring system

To explore the utility of the BEACON scoring system in an external cohort, pathotype, density, and aggregate grading was conducted on the AMP RA Phase II cohort by Prof Ellen Gravallese, Prof Brendan Boyce, and Prof Edward DiCarlo, with consensus scores being used for final classification of each sample (n=103). Figure 3.11 shows the relationship between pathotype and clinical variables. There were insufficient numbers for statistical testing of association with US GS, US PD, and DAS28-ESR response at 6 months. The only variable that was significantly associated with pathotype was CRP (p=0.04), with CRP being higher in the lymphoid pathotype than pauci-immune. However, this significance was lost following Benjamini-Hochberg correction for multiple comparisons.

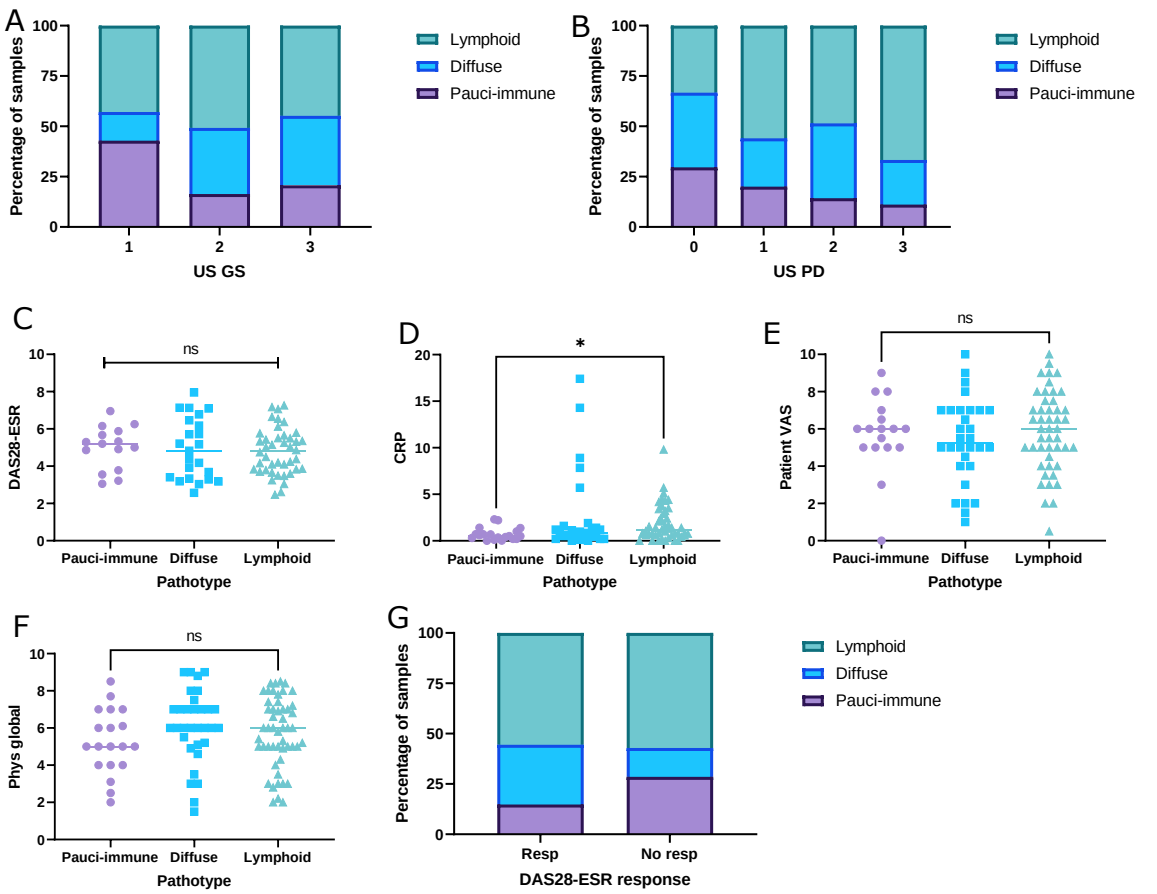


Figure 3.11: Relationship between pathotype and clinical variables in the AMP RA Phase II cohort. US GS (A, n=97), US PD (B, n=96), and DAS28-ESR response (G, n=41) show percentages. For DAS28-ESR (C), CRP (D), patient VAS (E), and physician global (F) bars show median values. Chi squared or Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; *p<0.05. Patient VAS = patient joint pain visual analogue scale, phys global = physician global assessment, DAS28-ESR response = EULAR response assessed at 6 months, response includes good and moderate responders.

BEACON density associated with US PD (p=0.01) and CRP (r=0.35, p=0.0005) in the AMP cohort but not with US GS, patient VAS, or physician global, which were associated with density in the BEACON cohort (Figure 3.12). BEACON aggregate grading was not significantly associated with any clinical variables in the AMP cohort, although there were insufficient numbers to test US GS, US PD, and DAS28-ESR response at 6 months (Figure 3.13).

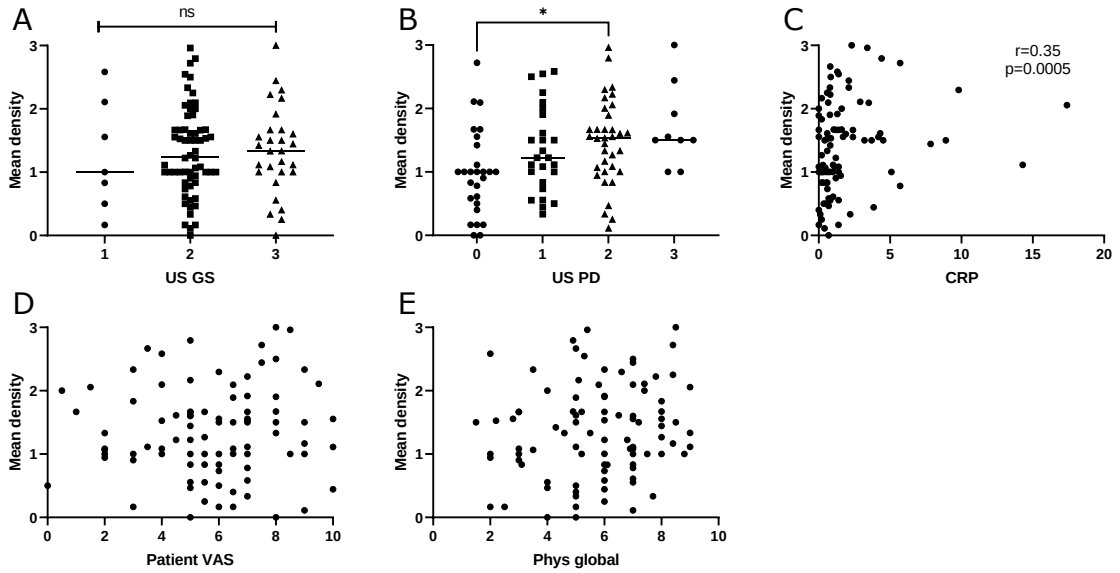


Figure 3.12: Relationship between BEACON density and clinical variables in the AMP RA Phase II cohort. (A) US GS, (B) UD PD, (C) CRP, (D) patient VAS, (D) physicial global. Bars show median values. Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; *p<0.05. Spearman rank used to assess correlation. Patient VAS = patient joint pain visual analogue scale, phys global = physician global assessment.

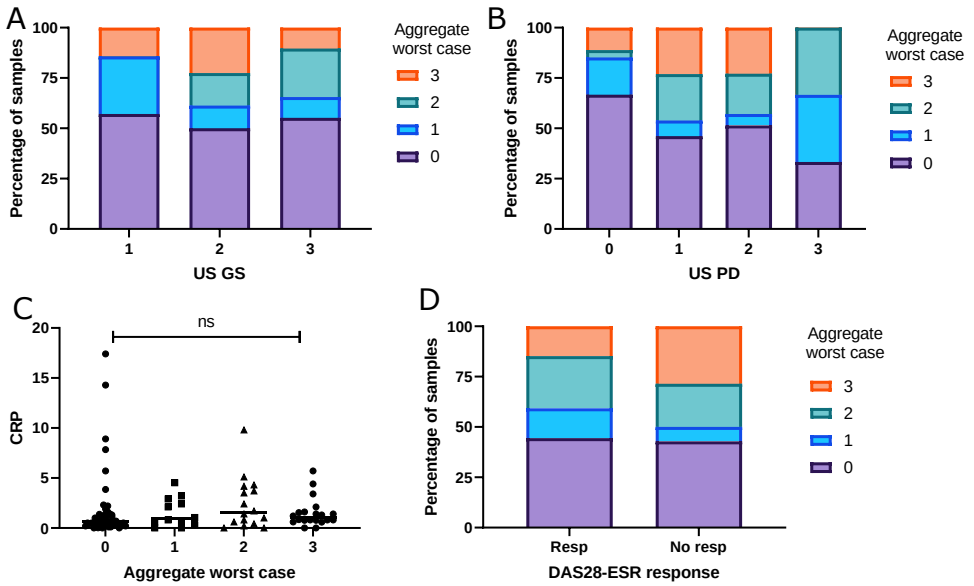


Figure 3.13: Relationship between BEACON aggregate and clinical variables in the AMP RA Phase II cohort. US GS (A, n=98), US PD (B, n=97), and DAS28-ESR response (D, n=41) show percentages. CRP (C) bars show median values. Chi squared or Kruskal-Wallis with post-hoc Dunn's multiple comparison tests. DAS28-ESR response = EULAR response assessed at 6 months, response includes good and moderate responders.

3.6 AI-ASSISTED SCORING

Although the BEACON scoring system showed good reliability between scorers, it was still subject to some differences and could be influenced by human error or bias. Furthermore, manual scoring is time-intensive, especially when there are large numbers of tissues that require scoring and has limitations in terms of the level of accuracy and detail that can be obtained when calculating areas or cell numbers. In order to improve on these issues, AI-assisted scoring that allows for automatic detection of features of the synovium was explored.

Initial proof-of-concept APPs (Analysis Protocol Packages) were developed by myself to identify tissue, lining layer, and aggregates, before calculating tissue

area, area and mean thickness of the lining layer, and aggregate area. Figure 3.14 shows the resulting images with associated results.

These APPs were then further developed by Dr Nasullah Khalid Alham alongside new APPs for the detection of lymphocytes and plasma cells, with manual annotation used for training being agreed on with consensus by Prof Clare Verrill, Dr Andrew Filer, and Prof Dagmar Scheel-Toellner. Figure 3.15 shows example images of the detection of tissue, lining layer, and aggregates and Figure 3.16 shows the detection of lymphocytes and plasma cells.

Validation of the lining layer and aggregate detection APPs was completed by Prof Clare Verrill, Dr Andrew Filer, and Prof Dagmar Scheel-Toellner. Validation of lymphocyte and plasma cell detection APPs is ongoing. 30 out of 41 (73%) annotations used as validation for lining layer detection were identified as being fully correct by all three reviewers, with a further 8 (93%) being classified as either correct or partially correct by all reviewers. Of the three remaining annotations, one was classed as incorrect by one reviewer and correct by the other two, one was difficult to classify, with reviews classifying it as 'incorrect', 'not sure', and 'partially correct', and the final one was identified as incorrect by two reviewers and partially correct by the final reviewer. There were therefore no annotations that all three reviewers identified as incorrect and only a single annotation (2%) that two of the three reviewers identified as incorrect.

37 annotations were used as validation of the aggregate detection APP, of which 24 (65%) were identified as being correct by all reviewers and 33 (89%) were identified as being either correct or partially correct by all reviewers. Of the remaining four annotations, one was classified as 'not sure' by one reviewer and

correct and partially correct by the other two, two failed to reach consensus among reviewers, being identified as correct, partially correct, and incorrect, and the final annotation was identified as partially correct by two reviewers and incorrect by the final reviewer. Overall, validation of the lining layer and aggregate detection APPs show that they have good performance and that most annotations are deemed fully correct by at least two reviewers (90% and 86%), with no annotations for either APP being deemed incorrect by all reviewers.

Once all APPs are validated, the whole pipeline will be run on tissues from the BEACON cohort and results will be compared to the manual BEACON scoring system and clinical variables. The use of this pipeline will also be tested in another cohort (the AMP RA cohort) to validate its utility outside of the BEACON cohort.

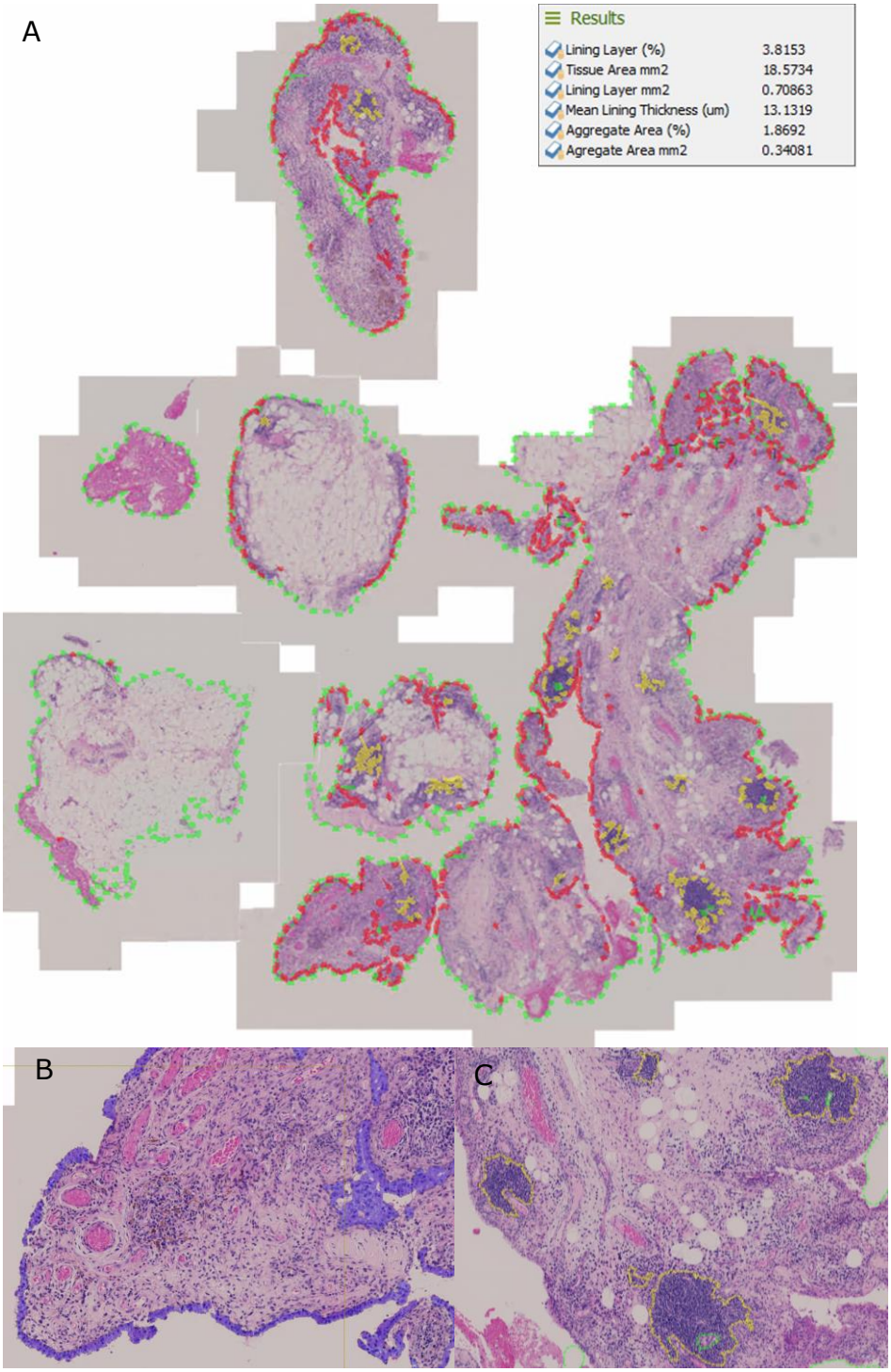


Figure 3.14: Tissue, lining layer and aggregate detection using Visiopharm. Features were detected by AI APPs trained on manual annotations. Post-processing was used to calculate the area of the tissue and features. (A) The whole process, with tissue detection (green), lining layer detection (red) and aggregate detection (yellow). (B) Detection of the lining layer in blue. (C) Detection of aggregates in yellow. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

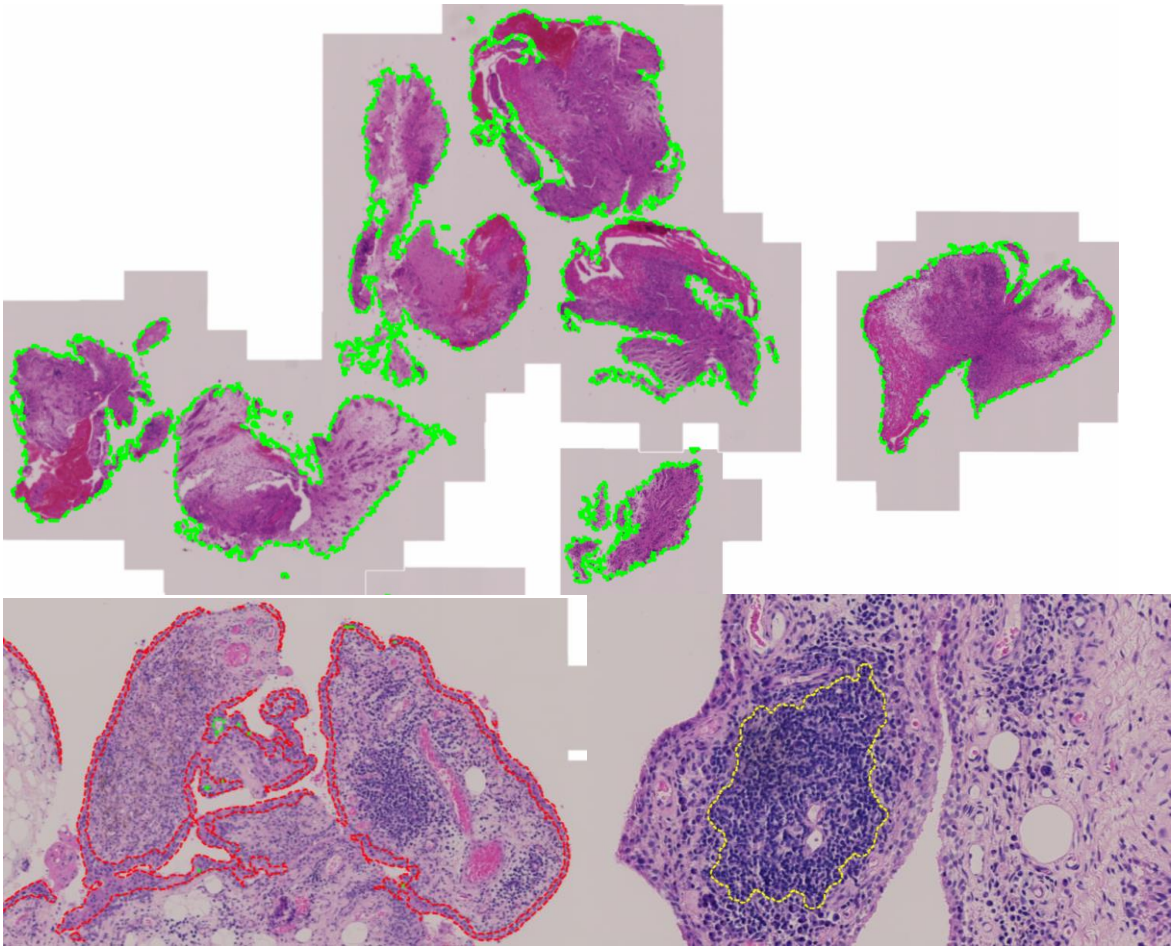


Figure 3.15: Detection of tissue (green), lining layer (red), and aggregates (yellow) using VIS. APPs were developed by Dr Nasullah Khalid Alham, by building upon APPs developed by myself in a proof-of-concept trial. Annotation for training of APPs and validation was carried out by Prof Clare Verrill, Dr Andrew Filer, and Prof Dagmar Scheel-Toellner. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

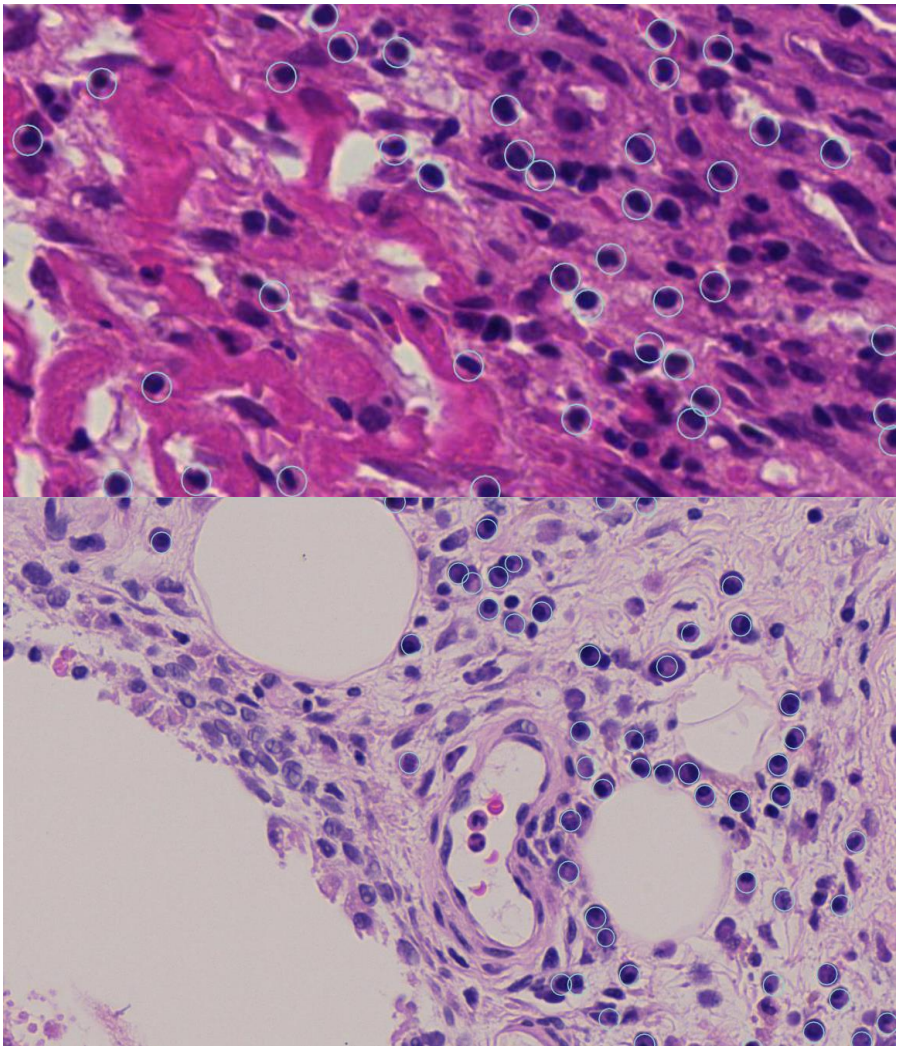


Figure 3.16: Detection of lymphocytes (above) and plasma cells (below) using VIS. APPs were developed by Dr Nasullah Khalid Alham with annotation for training and validation being carried out by Prof Clare Verrill, Dr Andrew Filer, and Prof Dagmar Scheel-Toellner. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

3.7 Discussion

The BEACON scoring system represents an alternative to other currently used synovial tissue scoring systems, allowing for the stratification of early inflammatory arthritis patients into pathotypes according to level and degree of organisation of the lymphocytic infiltrate using H&E stained tissue.

The BEACON scoring system showed good reliability between independent scorers both internally and externally, as shown by the relatively high Kappa and ICC scores achieved. This shows promise that the scoring system is reproducible across sites and the associated atlas should help to ensure reproducibility in the wider research community.

As a part of the validation, BEACON density and aggregate grades were compared with the Krenn inflammatory infiltrate score (Krenn et al., 2006). There was a very strong correlation between the density and infiltrate scores. This is likely because both scores are capturing the degree of infiltration of immune cells. However, the aggregate grade did not exhibit the same level of agreement with the other two scores. In the absence of aggregates there was a lower density score when compared to any other aggregate grade, however there was no difference in the density grade across different sized aggregates. Unsurprisingly, given the degree of similarity, this was also true for the Krenn infiltrate score. A plausible explanation for this is that the aggregate score is capturing information that the Krenn infiltrate score does not around the degree of organisation of the inflammatory infiltrate. However, this could also be due to the smaller number of samples that have high aggregate grades reducing the statistical power associated with these comparisons, as the majority of samples (54%) had an aggregate score of zero.

Higher BEACON density was associated with increased synovial hypertrophy (US GS) and systemic inflammation (CRP), and weakly with physician global assessment and patient joint pain (VAS) in the BEACON cohort. This may be due to the lymphocytic infiltrate enhancing production of pro-inflammatory mediators both directly and indirectly, resulting in increased synovial inflammation and

hypertrophy, as well as increased systemic inflammatory markers. One such mechanism is via the production of systemic IL-6, predominantly by macrophages and fibroblasts, which drives CRP production. Lymphocytes induce expression of IL-6 via the production of other pro-inflammatory cytokines, such as IL-17 produced by Th17 cells (Noack and Miossec, 2017). Alternatively, it may be that production of pro-inflammatory mediators, including chemokines, by activated synovial fibroblasts drives both increased CRP production and immune cell infiltration. However, activated synovial fibroblasts are likely to be present in all inflamed synovial tissue, with the main histological difference between low and high density tissues being the presence and intensity of the lymphocytic infiltrate. From the current data, this suggests that lymphocytes contribute to the increased production of CRP, even if they do not initiate it, although notably this does not take into account the presence of different fibroblast subtypes that can drive different pathological processes, which are not possible to identify using H&E staining alone (Croft et al., 2019). Aggregate grade was also associated with US GS and CRP, however the correlation with CRP was weak so may not be of biological relevance.

BEACON pathotype was not different across clinical groups, meaning it would not have utility as a diagnostic tool. This was also the case for the density and aggregate scores. This suggests that, although there is a large degree of heterogeneity, that heterogeneity in tissue structure is seen across different inflammatory arthritides. This may allow for the study of common mechanisms across diseases, potentially leading to the development of treatments for generic processes involved in inflammatory arthritis, regardless of diagnosis. This has already been shown to be possible, for example in the case of anti-TNF therapy,

which targets a common inflammatory cytokine and is effective in many inflammatory arthropathies (Bradley, 2008).

From this study we cannot confirm whether pathotype remains constant throughout disease course or is associated with different disease stages. However, the similar proportions of pathotypes in sdRA and ldRA suggest that duration of disease may not be the only explanation. It has been shown that pathotype can change following 6-month DMARD treatment, with transition to a less inflammatory pathotype being associated with response to treatment (Lewis et al., 2019). However, whether pathotype remains constant in the absence of treatment remains unknown.

Pauci-immune pathotype was associated with lower CRP and physician global assessment compared to diffuse and lymphoid pathotypes, and lower patient VAS compared with the diffuse pathotype in the BEACON cohort. There was also a greater proportion of pauci-immune pathotype in patients with low US GS. Overall, this may suggest a less inflammatory, milder disease in the pauci-immune group but there was no association with reduced disease activity (DAS28). Furthermore, pauci-immune pathotypes have previously been shown to be less responsive to treatments, such as TNFα-blockade (Nerviani et al., 2020), which may result in poorer long-term outcomes. This could be because distinct disease mechanisms are involved in different pathotypes, which may mean that they require different targets for successful treatment.

Pathotype, as described in previous studies, has been shown to have some utility in predicting response to treatment in RA, with more inflammatory phenotypes generally resulting in better response (Humby et al., 2019b, Dennis et al., 2014,

Lewis et al., 2019, Nerviani et al., 2020). In this study, there was a trend towards increased lymphoid pathotype in responders and increased pauci-immune pathotype in non-responders, determined using EULAR DAS28-ESR response at 12 months (van Gestel et al., 1996). However, insufficient numbers were available to assign statistical significance. Higher mean aggregate score was also associated with DAS28-ESR response, although significance was lost following correction for multiple comparisons. These data show similar trends to previous studies, including those that relied on staining beyond H&E (Nerviani et al., 2020, Lewis et al., 2019, Humby et al., 2019b). Differences in response of pathotypes could be due to the target mechanisms of treatments for inflammatory arthritis. Many treatments for rheumatoid arthritis currently target immune and inflammatory processes, so it may be that currently we are not targeting the mechanisms driving pauci-immune disease subtypes. Further study of this pathotype to identify new treatment targets may improve response rates in the future. As disease is likely less driven by immune infiltrate in this patient subset, current research efforts into synovial fibroblasts may also find new targets that prove to be more effective (Siebert et al., 2020, Croft et al., 2019, Montero-Melendez et al., 2020, Diller et al., 2019).

Validation of the BEACON scoring system was sought in an external cohort, namely the AMP RA Phase II cohort. In agreement with the BEACON cohort, both pathotype and density were associated with CRP, with lymphoid pathotypes having higher CRP than pauci-immune and density positively correlating with CRP. However, many of the associations between pathotype, density, and aggregate scores and clinical variables seen in the BEACON cohort were not seen in the AMP cohort. This was partly be due to insufficient numbers largely

preventing the use of contingency statistics in the case of pathotype and aggregate scoring, with data on mean aggregate grading not being available to overcome this, as was used for the BEACON cohort. Indeed, some similar patterns were seen, with pauci-immune pathotype tending to be higher in low US GS scores, and increased presence of aggregate grade 0 in patients with low US PD scores. However, there were some associations that did not seem to be present in the AMP cohort despite this, including patient VAS and physician global that showed no relationship with pathotype or density.

These differences are likely due to the difference in cohort composition, as BEACON is an early inflammatory arthritis cohort with a range of disease activities whilst the AMP cohort is a mixed RA cohort, including patients at different disease stages, but all with high disease activity. The BEACON scoring system was developed for use on early disease, so may not capture features of more advanced, late stage disease. Furthermore, the AMP cohort includes patients who have undergone treatment, which has been shown to have an impact on pathotype in previous studies (Lewis et al., 2019). Therefore, it may be beneficial to seek validation in another early inflammatory arthritis cohort, to confirm utility in these patient groups.

In summary, the BEACON scoring system represents an alternative to previously developed histology scoring systems for inflammatory arthritis. It allows for pathotype derivation from H&E stained tissue and captures the level and degree of organisation of the lymphocytic infiltrate. This tool shows similar associations to other previously developed systems without requiring further staining or extensive expertise in synovial histology.

Development of an AI-assisted scoring pipeline using VIS is currently underway, with APPs for the detection of tissue, lining layer, and aggregates having been developed and validated and APPs for lymphocyte and plasma cell detection developed and undergoing validation. This should allow for the detection of features of the synovium without human bias and allow for more reproducible results when compared to more subjective semi-quantitative scoring systems.

Validation of the lining layer and aggregate APPs showed good performance but also highlighted the challenges around reaching a consensus between reviewers. This is particularly true in the case of the aggregate score, for which two annotations were marked as correct, partially correct, and incorrect by different reviewers. This makes it complicated to determine ground truth for APP development but also highlights the difference in opinion between scorers, which would be overcome when using automatic detection of these features compared with manual scoring, resulting in more reproducible results.

Unfortunately, the pipeline for automatic detection of features has not been fully validated or run on the whole cohort, meaning there is not yet data available to investigate how this correlates with the manual BEACON scoring system or clinical variables. Upon final validation of the APPs, these questions will be explored in the BEACON cohort, and validation will be sought in an external cohort.

Overall, the manual BEACON scoring system provides a simple system that is easy to use and has good reliability between scorers. It also correlates with measures of local and systemic inflammation so has clinical relevance for patient stratification. An AI-assisted analysis pipeline is still in development but shows

promise for removing human bias and increasing the reproducibility of histology scoring in the future, while also allowing for the calculation of quantitative measures of features of the synovium.

4 RNA SEQUENCING: SOURCES OF VARIATION

4.1 INTRODUCTION

Given the inability to distinguish clinical groups from histological scoring alone, whole tissue RNA sequencing was undertaken to explore the similarities and differences in gene expression across the inflammatory arthritis cohort. The aim of this work was to explore pathogenic mechanisms in more detail and see how this related to the histological variables explored in the previous chapter.

Similar studies comparing histology with gene expression have been undertaken in RA alone, generally finding subgroups of RA patients with different levels of immune infiltrate and varying dominant cell types (Orange et al., 2018, Lewis et al., 2019). However, a comprehensive analysis of the heterogeneity across early inflammatory arthritides had not yet been undertaken, and should allow for the identification of common mechanisms that will increase our understanding of the processes driving inflammation in the joint, as well as disease-specific mechanisms, which may identify new therapeutic targets.

This chapter will describe the major sources of variation within the RNA sequencing dataset and whether they are associated with any clinical or histological variables, while also undertaking quality control (QC) and exploring confounding factors that may need to be corrected for in future analyses.

4.2 PATIENT CHARACTERISTICS

A summary of patient characteristics for the final cohort used in the analyses can be found in Table 4.1 and full details can be found in Appendix 9.1.

	Normal	Resolving	NonRA	sdRA	ldRA	Total
Number	5	14	13	18	35	85
Female (n)	3 (60%)	5 (36%)	7 (54%)	9 (50%)	20 (57%)	44 (52%)
Age	38	48	45	58	59	54
DAS28-ESR	NA	3.9	4.4	5.4	5.7	5.1
US GS	NA	1.4	1.7	1.9	2.2	1.9
US PD	NA	0.7	1.2	1	1.2	1.1

Table 4.1: Summary of patient characteristics. Values show mean unless otherwise stated. DAS28 = disease activity score, ESR = erythrocyte sedimentation rate, US GS = ultrasound greyscale, US PD = ultrasound power Doppler.

In order to confirm the disease trajectory of the resolving patients when biopsied and as this was not routinely recorded, Prof Karim Raza and Dr Andrew Filer explored patients' clinical notes and blood test results to assess whether they were at a stage of worsening or resolving inflammation at the time the samples were taken. Patients were categorised as 'definite improving phase', 'likely improving phase', 'worsening phase', or 'unknown'. Of the 14 resolving patients, only one was in a worsening phase, eight were in a definite improving phase, three were likely improving, and two were unclear from their notes and so classified as unknown. Therefore, the majority (79%) of the resolving patients were most likely in a phase of actively resolving inflammation at the time of biopsy.

The normal samples were all uninflamed macroscopically, histologically, and on imaging when classified as normal prior to the commencement of this study. This was confirmed histologically on H&E stained images for the three normal samples that had available images in the present study (Figure 4.1).

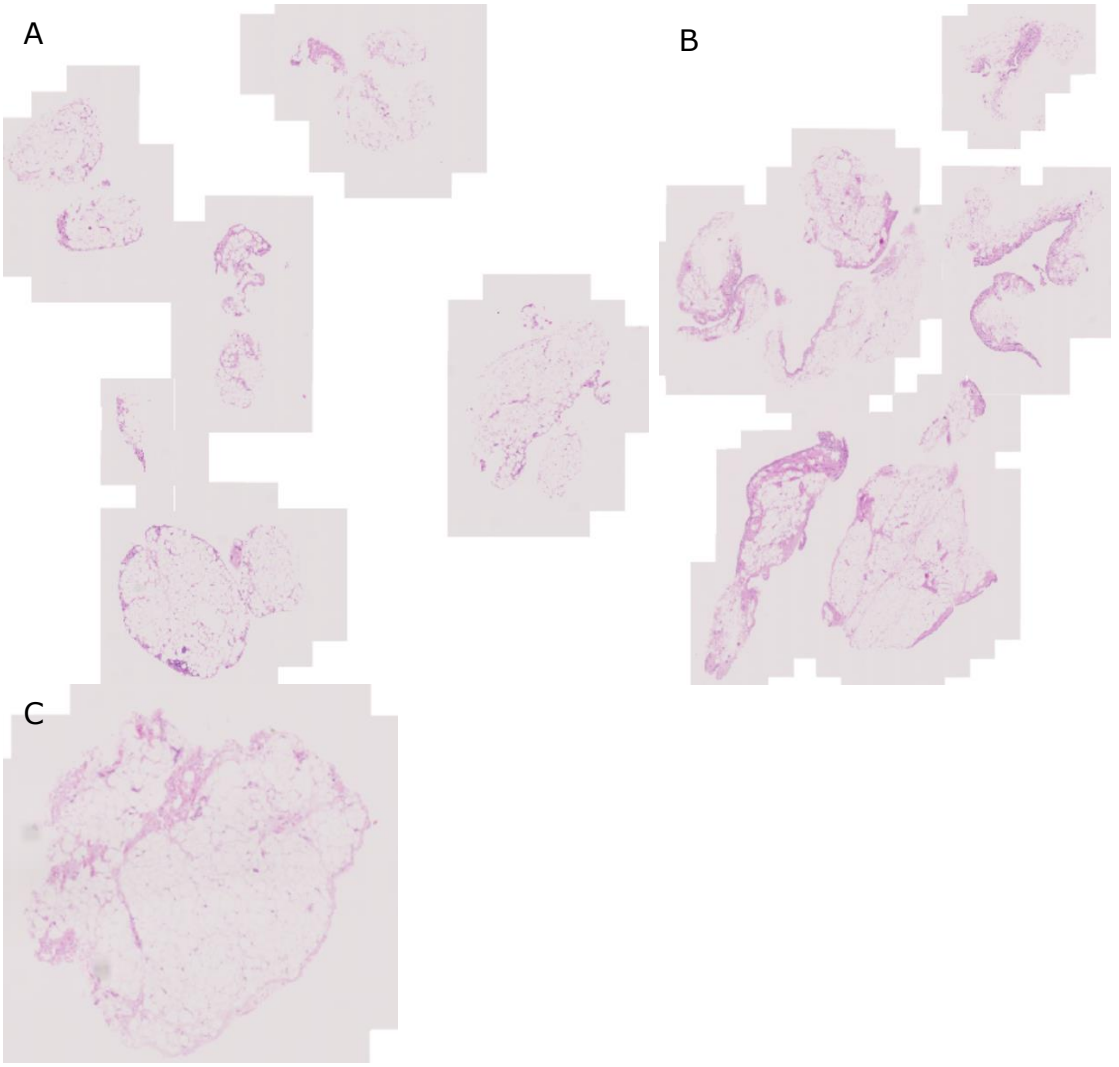


Figure 4.1: H&E stained images of the normal control samples. A-C are H&E images of the three of the normal samples that had images available in the present study. These show normal, uninflamed synovial tissue consisting predominantly of adipocytes. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

4.3 QC & EXPLORATION OF CONFOUNDERS

Exploration of the most variable genes in the RNA sequencing dataset was conducted to identify outlying samples. One sample was identified as being an outlier initially, having high expression of a number of genes that were not expressed in other samples. These genes were identified as being highly

expressed in skin (KRT1, KRT2, KRT5, KRT14, DSP, FLG, and FLG2), suggesting some skin contamination during the biopsy procedure. This sample was therefore removed from future analyses to prevent confounding of the results. 13 samples also showed exclusive high expression of a number of genes (Figure 4.2), which are highly expressed in skeletal muscle. The same samples also separated by principal component analysis (PCA) on principal components (PCs) one and two (Figure 4.3). These samples were removed from analyses at this stage, however will be discussed further later in the chapter.

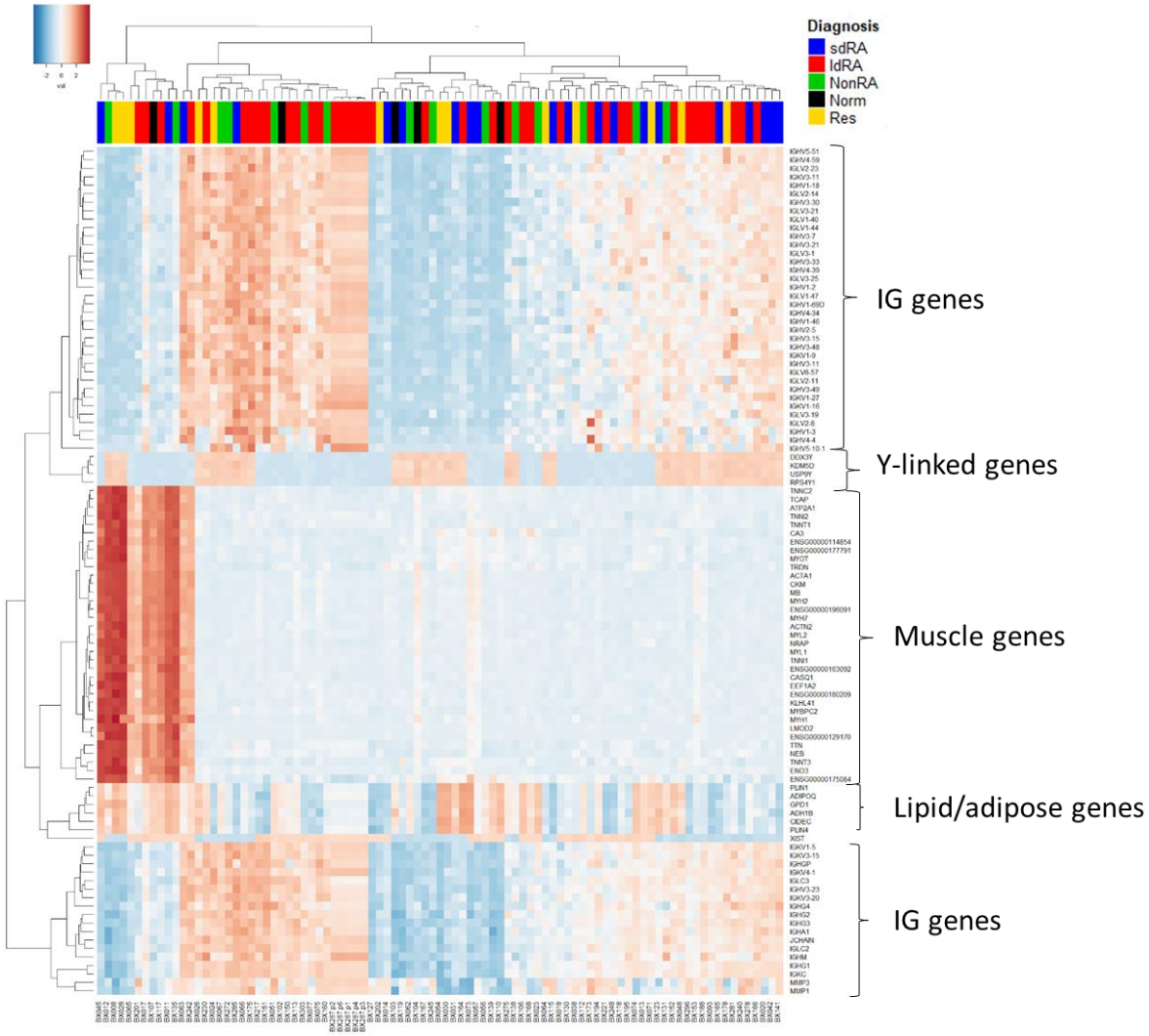


Figure 4.2: Heatmap of the top 100 most variable genes. Top 100 most variable genes are shown with hierarchical clustering on columns and rows. Bar along the top shows clinical group. Labels capture general information about the genes. IG = immunoglobulin.

When exploring the top most variable genes and PCA, it became apparent that a large amount of variation was attributed to patient sex, with Y-linked genes, XIST and TSIX coming up as highly variable and samples being split by sex in PC4 (Figure 4.2 and Figure 4.3). After removal of the Y chromosome, XIST and TSIX, sex was no longer associated with any PC and no sex-related genes came

up as being highly variable (Figure 4.5 and Figure 4.6). These genes were therefore removed from future analyses.

Another large source of variation was immunoglobulin variable (IGV) genes (Figure 4.2). This was likely an overrepresentation of their prevalence as IGV genes, that is V-, D-, and J-gene segments, are all counted individually, resulting in a large number of genes associated with IGs. IGV genes were removed from this and future analyses, with IG constant genes retained to ensure that the overall IG signal was not lost.

To ensure confounders were corrected for in the RNA sequencing differential expression design, the impact of potential confounding and important biological factors on the variation of the overall RNA sequencing data was explored. Linear modelling and stepwise regression were used to test the influence of both biological and confounding factors on varying forms of the data, including most variable genes and PCs (Table 4.2). Variables that came up most frequently in the different tests were RIN, DAS28-ESR, US GS, and a plasma cell score. The plasma cell score was a histology-based score that was graded as described in the study by Orange et al. (2018). Of these, the only non-biological variable, and therefore potential confounding factor, was the RIN score.

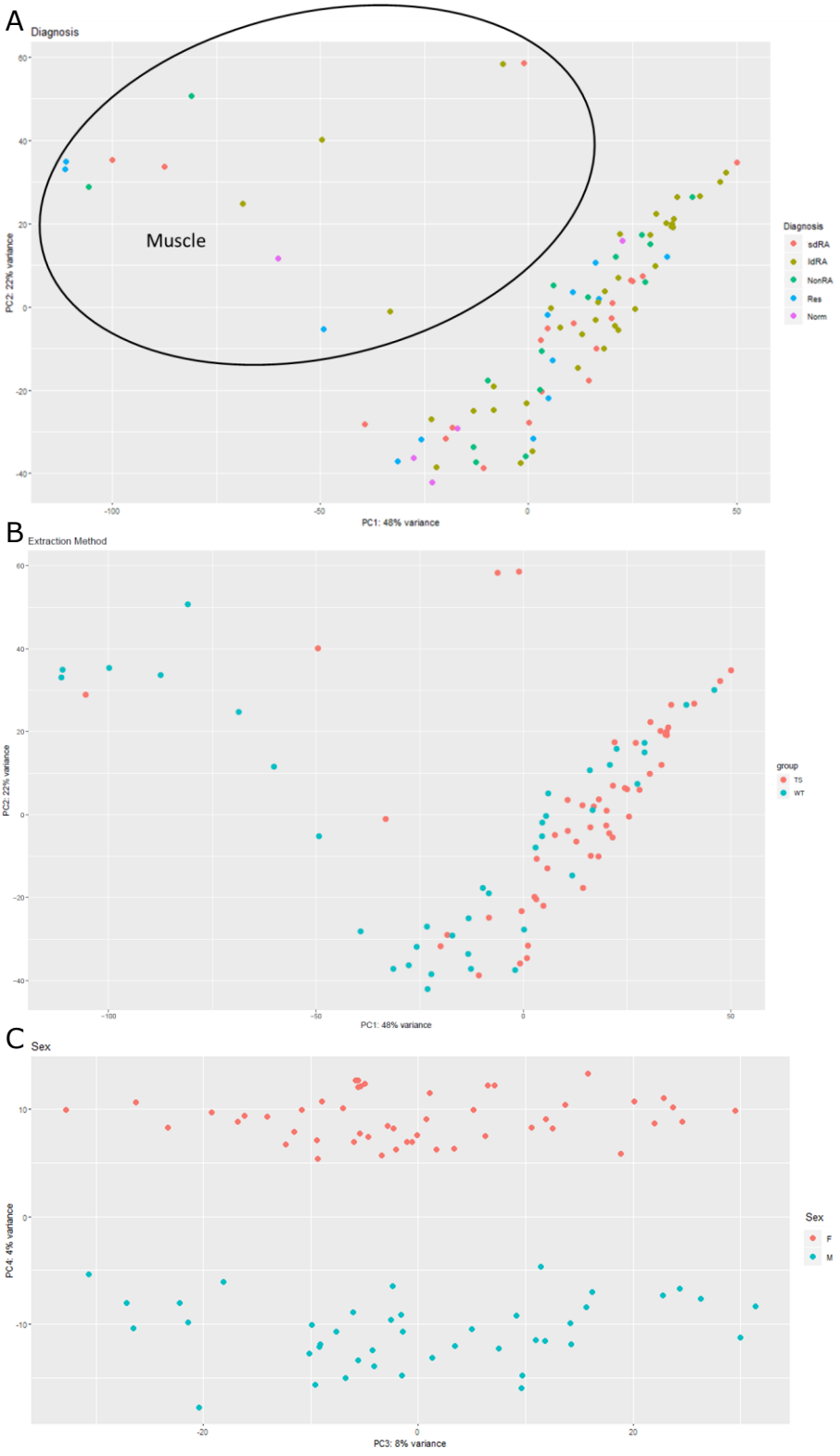


Figure 4.3: PCA plots showing muscle separation (A) and extraction method (B) on PC1/2 and sex on PC4 (C). PCA based on top 500 most variable genes. Extraction method refers to RNA that was extracted from OCT-embedded frozen block thick sections (TS) or snap-frozen whole tissue (WT). F = female, M = male.

Design Element	PC1	PC2	PC3	PC4	PC1-4	100 genes	1000 genes
RIN			x	x	x	x	x
DAS28-ESR			x	x		x	x
Plasma cell score			x	x		x	x
US GS	x		x			x	x
Biopsy joint	x					x	x
Biopsy side		x				x	x
DAS28-CRP				x		x	x
Density grade	x					x	x
Extraction method			x			x	x
Lining layer thickness		x				x	x
NSAID therapy			x			x	x
Age	x						x
Aggregate grade						x	x
ACPA						x	x
Diagnosis						x	x
Pathotype					x		x
Library prep batch						x	x
Pred therapy						x	x
RF		x				x	
US PD					x	x	
Sex							x

Table 4.2: Factors that were selected during stepwise regression analysis. Stepwise regression analysis using forward selection based on p-value was conducted on the top 100 and 1000 most variable genes, on individual PCs, and on combined PCs (PC1-4). 'x' indicates the inclusion of the variable in the regression model using the corresponding data input. Plasma cell score = histology-based score as described in (Orange et al., 2018). NSAID therapy = use of non-steroidal anti-inflammatory drugs at baseline. Pred therapy = use of Prednisolone at baseline. ACPA = anti-citrullinated protein antibodies positivity. RF = Rheumatoid factor positivity.

As RIN score had an impact on the overall variability of the data, we next explored its relationship to biologically important factors, such as clinical group, ultrasound variables, and BEACON histology variables (Figure 4.4). RIN was not variable across clinical group, meaning this should not have an impact on any clinical comparisons. However, RIN was associated with histology variables, with lymphoid tissues resulting in better RIN scores than pauci-immune tissue

(p=0.005) and RIN correlating with mean density (r=0.41, p=0.0001). RIN also weakly correlated with disease activity (DAS28-ESR: r=0.29, p=0.007). Due to concern about removing the biologically relevant variation that was associated with RIN and the consistency of RIN across clinical groups, RIN was not corrected for in the design.

Another potential confounding factor, which appeared to separate on PC1/2 was whether whole tissue (WT) or thick sections (TS) were used for RNA extraction (extraction method) (Figure 4.3). Therefore, relationship of extraction method with other variables was also explored (Figure 4.4). Interestingly, RIN was significantly higher in tissue from TS than from WT, so these two confounding factors may be related. However, the extraction method used was not equally distributed across clinical group, due to differences in the types of tissue that were available. This is particularly the case for the normal samples, for which we did not have access to OCT-embedded frozen tissue for TS. There was a significant association between extraction method and pathotype (p<0.0001), with increased use of TS in lymphoid and WT in pauci-immune, and density was higher in tissues from TS (p<0.0001). For some samples, both WT and TS were used for RNA extraction, with the best quality RNA then being used for sequencing. The difference in RIN between TS and WT (WT RIN – TS RIN) from these samples was calculated and Wilcoxon signed rank test performed for each pathotype to test for non-zero median values. The diffuse pathotype was the only one with a non-zero median value (p=0.008), suggesting an improved RIN from TS in diffuse tissue. This may mean that a bias was introduced when selecting the best RIN scores from samples that had both tissue types. Again, given this link between histology scoring methods and extraction method, to

avoid removing biologically relevant signals it was decided that extraction method would not be corrected for in the design.

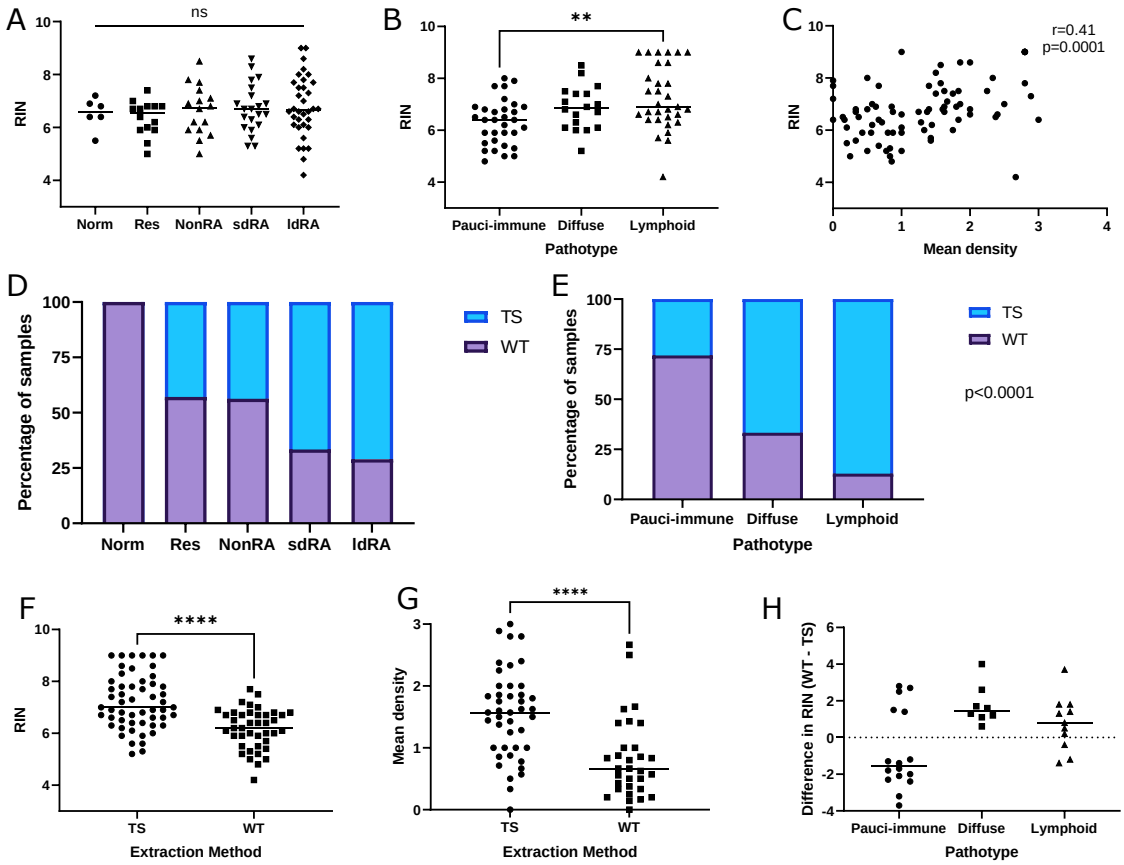


Figure 4.4: Relationship between RIN, extraction method, clinical and histological variables. (A) RIN vs clinical group, (B) RIN vs pathotype, (C) RIN vs BEACON density, (D) extraction method across clinical group, (E) extraction method across pathotype, (F) RIN vs extraction method, (G) BEACON density vs extraction method, (H) difference in RIN (WT-TS) vs pathotype. RIN score was measured by Genomics Birmingham prior to library preparation. Extraction method refers to RNA that was extracted from OCT-embedded frozen block thick sections (TS) or snap-frozen whole tissue (WT). Pathotype (n=81) and clinical group (n=95) show percentages. Chi squared, Spearman rank or Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

4.4 Exploring variation

After removing the Y chromosome, XIST, TSIX, IGV genes, and the samples containing skin or muscle, a more in-depth exploration of the remaining sources

of biological variation was undertaken to understand what drives variation in this mixed inflammatory arthritis cohort. The most variable genes and PCA were explored again to understand what was now driving the majority of the variation in the data (Figure 4.5 and Figure 4.6). Although samples did not cluster by clinical group, in the heatmap of the top 100 most variable genes, they broadly clustered by pathotype, with the majority of the pauci-immune samples clustering together. Pauci-immune pathotype also separated from diffuse and lymphoid on PC1/2, suggesting that histologically derived pathotype is associated with variation in the RNA sequencing data. When exploring the most variable genes and the genes that contribute to PC1/2, there was a large signature of immune-related genes, including a number of IG genes, and a potentially adipose signature, including the genes ADIPOQ, PLIN1, and PLIN4. These may reflect the different cell types that are predominant in different pathotypes. There were also some genes involved in tissue destruction, including the matrix metalloproteinases MMP1 and MMP3.

A scree plot was produced to show how much variation was accounted for in each PC (Figure 4.6). This showed that the vast majority of variation was associated with PC1 but that exploration up to PC4 may be worthwhile, after which the associated variation dropped to below 5%. Exploration of the gene contribution to PC3/4 showed a chemokine signature, including CXCL9, CXCL10, and CXCL11, in addition to another signature along PC3 involving genes such as SCUBE1, GPR1, and CLIC5. Exploration of these PC3 genes in the publicly available AMP RA Phase I dataset (Zhang et al., 2019) suggested they may be predominantly expressed in synovial fibroblasts, in particular lining layer fibroblasts (Figure 4.7). This potential fibroblast aspect was not captured by

BEACON pathotype as there was no clustering by pathotype on PC3/4. To follow up on this potential lining layer fibroblast signature, correlation of median lining layer thickness and PC3 was explored, with a moderate correlation being shown (r=-0.42, p=0.0005).

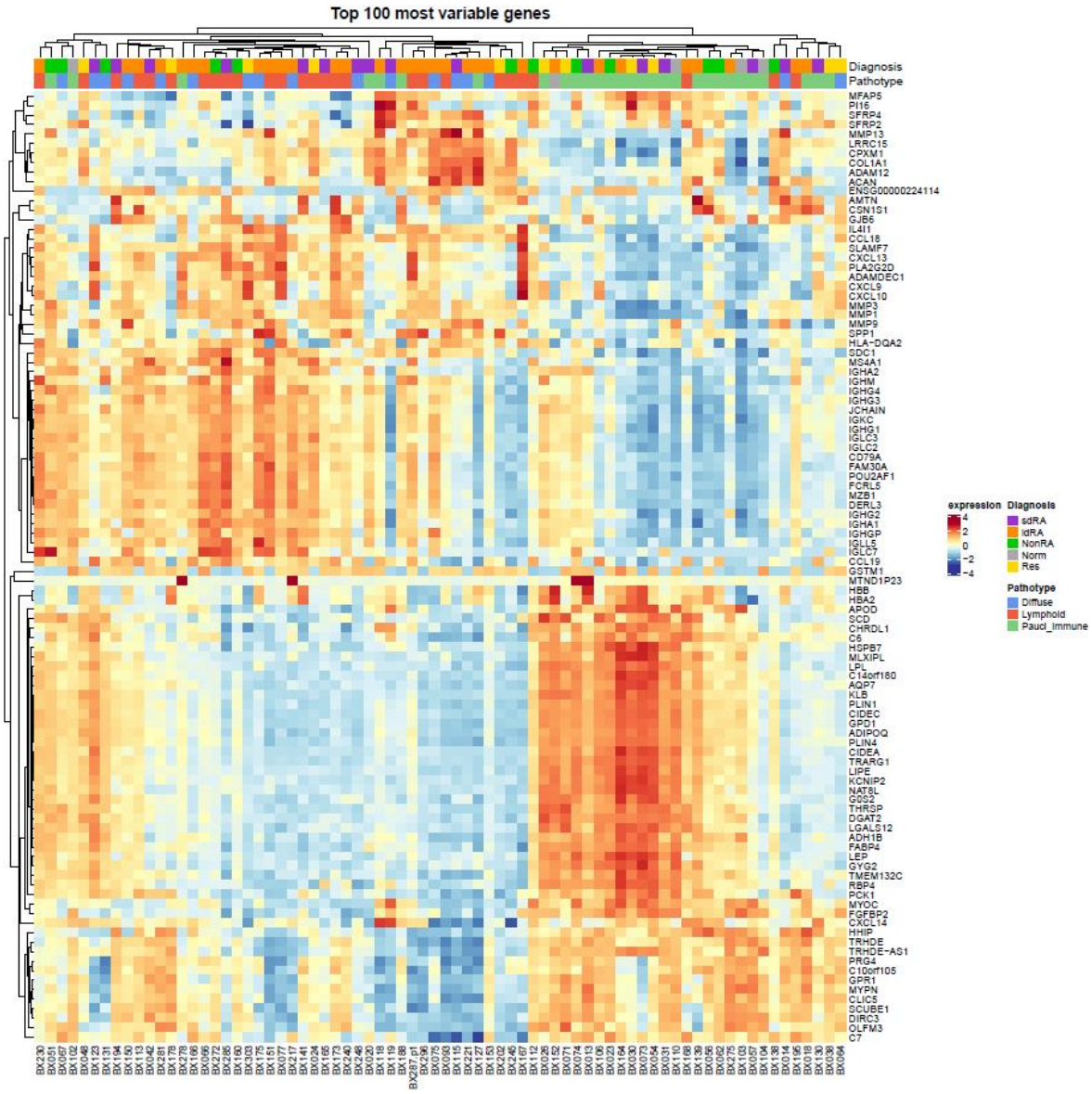


Figure 4.5: Heatmap of top 100 most variable genes after removal of samples containing skin or muscle, sex-related genes, and IGV genes. Top 100 most variable genes are shown with hierarchical clustering on columns and rows. Bars along the top show clinical group and pathotype.

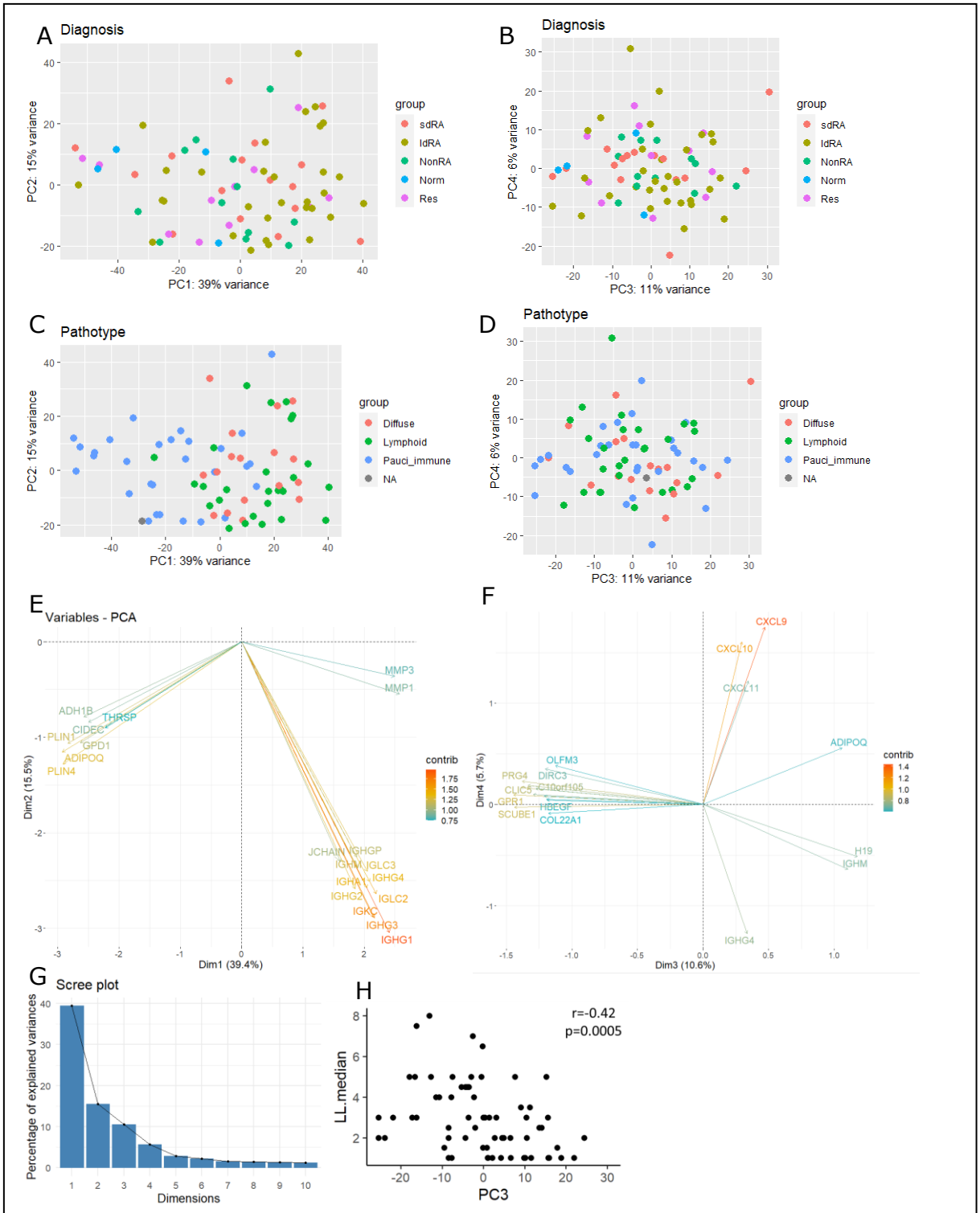


Figure 4.6: PCA plots for PC1/2 and PC3/4 and scree plot. Clinical group (A & B) and pathotype (C & D) are shown on PCA plots. PCA was conducted on top 500 most variable genes. Gene contribution plots (E & F) show the top 20 genes. Scree plot (G) shows variance associated with each PC. Correlation between lining layer thickness and PC3 (H) assessed using Spearman rank. LL.median = lining layer median, absolute number of cell thickness of the lining layer from histology averaged across biopsy fragments.

Figure 4.7: Exploration of genes contributing to PC3 in the AMP Phase I RA dataset. Genes from gene contribution plot in Figure 4.6. Single cell RNA sequencing data of selected cellular populations (monocytes, fibroblasts, B cells & T cells). M1 = IL1B+ pro-inflammatory monocytes, M2 = NUPR1+ monocytes, M3 = C1QA+ monocytes, M4 = IFN-activated monocytes, F1 = CD34+ sublining fibroblasts, F2 = HLA+ sublining fibroblasts, F3 = DKK3+ sublining fibroblasts, F4 = CD55+ lining fibroblasts, B1 = IGHD+ CD270 naive B cells, B2 = IGHG3+ CD27- memory B cells, B3 = autoimmune-associated cells (ABC), B4 = Plasmablasts, T1 = CCR7+ CD4+ T cells, T2 = FOXP3+ Tregs, T3 = PD-1+ Tph/Tfh, T4 = GZMK+ CD8+ T cells, T5 = GNLY+ GZMB+ CTLs, T6 = GZMK+/GZMB+ T cells. Data available from https://immunogenomics.io/ampra/ (Zhang et al., 2019).

4.4.1 UMAP clustering

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction was then explored to look the variation captured in the top 1000 genes in one

figure (Figure 4.8). Similarly to PCA, clinical group did not cluster on the UMAP, while there was some separation by pathotype.

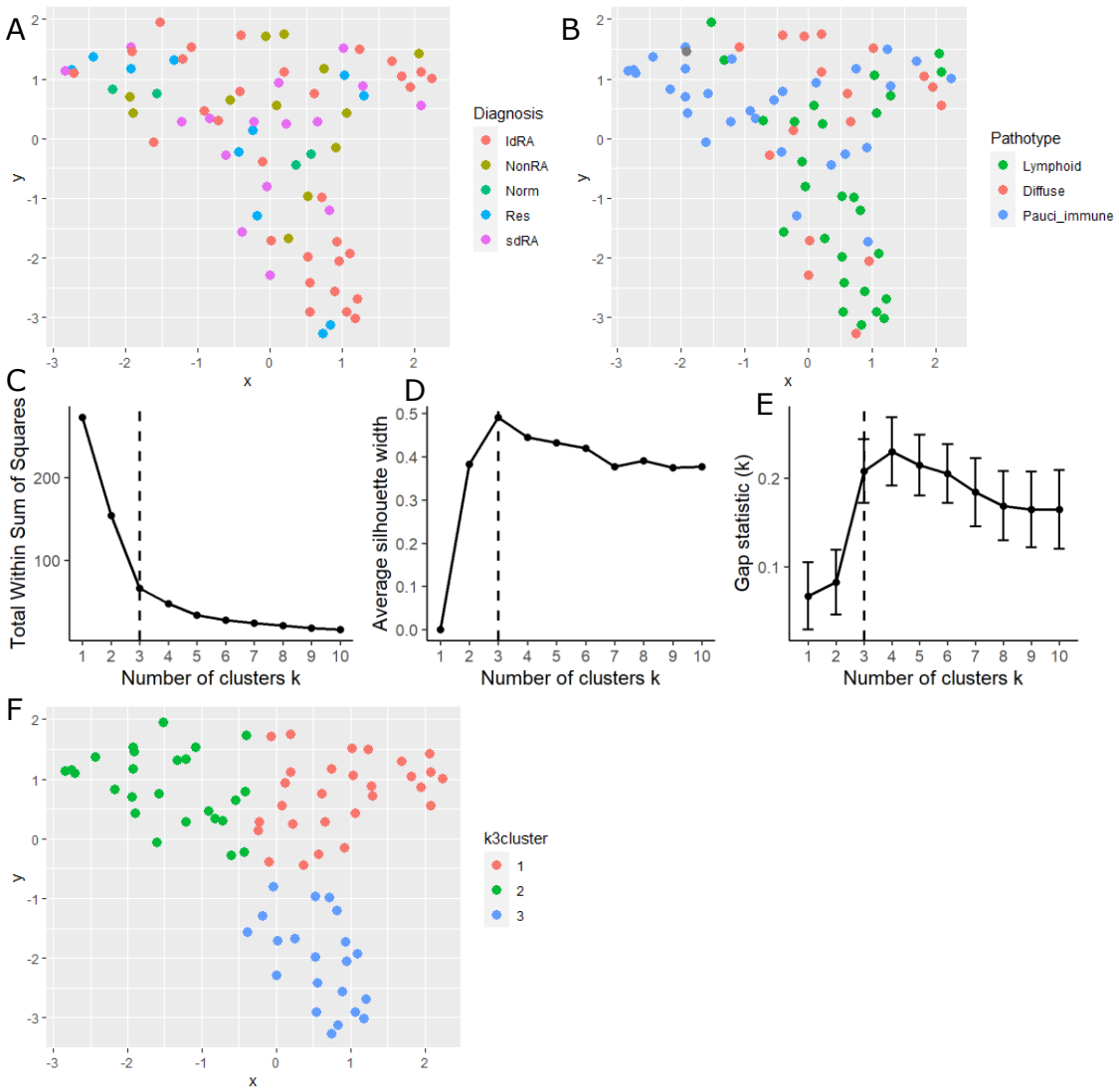


Figure 4.8: UMAP plots and k-means clustering. UMAP plots show clinical group (A) and pathotype (B). Plots showing elbow (C), average silhouette (D), and gap statistic methods (E) for cluster number selection show that three clusters are optimum. (F) UMAP plot showing the three clusters from k-means clustering.

K-means clustering was performed on the UMAP to explore how the samples clustered and to aid interpretation. In order to decide how many clusters would be most appropriate a number of techniques were used, including elbow, average

silhouette, and gap statistic methods (Figure 4.8). From this, three clusters were identified as being optimal.

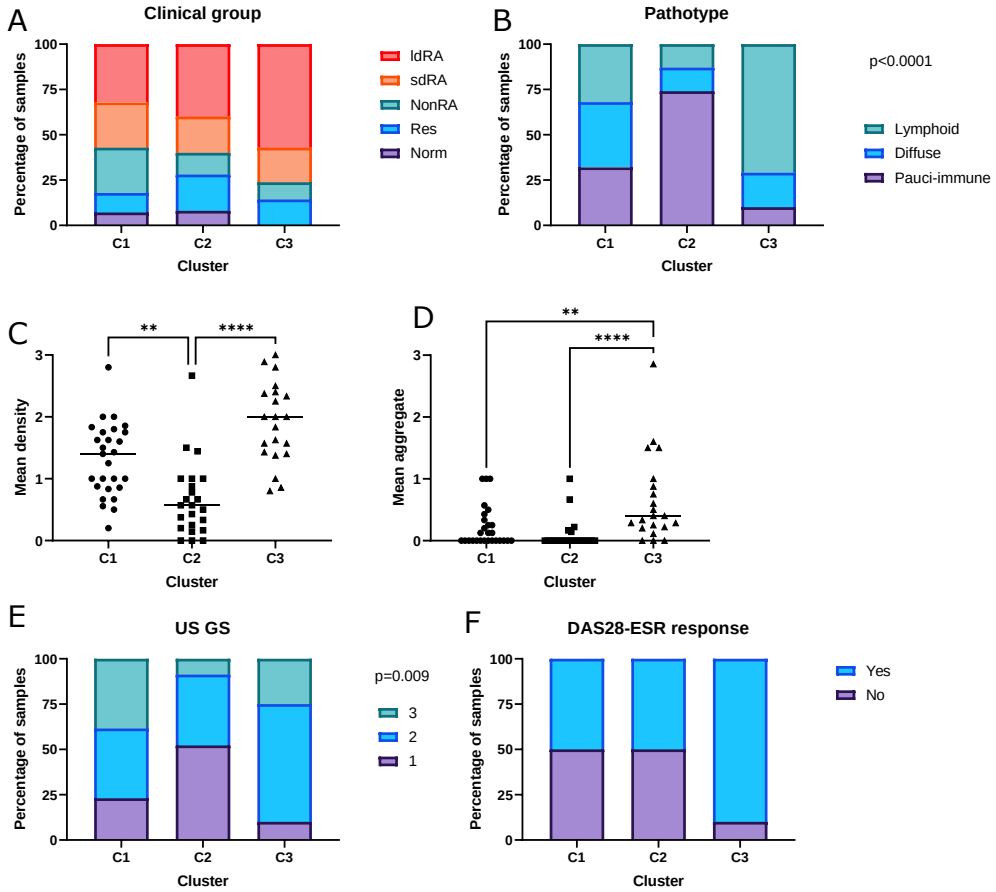


Figure 4.9: Relationship between UMAP clusters, clinical and histological variables. Clusters from k-means clustering of UMAP data. Clinical group (A, n=74), pathotype (B, n=73), US GS (E, n=69), and DAS28-ESR response (F, n=18) show percentages of samples. For density (C) and aggregates (D) bars show median values. Pathotype, density, and aggregate grades are from the BEACON scoring system (described in Chapter 3.2). DAS28-ESR response = EULAR DAS28-ESR response at 12 months. Chi squared or Kruskal-Wallis with post-hoc Dunn's multiple comparison tests; **p<0.01, ***p<0.001, ****p<0.0001.

Histological and clinical characteristics were then explored within these three clusters (Figure 4.9). There was a trend towards increased ldRA in cluster 3 (C3), although this was not statistically significant. Pathotypes differed across the three clusters (p<0.0001), with C2 being predominantly pauci-immune, C3 being

mostly lymphoid, and C1 being a mixed group. Furthermore, there was a difference across density and aggregate grades, with C2 having lower density grade compared to C1 and C3 ($p=0.003$ and $p<0.0001$) and C3 having higher aggregate grade than C1 and C2 ($p=0.01$ and $p<0.0001$). US GS also showed a similar pattern ($p=0.009$), with C3 tending towards higher US GS and C2 having a larger proportion of grade one. Relationship to EULAR DAS28-ESR response (good and moderate versus poor response) at 12 months was also explored within a subset of RA patients, with C3 showing a trend towards increased response rates but there were insufficient numbers to test this statistically.

4.4.1.1 Gene exploration

To explore the genes associated with each cluster, differential expression analysis on each cluster compared to the other two was undertaken. Taking the genes upregulated in each cluster ($padj<0.05$, $LFC>2$) resulted in 876 genes, of which 49 were upregulated in C1, 530 in C2, and 297 in C3 (Figure 4.10). Exploration of the heatmap of DEGs confirms that C2 and C3 are most distinct, with C1 having fewer upregulated DEGs compared to the other two clusters. This is further confirmation of C1 being a mixed cluster, as was suggested when exploring histological and clinical variables.

GO biological processes associated with genes upregulated in each cluster were then explored, finding skeletal system morphogenesis, extracellular matrix organisation, and extracellular structure organisation as the top processes in C1, antibiotic metabolic process, oxygen transport, and retinol metabolic process in C2, and B cell activation, regulation of lymphocyte activation, and immune response-activating cell surface receptor signalling pathway in C3 (Figure 4.11).

A list of all significant GO biological processes associated with each cluster can be found in Appendix 9.5.1.

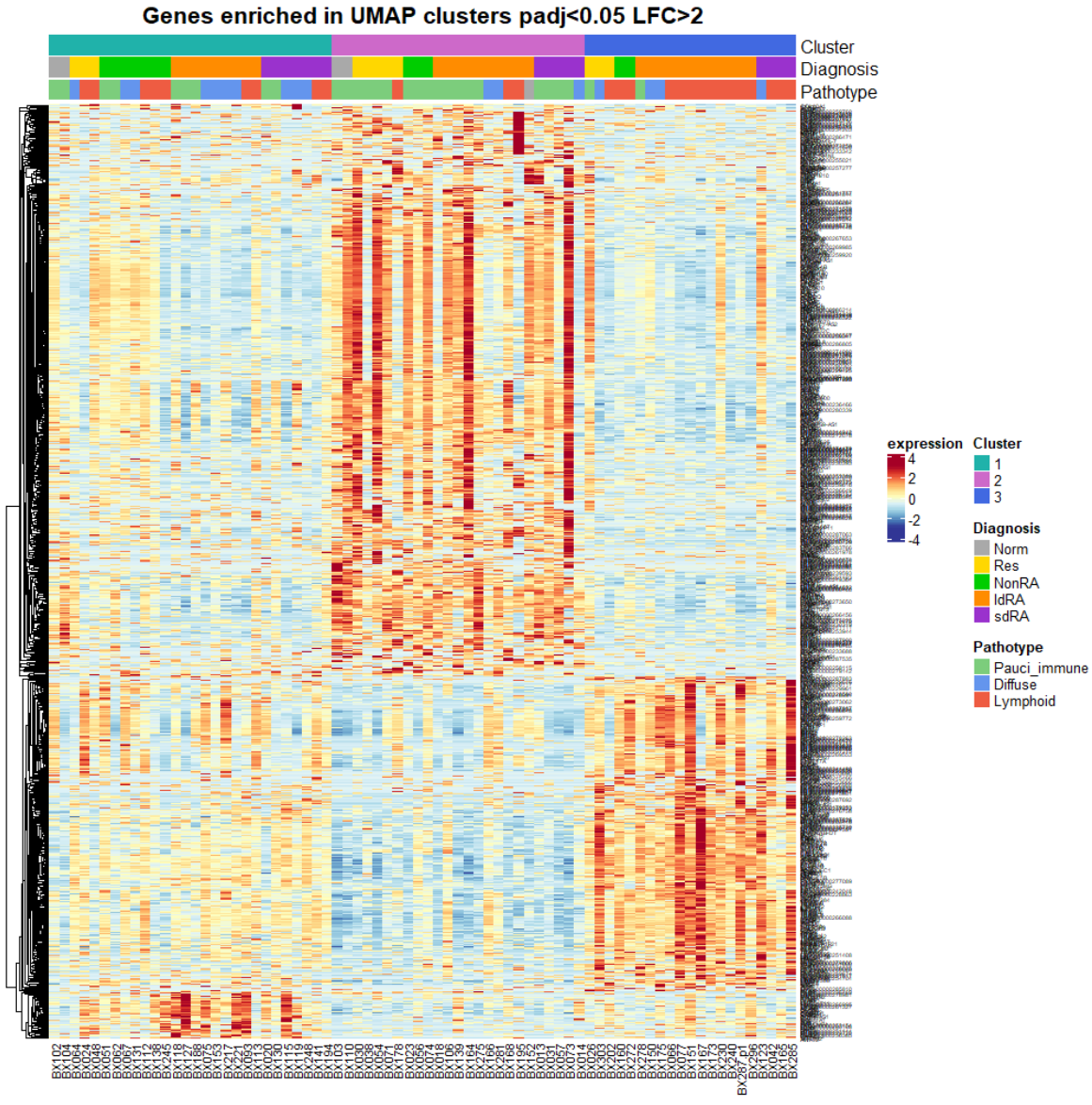


Figure 4.10: Heatmap of the DEGs upregulated in each cluster compared to the others. Padj<0.05, LFC>2. All upregulated genes shown on one plot: 49 genes were upregulated in C1, 530 in C2, and 297 in C3. Bars along the top show UMAP cluster, clinical group, and BEACON pathotype.

This further highlights the strong adaptive immune signature present in C3, which is in line with the large proportion of samples in this group with the

lymphoid pathotype that have high density and aggregate grades. Investigation of the 10 top most significant genes upregulated in C3 further confirm this, with genes involved in T cell priming (SEMA4A), inflammasomes (GBP5 and AIM2), NK cell activation (CALHM6), macrophage polarisation (SNX10), and other inflammatory processes (ADAM8) (Koda et al., 2020, Ebihara et al., 2010, Shenoy et al., 2012, Lugrin and Martinon, 2018, Bartsch and Schlomann, 2013, You et al., 2016). However, three of the most significant genes, namely CD274 (PDL1), IL4I1, and SILEC10, were immune inhibitory, with CD274 being an inhibitor of T cell activation and cytokine secretion, IL4I1 being an inhibitor of T cell proliferation and being able to promote Treg differentiation, and SILEC10 being an inhibitor of the innate immune response (Freeman et al., 2000, Boulland et al., 2007, Cousin et al., 2015, Zhang et al., 2015a, Munday et al., 2001).

The genes upregulated in C1, which was the mixed cluster, were associated with pathways involved in extracellular matrix organisation and a number of the genes were collagen genes, with 3 of the top 10 most significant genes encoding collagen chains (COL1A1, COL10A1, and COL11A1). Within the top 10 genes, there were a number of genes associated with tissue destruction or repair, including CTHRC1, ADAM12, POSTN, and CHRDL2, alongside a notable link to Wnt signalling, with CTHRC1, ALPK2, POSTN, and KIF26B being found to be involved in this pathway (Hofsteen et al., 2018, Myngbay et al., 2019, Zhang et al., 2021a, Susman et al., 2017). Overall, this suggests a tissue remodelling phenotype in this cluster.

The genes upregulated in C2, which is predominantly composed of the pauci-immune pathotype, were associated with multiple metabolic processes, including

antibiotic metabolism, retinol metabolism, and allied processes, such as oxygen transport, bicarbonate transport, and cellular detoxification, alongside cell-cell adhesion. Cell adhesion was also associated with the low inflammatory subtype identified by clustering RNA sequencing data from RA synovial tissue in the study by Orange et al. (2018).

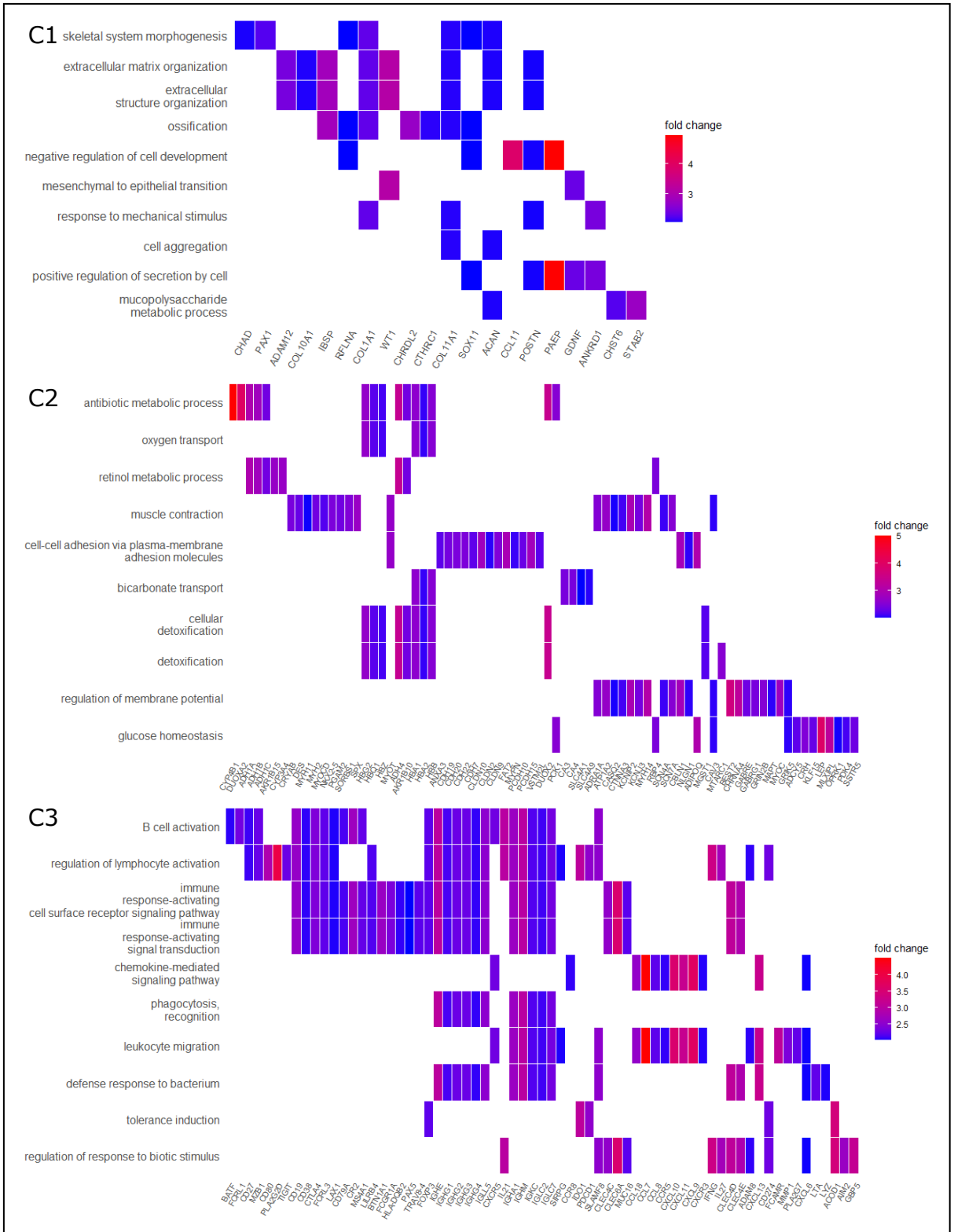


Figure 4.11: GO biological process analysis on genes upregulated in UMAP clusters. Top 10 processes are shown for C1 (top), C2 (middle), and C3 (bottom). DEGs from unadjusted p<0.05, LFC>2. A list of all significant GO biological processes can be found in Appendix 9.5.1.

The top 10 most significant upregulated genes in C2 were explored in the literature. CPAMD8 has a potential role in innate immunity, although has not been extensively studied to date (Li et al., 2004). LTBP4 is required for TGF-β signalling as it stabilises the receptor complex and TGF-β has complex and wide reaching impacts on immune processes, having the ability to induce T cell tolerance but also having roles in the differentiation of T cell subsets by inhibiting Th1 and Th2 cell formation and driving Treg and Th17 cell formation (Su et al., 2015a, Sanjabi et al., 2017). TGF-β is also a negative regulator of B cells and NK and can induce tolerance in DCs (Sanjabi et al., 2017). TGF-β pathways were also associated with the low inflammatory subtype in the study by Orange et al. (2018).

MAOA is able to drive oxidative breakdown of amines and has a suggested role in M2 anti-inflammatory macrophages, where it may increase production of hydrogen peroxide (Cathcart and Bhattacharjee, 2014). DUOX2 also produces hydrogen peroxide, a reactive oxygen species (ROS) that causes oxidative stress. Oxidative stress has been found to increase cartilage and joint destruction in RA (Wruck et al., 2011, Mirshafiey and Mohsenzadegan, 2008).

CYP4B1 is involved in xenobiotic metabolism, fatty acid oxidation, and has been found to be decreased during acute allergic inflammation and increased following resolution in a mouse model (Stoilov et al., 2006, Thesseling et al., 2020). However, the function of CYP4B1 in humans has been questioned, with it being suggested that it may not be as active as it is in animal models (Zheng et al., 1998, Thesseling et al., 2020).

HLF expression has been found to be increased following treatment with methotrexate, potentially being involved in methotrexate-induced synovial fibroblast apoptosis (Suzuki et al., 2018). However, another study found that HLF was able to inhibit apoptosis in mouse epidermal cells and human keratinocytes (Waters et al., 2013). The role of HLF is therefore unclear but may have differing roles in different cell types or microenvironments.

RAMP2-AS1 is a lncRNA that has been suggested to increase senescence and inhibit sprouting in endothelial cells, potentially via increasing the expression of RAMP2 (Lai et al., 2021). Furthermore, roles in angiogenesis, innate immunity, and inflammation have been suggested, with inhibition bring pro-inflammatory (Lai et al., 2021).

HIF3A has been described to be a negative regulator of cellular response to hypoxia, although has many different splice variants that can have differing roles in different tissues (Hara et al., 2001, Maynard et al., 2005, Duan, 2016). It is not possible from the current analyses to know which splice variants are upregulated in C2. The remaining most significant upregulated genes were a pseudogene (HMGN2P15) and an uncategorised gene (ENSG00000280339).

Exploration of the top 10 most significant genes in the literature did not result in a clear gene expression signature for C2, as was the case for the other two clusters. However, in the literature many of these genes have been associated with broadly anti-inflammatory functions, with some being found to decrease during inflammation or increase following treatment.

Exploring the genes that were downregulated in each cluster compared to the other two further highlighted the clear differences between C2 and C3 and the

more separate, mixed nature of C1. Downregulated pathways in C2 broadly overlapped with the pathways that were upregulated in C3 and visa versa. Namely B cell activation, cytokine-mediated signalling pathway, phagocytosis, leukocyte migration, and defense response to bacterium were all downregulated pathways in C2 and upregulated in C3 and cell-cell adhesion via plasma-membrane adhesion molecules was downregulated in C3 and upregulated in C2 (Figure 4.12). Furthermore, 153 of the 266 genes genes downregulated in C3 were upregulated in C2 and 176 of the 297 genes downregulated in C2 were upregulated in C3.

The pathways that were associated with genes downregulated in C1 were predominantly associated with metabolic processes, including lipid catabolic processes, glucagon secretion, polyol biosenthetic process, lipid droplet organisation, and negative regulation of amine transport (Figure 4.12).

Exploration of the top 10 most significant downregulated genes in C1 was undertaken. Two of these genes, namely C15orf62 and RHOXF1-AS1, have not been studied previously so their function is currently unknown. ALDH2 was the most significant downregulated gene, which has been found to reduce oxidative stress and inhibit inflammation, with this being seen as a potential target for the treatment of knee OA (Pan et al., 2021, Cao et al., 2022).

CHIT1 is expressed by a number of immune cells, predominantly macrophages, and is upregulated in fibrotic lung disease but downregulated in ulcerative colitis (Mazur et al., 2021). It has been suggested that CHIT1 drives alternative macrophage activation, to increase fibrosis, but is not associated with classical macrophage activation (Lee et al., 2022b).

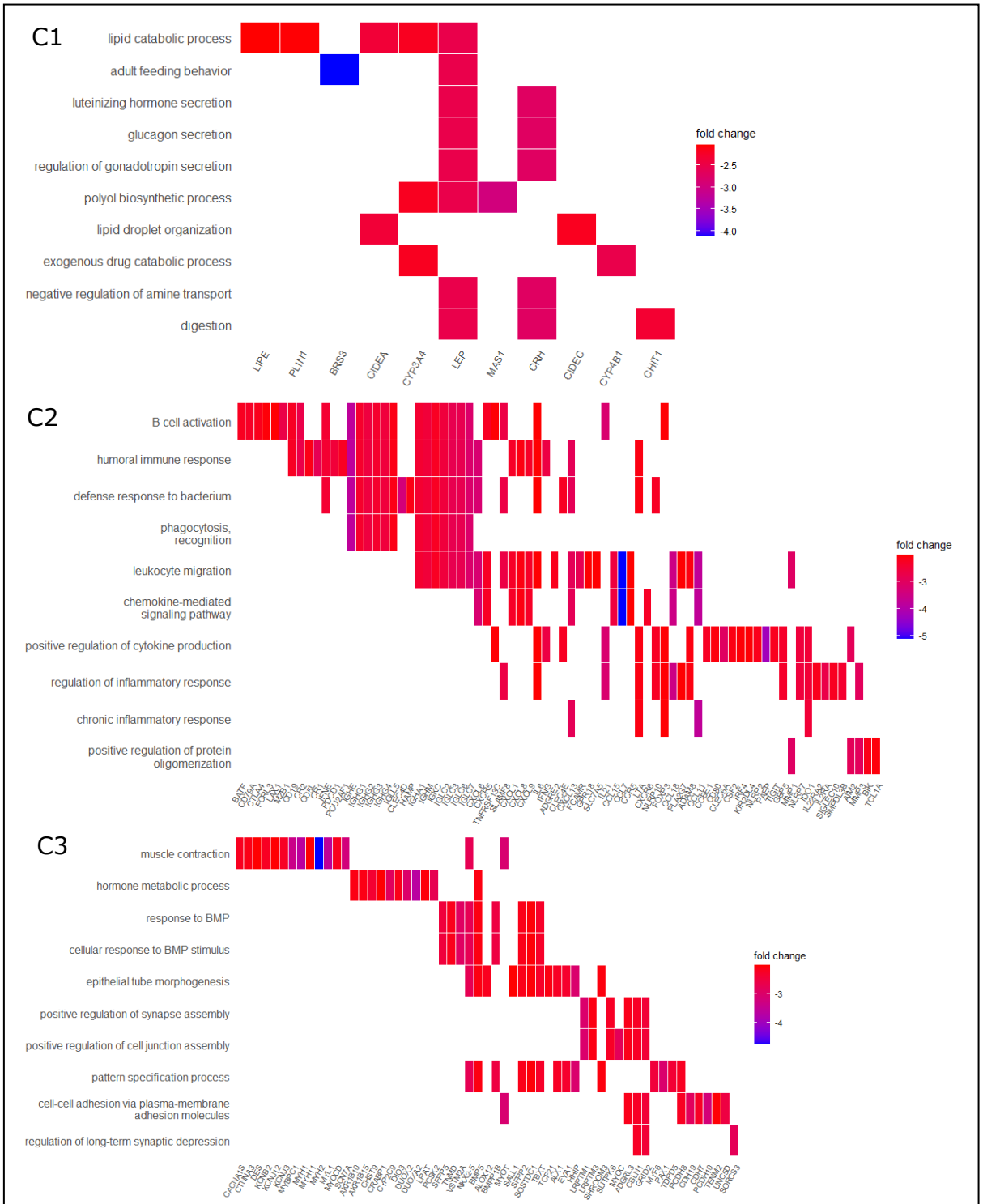


Figure 4.12: GO biological process analysis on genes downregulated in UMAP clusters. Top 10 processes are shown for C1 (top), C2 (middle), and C3 (bottom). DEGs from unadjusted p<0.05, LFC>2. A list of all significant GO biological processes can be found in Appendix 9.5.1.

CEBPA is associated with adipogenesis, increasing the expression of enzymes involved in lipid metabolism, and is thought to be required for adipocyte development (Kwak et al., 2012, Lee et al., 2022a). PNPLA2 is another downregulated gene associated with adipocytes and lipid metabolism, being a key enzyme in lipolysis and associated with reduction in lipid droplet accumulation (Yang and Mottillo, 2020, de la Rosa Rodriguez and Kersten, 2020). CEBPA has also been found to inhibit IL6 expression in chrondrocytes, with it being downregulated in OA, and may have a role in osteoclast differentiation (Makki and Haqqi, 2017, Chen et al., 2013). FFAR4 (GPR120) has also been found to inhibit IL6, as well as IL8 and IL1β, and was found to reduce cartilage damage and ECM degradation (Xu et al., 2020).

SNX10, which was also found to be downregulated in C3, has roles in macrophage polarisation and has previously been associated with rheumatoid arthritis, increasing bone erosion via osteoclast activation and the increased secretion of MMP9 (Zhou et al., 2017). This is perhaps contrary to previous findings in this cluster, as tissue remodelling-associated genes are generally upregulated.

BATF2 (SARI) expression is involved in the regulation of immune cells, particularly in the activation of dendritic cells and differentiation of T cells, driving formation Th1 and Treg cells and inhibiting Th17 differentiation (Yokoyama-Kokuryo et al., 2020, Li et al., 2021). It has been associated with response to abatacept in RA, is protective in colitis, and has anti-tumour suppressive roles (Yokoyama-Kokuryo et al., 2020, Zhang et al., 2020b). In ocular disease, BATF2 has been found to inhibit both vascularisation and inflammation (Zhang et al., 2020b).

SRCIN1 (p140Cap) is associated with good prognosis in many tumours due to the reduction of migration and cell growth (Salemme et al., 2021). Silencing increases invasiveness and cell spreading, suggesting an inhibitory role in metastasis formation (Di Stefano et al., 2007). SRCIN1 has also been found to inhibit angiogenesis and proliferation in endoethelial progenitor cells (Wang et al., 2019b).

Overall from the literature, the genes downregulated in C1 are predominantly involved in lipid metabolism, inflammation, or destructive tissue remodelling processes, with many of them being inhibitory of inflammation and tissue remodelling suggesting upregulation of these processes in this cluster. This is broadly in agreement with the upregulated genes and processes, although highlights a potential downregulation of lipid catabolic processes that was not previously highlighted.

4.5 MUSCLE CORRECTION

Skeletal muscle is sometimes captured during the biopsy process because of proximity (as can be seen in Figure 4.13 image BX201), particularly round the knee joint, but is not considered anatomically part of the synovium. As previously discussed, upon exploration of the top 100 most variable genes, a cluster of skeletal muscle-related genes, including MYH2, MYH7 and ACTA1, were highly expressed by 13 samples (Figure 4.2). The same samples also clustered separately on PCA (Figure 4.3). H&E sections from these samples were examined to see whether muscle could be seen histologically (Figure 4.13). This was the case for many of these samples, even though different biopsy pieces were used for the RNA sequencing than histology. In two of these samples there was a very

high proportion of muscle compared to synovial tissue in the H&E section so these were removed from future analyses.

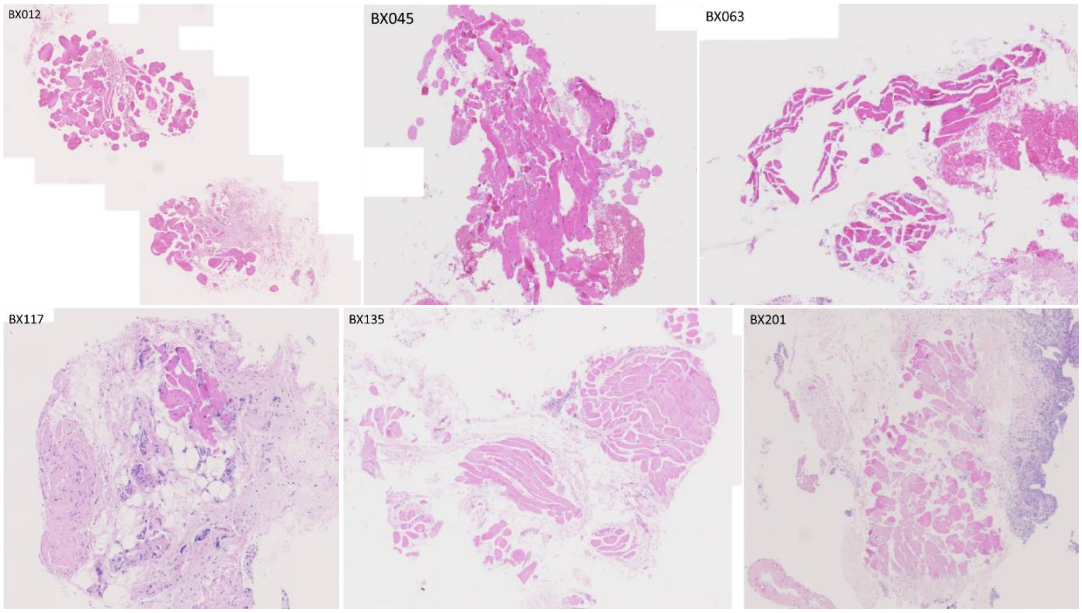


Figure 4.13: Representative H&E images showing muscle in synovial biopsies. Samples identified as having muscle present in RNA sequencing data were explored in H&E stained tissue, with many of them having muscle visible in the tissue. Biopsy fragments used for H&E staining are not the same as biopsy fragments used for RNA sequencing, although they were taken during the same biopsy procedure. H&E images were produced as described in Chapter 2.2, with fixing, sectioning, and staining being undertaken by the University Hospitals Birmingham pathology department and imaging being done by Dr Jennifer Marshall or Dr Triin Major.

A number of methods were tested to remove or correct for this, including removing the remaining 11 samples that were responsible for the muscle signal, and labelling these samples as containing muscle to allow for correction in the differential expression design. Removal of the samples containing muscle reduced numbers within the dataset considerably and so reduced the overall statistical power. Additionally, when using both these methods, there were still some muscle-related genes and pathways being found in later analyses. It was

therefore hypothesised that there may be further lower-level muscle present in other samples.

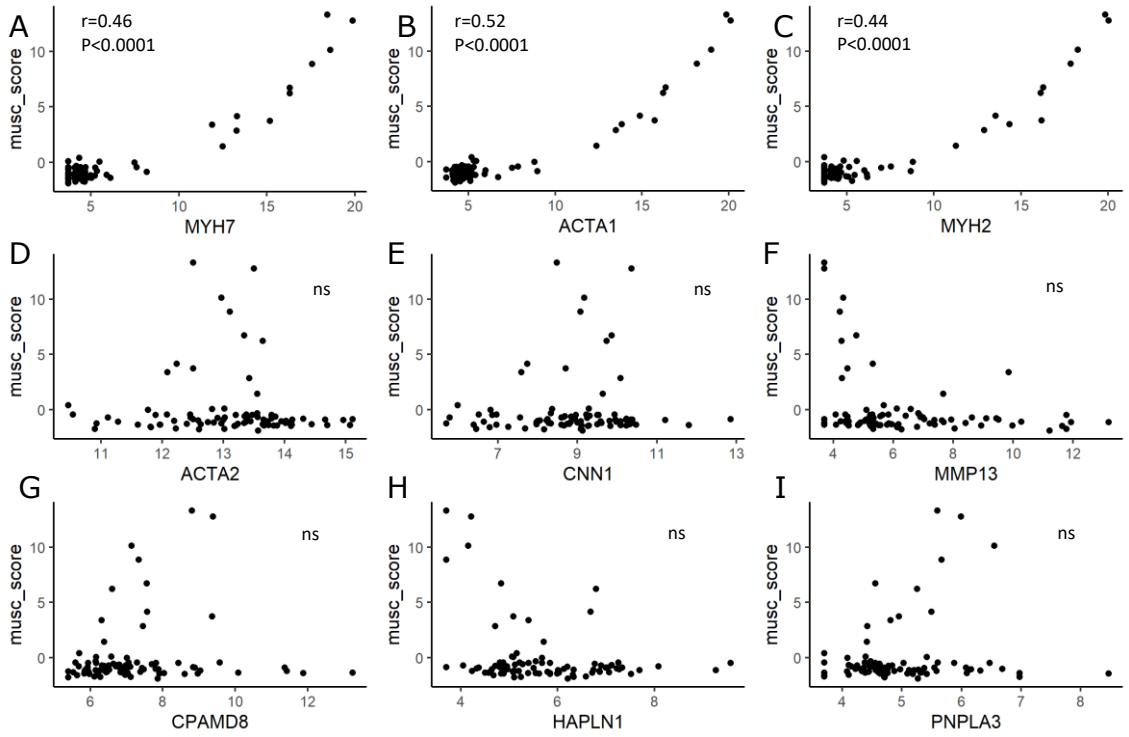


Figure 4.14: Relationship between skeletal muscle score and genes expressed in skeletal muscle, smooth muscle, and unrelated genes. Gene expression shown as log values. MYH7 (A), ACTA1 (B) and MYH2 (C) are expressed in skeletal muscle. ACTA2 (D) and CNN1 (E) are expressed in smooth muscle. MMP13 (F), CPAMD8 (G), HAPLN1 (H) and PNPLA3 (I) are unrelated genes. Musc_score refers to a skeletal muscle summary score that averages expression of genes that are most selective for skeletal muscle. Spearman rank used to assess correlation.

In order to capture all skeletal muscle related gene expression within the dataset, a 'muscle score' was created. Genes that were most uniquely expressed in skeletal muscle tissue were identified using the Human Protein Atlas (IDI2, DUPD1, MYH1, LRRC30, MYH4, SMTNL1, ACTN3, PPP1R27, MYADML2, ANKRD23, UCP3, and CHRNA10) (Uhlen et al., 2015, Yu et al., 2015, Fagerberg et al., 2014, Lindskog et al., 2015). A summary score was calculated that averaged skeletal muscle-specific gene expression in each sample using a technique

developed by Wagle et al. (2018). Figure 4.14 shows how the resulting muscle score correlated with muscle genes that had previously been identified in our analyses, genes that are expressed in smooth muscle, and unrelated genes of interest. This was to ensure the muscle score was specific to skeletal muscle and would not remove relevant gene signatures. There was a strong correlation with skeletal muscle genes MYH7, ACTA1, and MYH2 ($r=0.46$, $p<0.0001$; $r=0.52$, $p<0.0001$; and $r=0.44$, $p<0.0001$, respectively) but no correlation with smooth muscle genes (ACTA1 and CNN1) or other genes of interest, suggesting good skeletal muscle specificity. This muscle score was therefore included in the design for differential expression analysis in order to correct for the presence of muscle.

4.6 Discussion

In this dataset, there were five normal control samples, which underwent exploratory arthroscopy due to the occurrence of joint pain with no evidence of synovial abnormalities. This was confirmed by examining the histology images available in the present study, where there was no visible inflammation, however the presence of joint pain in these patients does raise questions as to whether they truly represent normal synovium. Given the ethical implications of taking a synovial biopsy from healthy patients, access to uninflamed tissue with no perceived abnormalities based on clinical examination, imaging, and microscopy is the closest to normal as can reasonably be obtained. However, consideration must be given to the potential that the synovium from these controls is not entirely normal, particularly if pathways relating to pain are found to be significant in analyses.

Exploration of the overall variation within the dataset allowed for the identification of confounding variables and overrepresented genes. One such example is the removal of the IG variable genes to prevent overrepresentation, which was also done in a similar study by Orange et al. (2018). In the same study they also removed genes from both the X and the Y chromosome to remove sex biases. However, in this study we decided to retain the majority of the X chromosome, just removing XIST and TSIX, due to there being a sex bias in the prevalence of many arthritides, including RA (Buckwalter and Lappin, 2000). By being less stringent with removing all sex-related differences and not including this as a part of the differential expression design it was hoped that pathologically relevant sex-related differences would be retained.

Generally in this study we decided to have a minimal differential expression design to avoid over-correcting for potential confounding factors. This was because the variables that were identified as potential confounders, RIN and extraction method, were both heavily associated with histological variables. Correcting for these would have risked masking of biologically relevant signals.

There are a few potential reasons why RIN is associated with histological variables. It may be that absolute cell number in the tissue could have an effect on the RIN, as more cells likely results in more total RNA and higher RNA concentrations, which may result in better RNA quality overall. Lymphoid and diffuse tissues are likely to have increased cell number in the same sized tissue due to the presence of large numbers of infiltrating immune cells, and it is highly likely that increasing density score, which correlates with RIN, correlates with cell number.

It may also be that better quality RNA can be obtained from infiltrating immune cells than from resident cells. This may particularly be the case for adipocytes, which require non-standard methods to ensure good quality RNA extraction (Janke et al., 2001). Exploring the relationship between tissue composition and RIN may aid future research efforts and ensure that certain tissue types are not discriminated against, resulting in a biased view of inflammatory arthritis. However, this was beyond the scope of the current study.

The use of TS from OCT-embedded frozen blocks or snap-frozen WT, termed the extraction method, was also identified as a potential confounder. As WT was the default source of tissue and TS were added to increase sample numbers and improve on RIN scores where possible, this may have introduced a bias in the type of tissue used for each method. Extraction method was associated with histological variables, with there being a greater use of WT in pauci-immune samples. This is likely due to the tissue composition of pauci-immune samples, lacking infiltrating immune cells and potentially containing a larger proportion of adipocytes. This may mean that there is less RNA within a thick section, resulting in a lower RIN score and therefore resulting in the preferential use of WT. This hypothesis is supported by the RIN improving in tissues with a diffuse pathotype when using TS compared to WT from the same biopsy. Neither the lymphoid nor the pauci-immune samples had significantly non-zero medians, although the pauci-immune samples did have a negative median, meaning there was a trend towards better quality RNA from WT than TS in this group. This difference in quality across pathotypes from different sources of tissue for RNA extraction is something that should be considered and taken into account going forward to avoid creating a bias towards the study of diffuse and lymphoid pathotypes.

A considerable number of samples (n=13) were identified as having a skeletal muscle signature, which presents a number of challenges in whole tissue RNA sequencing data. Interestingly, muscle was also visible in the H&E stained sections for most of these samples, despite them being from distinct biopsy fragments. This, along with the association of pathotype with overall variability, suggests that generally pathotype derived from 6-8 distinct biopsy fragments is representative of the overall biopsy joint, thus validating the approach taken. This is also supported by previous studies looking at within joint variability, which were consulted when developing the original protocol (Kennedy et al., 1988, Dolhain et al., 1998, Boyle et al., 2003).

One potential approach to account for the presence of muscle could have been to remove the genes that appeared in the most variable gene analysis that were associated with muscle. However, this would be under the assumption that these were the only genes that had altered expression due to the presence of skeletal muscle, which did not seem like a reasonable assumption. Other approaches were therefore explored.

Samples that had dominant expression of skeletal muscle genes and that separated from the main cluster in PCA analysis could be removed, to remove the muscle signature altogether. This may result in a cleaner dataset overall but also considerably reduces the total number of samples, thus decreasing statistical power in later analyses. Moreover, there is a chance other samples may also have small amounts of skeletal muscle present in the tissue, which would not be taken into account with this method. Highlighting the samples containing muscle could also be used to correct for its presence in the differential

expression design, however this has the same challenge that any low level presence may not be accounted for.

The approach used in this study was to create a summary muscle score that should capture all gene expression associated with skeletal muscle that was used in the differential expression design. This was done by selecting genes that are most specific for skeletal muscle from the Human Protein Atlas (Uhlen et al., 2015, Yu et al., 2015, Lindskog et al., 2015, Fagerberg et al., 2014) and averaging the gene expression. To confirm this was specific for skeletal muscle, the score was correlated against expression of skeletal muscle genes identified in the most variable genes, smooth muscle genes, and unrelated genes from the dataset. This showed a high level of specificity, correlating with the skeletal muscle expressed genes exclusively. However, as this was only tested on a subset of genes it is possible there may be correlation with unknown off-target genes. In addition, it is unclear exactly what influence correcting for the muscle score will have on gene expression in each sample. Although this is the approach taken in this study, other approaches may be worth exploring further to test the reproducibility of the results. This was done for some of the downstream analyses, which suggested that using the muscle score and removing the samples largely produced similar results (data not shown).

Perhaps unsurprisingly, cellular composition had a large impact on overall variability within the dataset, with immune-related signatures capturing the bulk of variation in PCA, and this broadly being associated with histological pathotype. Another dominant signal included adipose related genes, such as ADIPOQ, PLIN1, and PLIN4. This signal was predominantly associated with the pauci-immune cluster. ADIPOQ encodes the adipokine adiponectin, which is known to have a

role in inflammation and is increased in plasma and synovial fluid in RA (Szumilas et al., 2020). PLIN1 and PLIN4 encode perilipins, which coat lipid droplets in adipocytes (Wolins et al., 2005). PLIN1 has been shown to decrease in expression in collagen-induced arthritis (CIA), a mouse model of RA (Arias de la Rosa et al., 2018). The presence of adipocytes in the synovium has been explored less than the presence of immune cells, despite one study by Wang et al. (2019a) suggesting that adipocytes may make up 12.5% of healthy synovial tissue, with this percentage decreasing in RA. The presence of a strong adipose signal in a mixed inflammatory arthritis dataset suggests this may be worthwhile exploring in future studies, as this may highlight new pathogenic mechanisms that could be targeted therapeutically.

After exploration of the genes contributing to PC3 in the publically available AMP Phase I RA dataset (Zhang et al., 2019), another signal appeared to be associated with lining layer fibroblasts. Furthermore, there was overlap in many of the genes associated with the lining layer fibroblast subset identified in the study by Croft et al. (2019) (CLIC5, COL22A1, PRG4, and HBEGF). This was not associated with any particular pathotype, however PC3 did correlate with histological lining layer thickness.

Clustering of the samples on the UMAP data found three clusters. Although there was a slight trend towards more ldRA in one cluster (C3), there was again no significant separation by clinical group across the three clusters. However, as expected there were no normal samples in C3, which was the cluster with heavy immune infiltration. This is similar to the findings in the previous chapter when exploring histology and in the PCA, which suggests that broad exploration of overall variation in synovial tissue is unable to separate clinical groups.

When exploring the three clusters and their relationship to BEACON pathotype, there were low, mixed, and high inflammatory clusters, as shown by the predominantly lymphoid C3, predominantly pauci-immune C2, and the relatively even proportions of the three pathotypes in C1. C3 was defined by the presence of aggregates and C2 by the lack of infiltrating immune cells, while C1 had infiltrating immune cells but in the absence of large aggregates. This was further confirmed by exploration of gene expression, with pro-inflammatory genes being associated with C3 and, from exploration of genes in the literature, broadly anti-inflammatory genes being associated with C2, although gene exploration identified a tissue remodelling phenotype and potential downregulation of lipid catabolic processes in the mixed cluster C1, which was not identified from clinical and histological exploration alone.

It could be interesting to explore the presence and severity of erosions across these clusters, to elucidate whether the tissue remodelling phenotype in C1 is associated with increased bone erosion, but this data was not readily available for the current study. Exploration of the cell types, if any, that are associated with these clusters may help to explain the gene signatures seen, for example the presence or absence of adipocytes in C1 compared to the other two clusters may help to explain the differences in lipid catabolism, although staining for lipid droplets may also help to answer this question if the signature is due to expression in other cell types. Identification of the predominant cell types present within each cluster could be achieved by exploring where the genes associated with each cluster are expressed in a single cell dataset, by staining for specific cell types in tissue sections, in particular those that cannot be easily identified by H&E staining, such as macrophages, or using techniques such as

CIBERSORTx (Newman et al., 2019) to computationally deconvolute cell types in the present dataset.

Interestingly, despite being conducted in a mixed inflammatory arthritis cohort, this broadly replicates what has been seen in similar previous studies in RA, with high, mixed and low inflammatory phenotypes being identified (Orange et al., 2018, Lewis et al., 2019). Furthermore, the presence of immune-related pathways, including immune cell signalling and chemokine pathways, was found in the high inflammatory subtype and cell adhesion and TBF-β signalling were found to be associated with the low inflammatory phenotype in the study by Orange et al. (2018), in agreement with the present study. Eight of the top ten most significant upregulated genes in C3 were also found to be significantly differentially expressed between the high inflammatory cluster and the others in this study. However, none of top ten genes from C1 or C2 overlapped with the mixed or low inflammatory clusters, despite the overlap in pathways seen between the low inflammatory subtypes, although the mixed subtype only resulted in 3 significantly DEGs in the study by Orange et al. This slight difference in agreement is likely due to the mixed cohort of the present study compared to the predominantly RA (123 RA and 6 OA) cohort in the Orange et al. study, although could also be due to differences in disease stage, as the present study includes early inflammatory arthritis and the Orange et al. study used samples from arthroplasty at late stage disease (Orange et al., 2018).

Wnt signalling was identified as being associated with C1, the mixed inflammatory cluster, via exploration of the top ten most significantly upregulated genes. Interestingly, Wnt pathways were associated with pauci-immune fibroid and diffuse-myeloid pathotypes in the study by Lewis et al.

(2019) suggesting some potential overlap between C1 and their pathotype definition. In addition, immune-related processes, in particular B cell activation, were found to be associated with C3 in the present study and the lympho-myeloid pathotype in the Lewis et al. study, as might be expected due to the overlap between the definition of the lymphoid pathotype in BEACON scoring and lympho-myeloid in the PEAC study, namely the presence of lymphocyte aggregates.

Altogether, overall variation in the dataset was driven predominantly by the cell types present within the tissue. However, the level of heterogeneity in cellular composition across inflammatory arthritides is, in itself, an interesting observation. This may suggest that there are common mechanisms driven by specific cell types across clinical groups, rather than one specific cell type causing any particular disease. It may also be that broad differences in cell type mask any more subtle changes in cellular subtypes or mechanisms across clinical groups. Therefore, in the next chapter, specific comparisons will be made to explore whether there are any differences that can be uncovered when not looking at broad variation within the data.

5 RNA SEQUENCING: CLINICAL COMPARISONS

5.1 INTRODUCTION

Following exploration of the sources of variation and confounders, specific comparisons were undertaken to explore pre-set questions around inflammatory arthritis. Five hypotheses were formed at the beginning of the study, which detailed questions that have clinical and mechanistic relevance and allow for the advancement of knowledge of inflammatory arthritis. These hypotheses were as follows:

1. Tissue biomarkers will discriminate between short duration RA and other groups presenting to BEACON at baseline with short duration.
2. There are genes and pathways that differ between short duration and long duration RA.
3. Biomarkers for good or poor outcomes in RA can be found in gene expression data.
4. There are distinct mechanisms that can be found in gene expression data between early RA patients and patients destined to resolve.
5. Early RA and resolving signatures signature can be identified by comparison with normal synovium.

In this chapter, hypotheses 1-3 will be explored, with the final two hypotheses being discussed in the next chapter exploring mechanisms of resolution.

The first hypothesis was set to explore the potential use of biomarkers to distinguish early RA from other clinical groups that present at clinic. This would

enable early identification of RA patients that do not yet meet current classification criteria and allow for the commencement of DMARD therapy early on. Furthermore, biomarkers specific for RA would prevent any other inflammatory arthritis patients from receiving RA treatment unnecessarily. This is particularly important for patients whose inflammatory arthritis is destined to resolve.

The second hypothesis explores the previously described 'window of opportunity' for RA treatment, which is the idea that commencement of treatment very early on in RA disease course results in better outcomes long term (van Nies et al., 2015, Monti et al., 2015). We hypothesised that differential gene expression is responsible for this difference in treatment response, and that this could be utilised to further our understanding of what determines response to treatment in RA. This could potentially result in new targets that improve responses in longer duration RA, while also increasing our knowledge of very early RA disease pathobiology.

The final hypothesis discussed in this chapter is around treatment response specifically in RA, irrespective of disease duration. This aims to explore mechanisms that differ between patients who respond following 12 months of treatment and those that do not. This may enable new targets to be explored that improve response rates in current non-responders, as well as highlight potential biomarkers for stratification of patients by response.

5.2 BIOMARKERS OF EARLY RA

The aim of exploring biomarkers of early RA (sdRA) was to aid identification of early RA patients when they present at clinic. A number of patients initially

present with undifferentiated arthritis (UA) who have inflammatory arthritis that does not meet classification criteria for any particular disease. In the present cohort, just over 10% of sdRA patients (2 out of 18) initially presented with UA but later met classification criteria for RA at 18-month follow-up, with 20% of patients (17 out of 85) in the whole cohort initially presenting with UA. This can result in delayed treatment and uncertainty for the patient. Therefore, it would be beneficial to have biomarkers that could distinguish early RA from other early inflammatory arthritides, to enable diagnosis of RA patients who initially present as UA earlier in the disease course. To explore this, differential expression was performed on sdRA compared to resolving and nonRA patients grouped together.

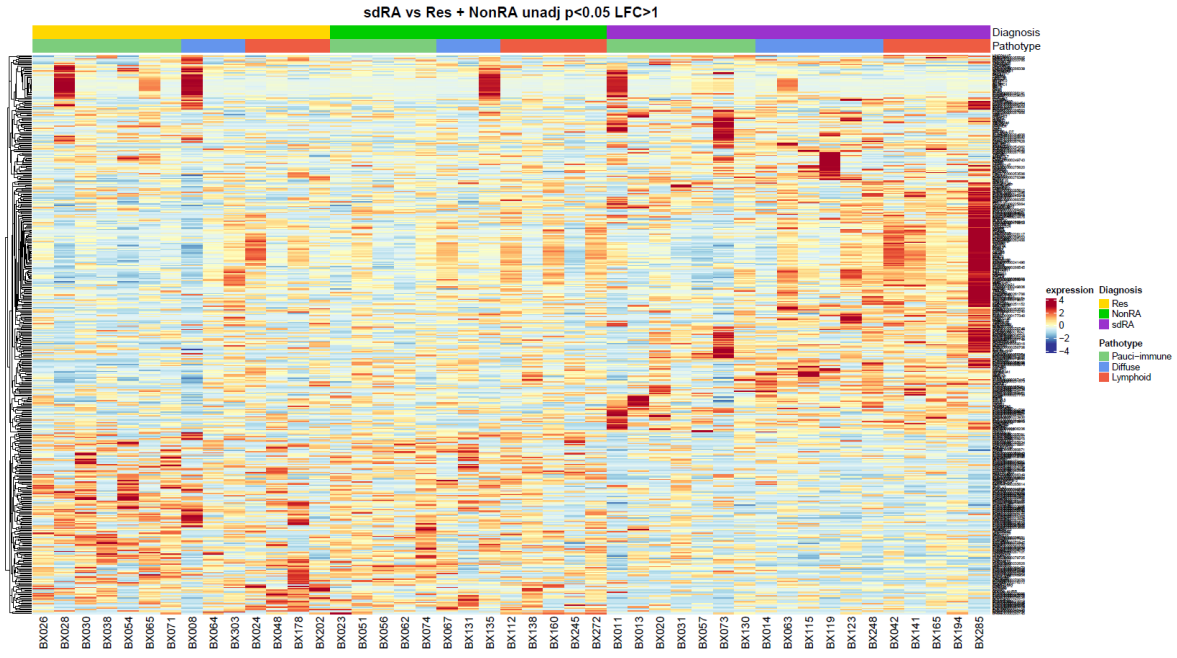


Figure 5.1: Heatmap of DEG from sdRA vs Res and NonRA. Unadjusted p<0.05, absolute LFC>1. Bars along the top show clinical group and BEACON pathotype.

This analysis discovered no significant differentially expressed genes. Cut-offs were loosened to include genes with an unadjusted p-value less than 0.05 and

absolute log-fold change greater than 1 both to explore how noisy the data were and to potentially allow for the exploration of pathways that could be associated with these genes. Figure 5.1 shows the resulting heatmap, including 481 genes (166 downregulated and 315 upregulated). Although there does appear to be some separation between sdRA and the other groups, the data show significant noise. Therefore, this comparison was not explored further. Other methods are likely to be required to find biomarkers due to the very heterogeneous nature of the groups including resolving and other persistent inflammatory arthritis.

5.3 MECHANISMS OF SHORT DURATION VS LONG DURATION RA

We hypothesised that short and long duration RA would be associated with different genes and pathways, due to the presence of a treatment 'window of opportunity' (van Nies et al., 2015, Monti et al., 2015). Therefore, ldRA was compared to sdRA using differential expression analysis. Five genes were identified as being differentially expressed (Figure 5.2), with four being upregulated in ldRA (MTCO3P12, ENSG00000203286, NOS2, and USP43) and one being downregulated (RPL3P7). RPL3P7 and MTCO3P12 were both predominantly driven by high expression in a single sample, so it is unlikely these genes would be biologically relevant. This therefore leaves only three genes that are upregulated in ldRA compared to sdRA. This led us to reject our hypothesis that differences in gene expression could be found between short and long duration RA using bulk, whole tissue RNA sequencing in this cohort.

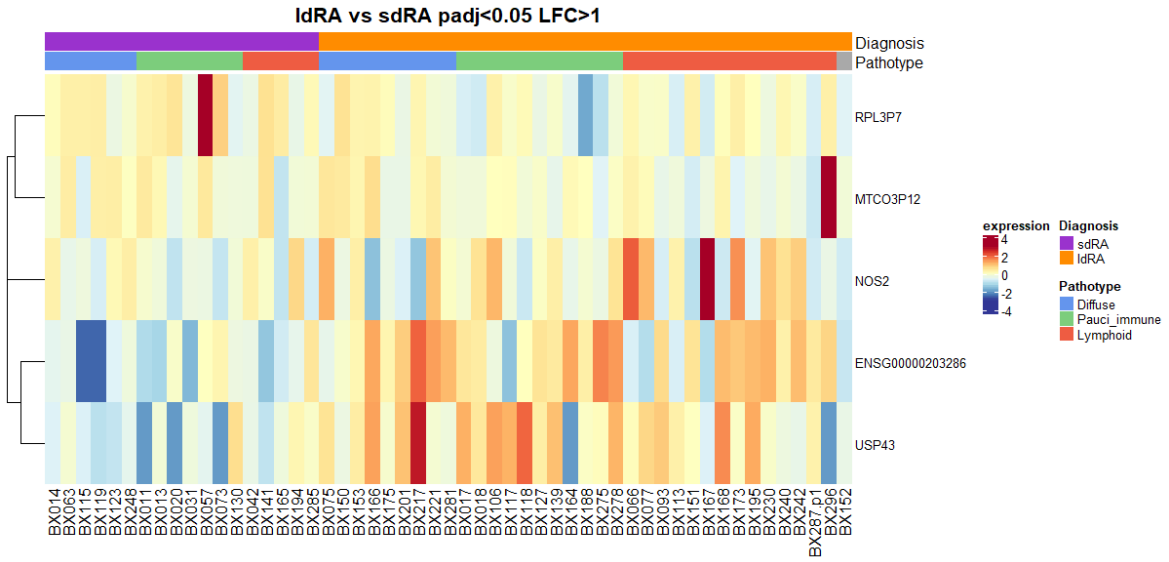


Figure 5.2: DEG between ldRA and sdRA. Adjusted p<0.05, LFC>1. Bars along the top show clinical group and pathotype.

5.4 Biomarkers of response

In order to identify biomarkers that are predictive of response to treatment in RA, EULAR DAS28-ESR response (good or moderate) at 12 months was compared to those with no response. The treatment regimen followed standard of care NICE approved protocols but specific treatment may have differed between patients and was not defined by a clinical trial approach so data obtained reflected overall response, irrespective of treatment received. These data were only available for a subset of RA patients in the cohort (3 sdRA and 15 ldRA).

21 genes were found to be significantly differentially expressed, with 10 upregulated in responders and 11 downregulated (adjusted p<0.05) (Figure 5.3). Genes that were upregulated in responders were generally immune-related and, from exploration of the genes present in the AMP Phase I RA dataset (Zhang et

al., 2019), predominantly expressed in monocytes and T cells, while downregulated genes were generally expressed in fibroblasts (Figure 5.4). Six of the DEGs were not present in this dataset, which may be because they are expressed by cell types that were not included or they have too low expression for detection by single cell RNA sequencing.

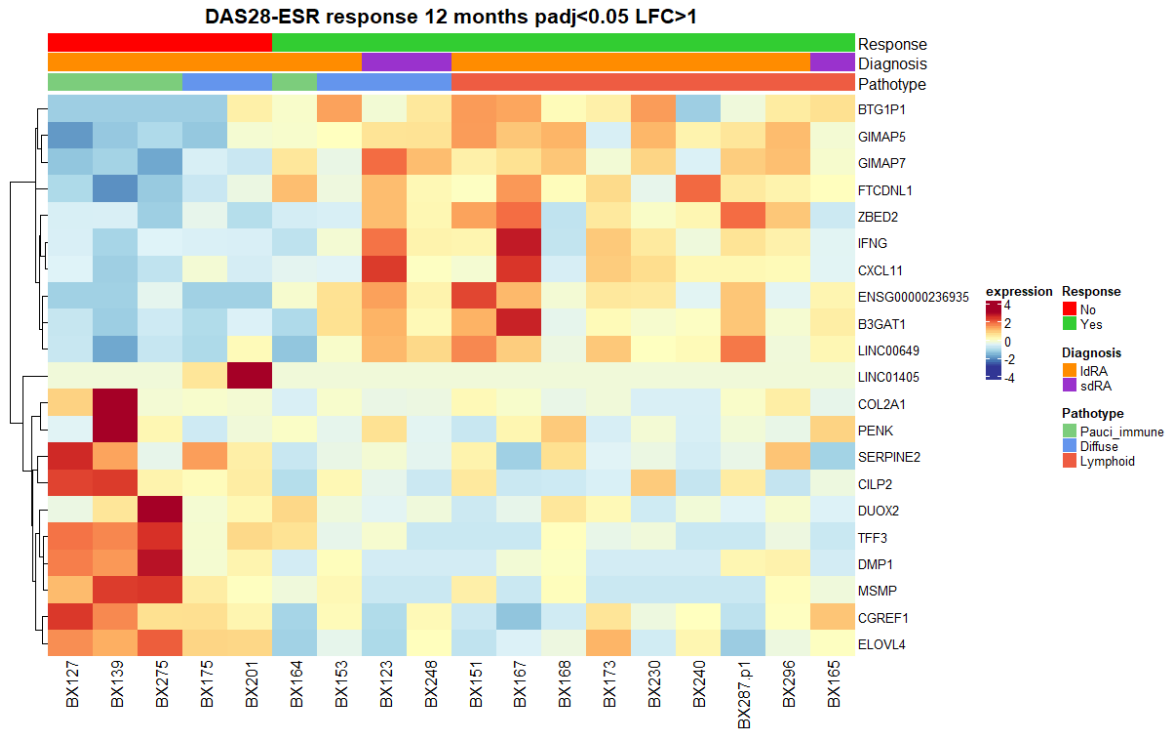


Figure 5.3: Heatmap of RA response DEGs. Adjusted p<0.05, absolute LFC>1. Bars along the top show EULAR DAS28-ESR response, clinical group, and BEACON pathotype. Response includes those with good and moderate response.

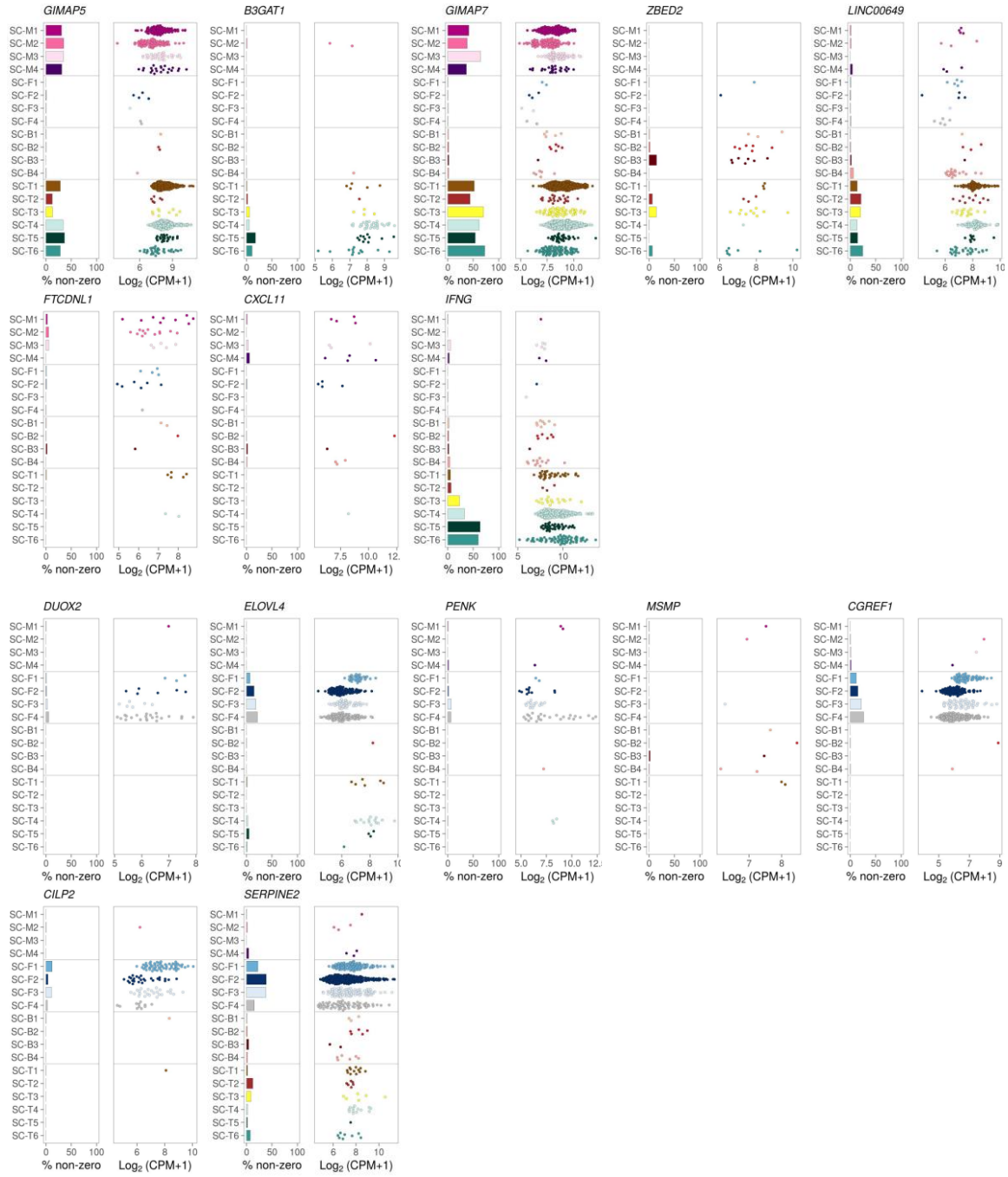


Figure 5.4: Exploration of RA response DEGs in the AMP Phase I RA dataset. GIMAP5, B3GAT1, GIMAP7, ZBED, LINC00649, FTCDNL1, CXCL11 and IFNG are upregulated in response. DUOX2, ELOVL4, PENK, MSMP, CGREF1, CILP2, and SERPINE2 are downregulated. Single cell RNA sequencing data of selected cellular populations (monocytes, fibroblasts, B cells & T cells). M1 = IL1B+ pro-inflammatory monocytes, M2 = NUPR1+ monocytes, M3 = C1QA+ monocytes, M4 = IFN-activated monocytes, F1 = CD34+ sublining fibroblasts, F2 = HLA+ sublining fibroblasts, F3 = DKK3+ sublining fibroblasts, F4 = CD55+ lining fibroblasts, B1 = IGHD+ CD270 naive B cells, B2 = IGHG3+ CD27- memory B cells, B3 = autoimmune-associated cells (ABC), B4 = Plasmablasts, T1 = CCR7+ CD4+ T cells, T2 = FOXP3+ Tregs, T3 = PD-1+ Tph/Tfh, T4 = GZMK+ CD8+ T cells, T5 = GNLY+ GZMB+ CTLs, T6 = GZMK+/GZMB+ T cells. Graphs from https://immunogenomics.io/ampra/ (Zhang et al., 2019).

As previous work in this study found that response may be associated with pathotype (Chapter 3.4) and UMAP clusters (Chapter 4.4.1), average expression of DEGs associated with response or non-response was explored across these in both RA groups, regardless of availability of treatment response data (Figure 5.5). In agreement with previous exploration, treatment response genes were significantly increased in C3 (high inflammatory) compared to C1 (mixed inflammatory/tissue remodelling) and C2 (low inflammatory) ($p=0.005$ and $p<0.0001$, respectively), with non-response genes being higher in C1 and C2 compared to C3 ($p=0.0002$ and $p=0.003$, respectively). Lymphoid pathotype was associated with significantly higher expression of response genes compared to pauci-immune ($p=0.0008$), with the pauci-immune pathotype having higher expression of the non-response genes than lymphoid ($p=0.01$). The diffuse pathotype had expression between lymphoid and pauci-immune and was not significantly different to either. This is in agreement with previous exploration of proportions of responders and non-responders between pathotypes and also with expression seen across cell types in the AMP dataset, which does not have response data but allows for exploration of gene expression in RA synovial monocytes, fibroblasts, B cells, and T cells.

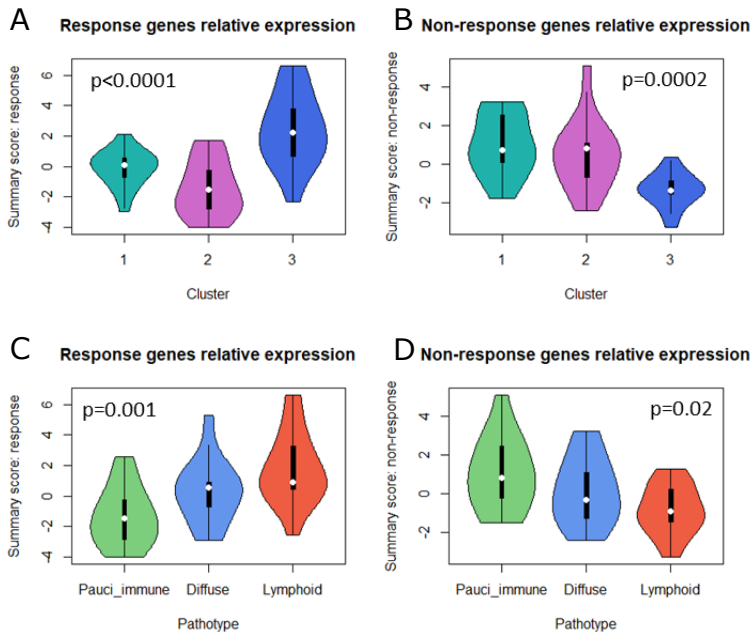


Figure 5.5: Average expression of treatment response and non-response genes across UMAP clusters and pathotypes. UMAP clusters (A & B) were derived in Chapter 4.4.1. Pathotype (C & D) is the histology BEACON pathotype score. Genes included in response (A & C, n=10) and non-response (B & D, n=11) are shown in Figure 5.3. Only sdRA (n=16) and ldRA (n=31) samples were included in this analysis (total n=47).

DEGs were also explored in the publically available PEAC RNA sequencing dataset (Lewis et al., 2019). The interactive website allows for comparison of gene expression against histology, including their lympho-myeloid, diffuse-myeloid, and pauci-immune fibroid pathotypes, clinical data, radiology, and response. Only one of the response DEGs (DUOX2) was found to be associated with EULAR DAS28-ESR response at 6 months in the PEAC dataset, with it being increased in non-responders in both datasets. However, similarly to the AMP data and current study, most of the genes upregulated in responders were increased in the lympho-myeloid pathotype and non-response genes were generally decreased in lympho-myeloid pathotypes, although expression in diffuse-myeloid and pauci-

immune fibroid pathotypes was more variable (Figure 5.6 and Figure 5.7). Of the 21 DEGs, four were not found in the PEAC dataset.

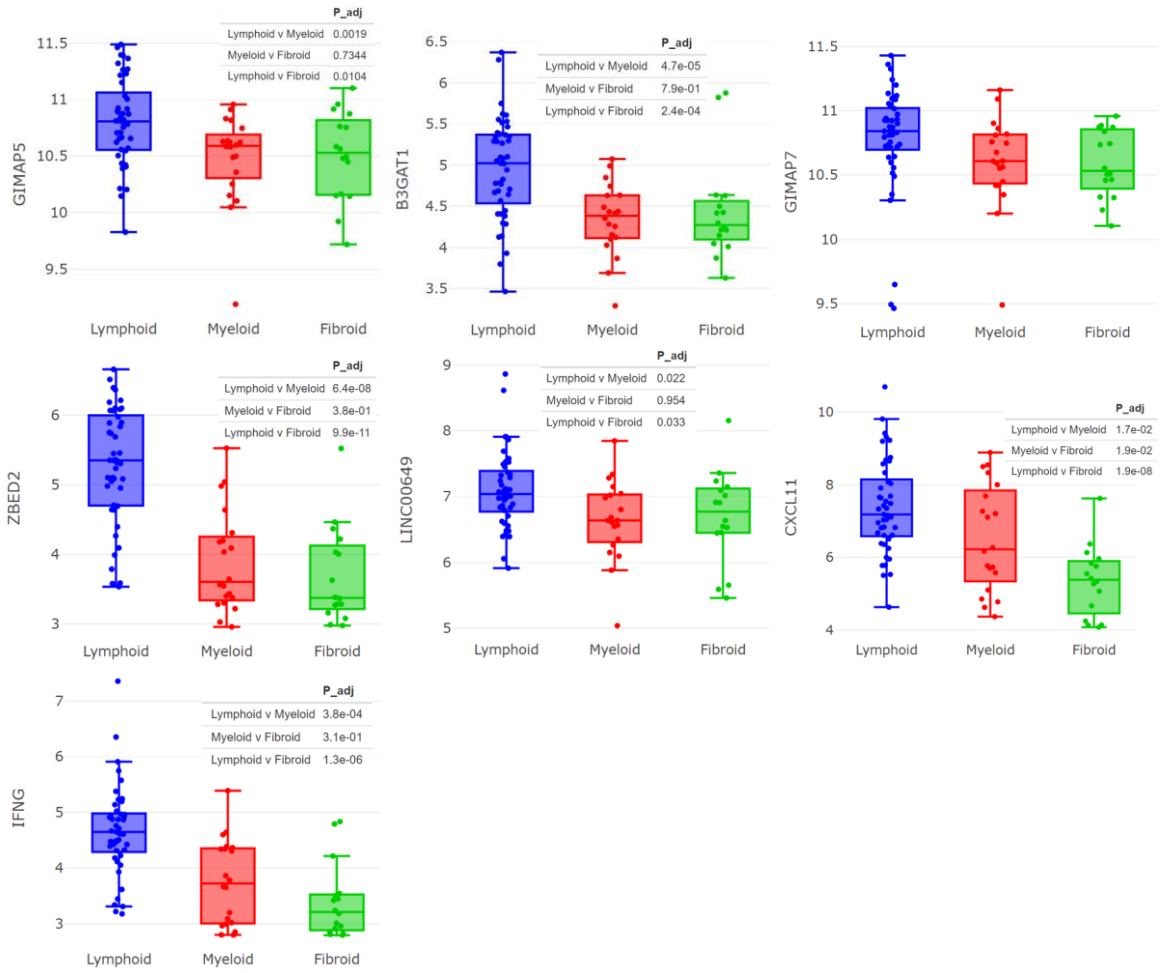


Figure 5.6: Upregulated DEG exploration in the PEAC dataset. Genes upregulated in responders were explored in the PEAC RNA sequencing dataset (Lewis et al., 2019) for expression in their pathotypes. Graphs are taken directly from the interactive website in the gene view section, accessable at https://peac.hpc.qmul.ac.uk/.

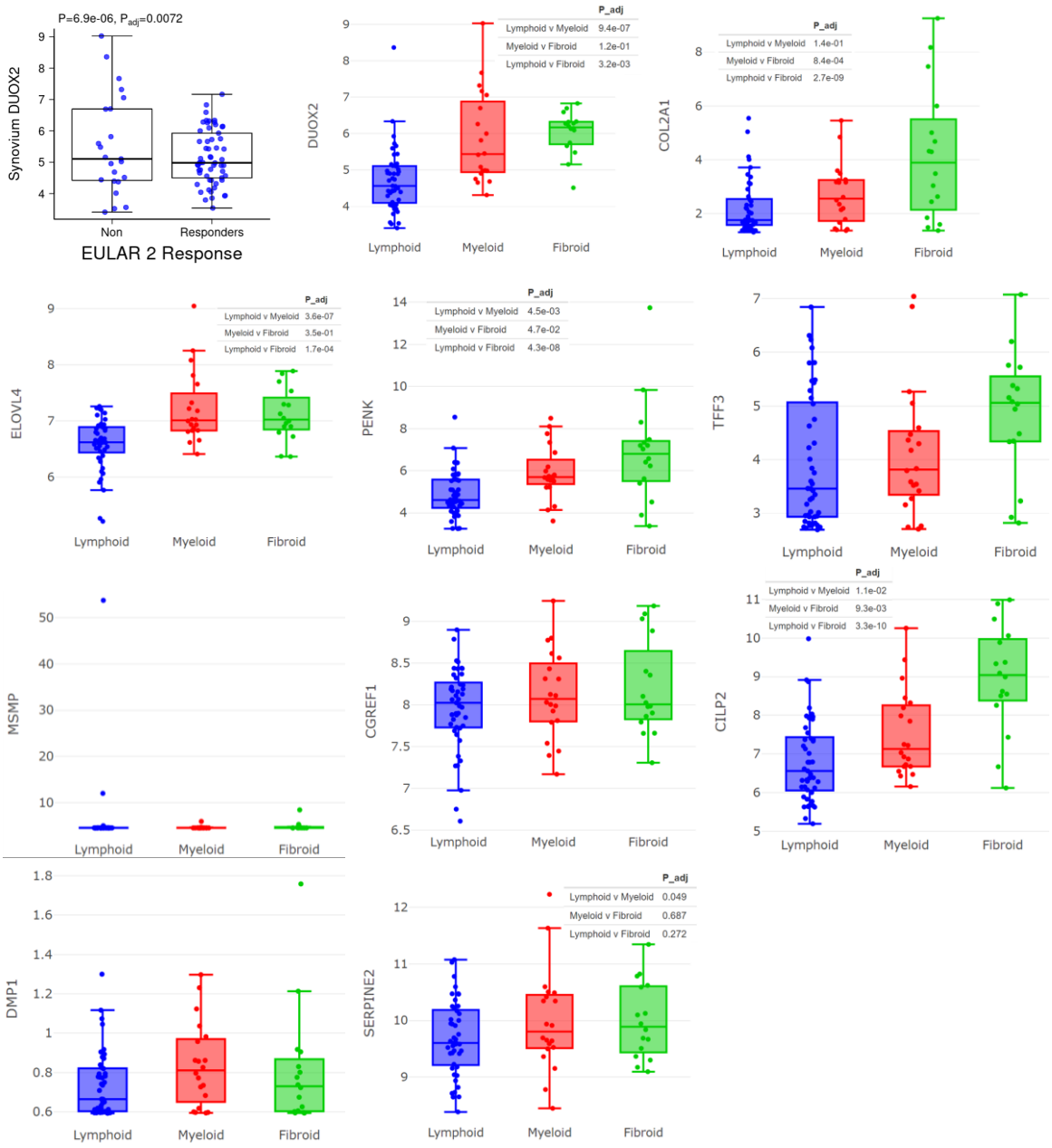


Figure 5.7: Downregulated DEG exploration in the PEAC dataset. Genes downregulated in responders were explored in the PEAC RNA sequencing dataset (Lewis et al., 2019) for expression in their pathotypes. Genes were also tested for association with DAS28-ESR EULAR response (good & moderate versus poor) at 6 months, DUOX2 was the only gene that had an association, so this graph is shown. Graphs are taken directly from the interactive website in the gene view section, accessable at https://peac.hpc.qmul.ac.uk/.

To further explore the differences between responders and non-responders, exploration of gene ontology (GO) biological processes that are associated with

the differentially expressed genes (DEGs) was explored. For this a less stringent cut-off using an unadjusted $p<0.05$ and absolute log-fold change greater than one (LFC>1) was used, resulting in 1288 genes. Genes that were upregulated (n=854) in responders were associated with predominantly immune-related processes, with the top processes being T cell activation, leukocyte cell-cell adhesion, and lymphocyte mediated immunity (Figure 5.8).

The processes associated with downregulated genes (n=434) were less clear in their biological significance, with ossification, collagen fibril organisation, skeletal system morphogenesis, and cartilage development being the most significantly associated processes (Figure 5.8). However, upon exploration of the genes associated with these processes in the AMP Phase I RA dataset (Zhang et al., 2019), and similarly to the significant DEGs, the majority of these genes were predominantly expressed in synovial fibroblasts, with many of them showing high expression across all fibroblast subsets (Appendix 9.6). Therefore, it is likely that these genes are representative of a synovial fibroblast signature rather than of ossification or collagen processes, as suggested by GO analysis. However, it is worth considering that the AMP dataset only includes pre-sorted monocytes, fibroblasts, B cells and T cells, and so excludes some cell types that are present in the synovium, such as endothelial cells, which may also express these genes. A list of all significant GO biological processes can be found in Appendix 9.5.2.

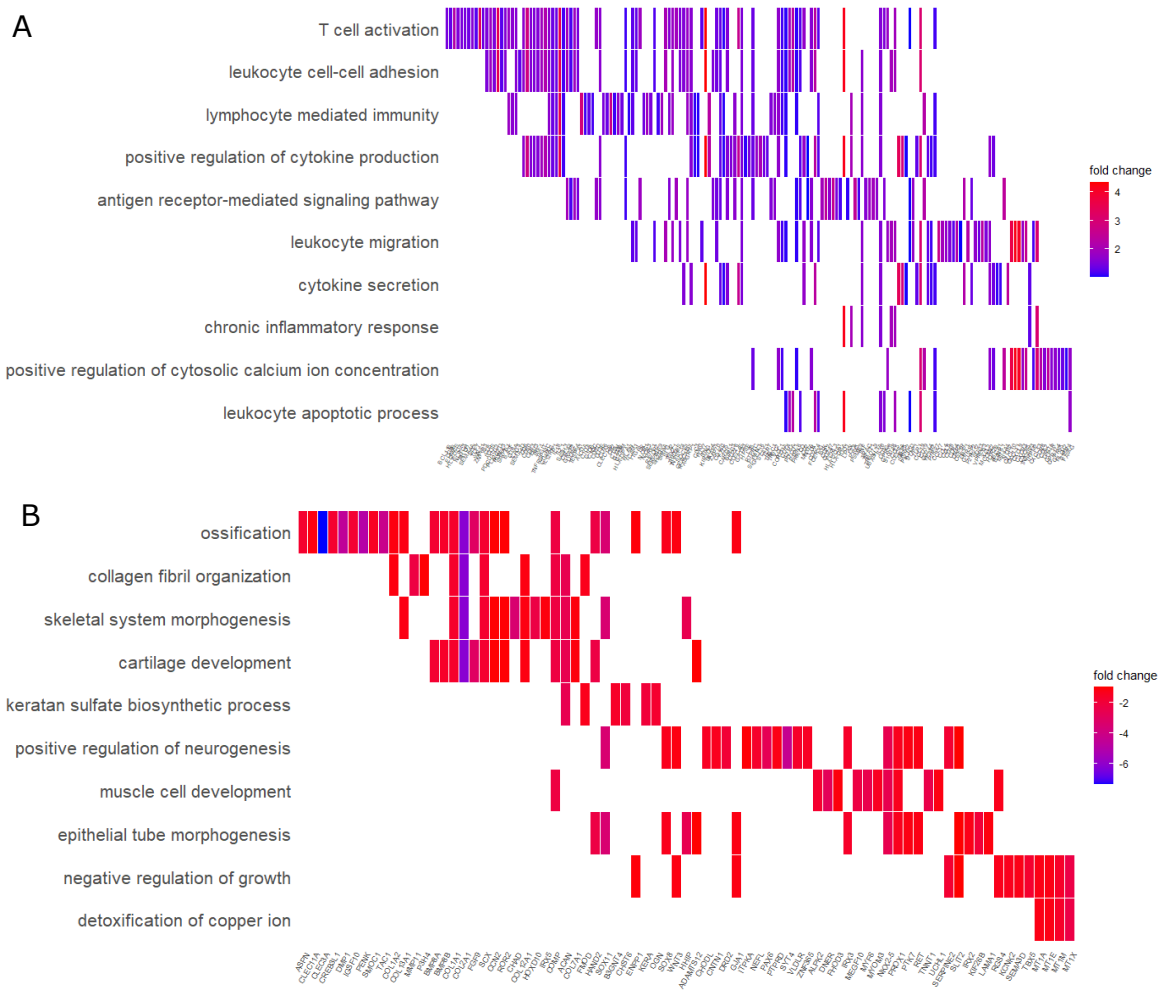


Figure 5.8: GO biological process analysis on RA response DEGs. Top 10 processes are shown. Shows processes from upregulated (A) and downregulated (B) genes. DEGs from unadjusted p<0.05, absolute LFC>1.

5.5 Discussion

5.5.1 Biomarkers of early RA

Differential expression analysis was unable to identify any significant DEGs between sdRA and other early inflammatory arthritides. Due to the very heterogeneous nature of the groups involved in this comparison, other methods may be better able to cope with the level of heterogeneity in this comparison. Weighted correlation network analysis (WGCNA) allows for the formation of gene

modules, which could be tested for differential expression between these groups (Langfelder and Horvath, 2008). This would reduce the number of multiple comparisons and increase statistical power, potentially overcoming some of the heterogeneity issues associated with this analysis. Another option could be the use of machine learning techniques, as was used in a previous study by Yeo et al. (2016), which also enables the capture of interactions between genes to identify gene expression profiles that distinguish between clinical groups. Alternatively, it may be that exploration of a subset of cells or the use of single cell RNA sequencing may be able to reduce some of the noise associated with whole tissue RNA sequencing and enable identification of biomarkers for very early RA.

5.5.2 Mechanisms of short duration vs longer duration RA

Exploration of the differential expression between short (symptom duration <3 months) and longer duration (symptom duration >6 months) RA resulted in three significant DEGs that may have biological relevance that were all upregulated in ldRA. ENSG00000203286 is a misc RNA, also known as AL441992.1. This gene has only been explored in one study, in which it was identified as an immune-related prognostic lncRNA for cervical cancer (Chen et al., 2020a). USP43, an ubiquitin peptidase, has also not been associated with rheumatoid arthritis directly, although a study into breast cancer suggested USP43 may be linked to the EGFR/PI3K/AKT pathway, which plays a role in RA pathogenesis (He et al., 2018, Yuan et al., 2013, Zhang et al., 2001). NOS2 (inducible nitric oxide synthase, iNOS), expressed in monocytes, has previously been associated with rheumatoid arthritis pathogenic mechanisms. It has been found to play a role in pro-inflammatory processes via the production of nitric

oxide (NO). NO may be involved in RA pathogenesis via its impact on bone erosion and its ability to drive T cells towards Th1 differentiation and IFNγ production (Nagy et al., 2010). It is also potentially linked to IL6 expression (Liu et al., 2018, Perkins et al., 1998, Grabowski et al., 1997).

Differential expression analysis did not give a clear separation between short and longer duration RA, leading to the conclusion that, from this dataset, we are unable to find differential mechanisms between these groups. Some previous studies have also not found differences between short duration and longer duration RA, finding no difference in the presence of immune cells by histology or immunoregulatory cytokines (Tak et al., 1997, Bucht et al., 1996). However, other studies were able to find differences in cytokines in both synovial fluid and tissue in early RA (Raza et al., 2005a, Yeo et al., 2016).

These results could suggest that, without the use of treatment, disease pathobiology remains fairly constant regardless of symptom duration, with similar mechanisms continuing to drive disease progression at both early and later stages, and with the predominant disease mechanisms depending more on pathotype than disease duration. This is further supported by the lack of significant differences between sdRA and ldRA when looking at histological variables, including BEACON density, aggregates, and pathotype. This may further highlight the potential for pathotype to inform treatment decisions, without a requirement for capturing disease at a specific point in time. However, there is still the possibility that the same mechanisms drive disease onset in all patients prior to symptom presentation, which cannot be elucidated from the current study.

It is also likely there are some differences between short and longer duration RA that have been masked by the noise associated with analysing whole tissue. There could also be differences on a level that cannot easily be detected in gene expression, such as at the level of protein translation, modification or degradation. It may be that other approaches that take a more specific view of different cell types or that investigate the proteome are better able to elucidate differences.

5.5.3 Biomarkers of response

Exploring the differential gene expression between RA responders and non-responders resulted in more differentially expressed genes and pathways than the previous two comparisons, with biological plausibility. There were a total of 21 significant DEGs. Exploration of these genes in the AMP Phase I RA dataset suggested an immune cell signature in responders and a fibroblast signature in non-responders. However, there are only selected cell types in this dataset so there may be other cells that express these genes that are not accounted for. This may also be the reason why six of the DEGs were not found in this dataset, although this could also be due to detection limitations for low expression levels in single cell RNA sequencing.

An immune cell versus fibroblast signature was also supported by the average expression of the genes increased in responders being significantly higher in the lymphoid pathotype compared to pauci-immune and vice versa, although fibroblast proportions are not accounted for in pathotype derivation. This was also true upon exploration in the PEAC dataset, in which genes increased in resolvers were generally increased in their lympho-myeloid pathotype. However,

it would be useful to follow this up in a single cell RNA sequencing dataset that includes all cell types found in synovial tissue to ensure no cell types are not accounted for and to assign expression more accurately to specific subsets.

IFNG and CXCL11 were both upregulated in responders. CXCL11 is induced by IFNγ, the protein product of IFNG, and binds CXCR3, which is expressed predominantly by activated T cells (Metzemaekers et al., 2017). CXCL11 induces immune tolerance via the induction of Th2 and Tr1 cells (Zohar et al., 2014). On the other hand, IFNγ is pro-inflammatory and generally associated with Th1 cells, although expression has also been found in Th17 cells, with IFNγ^{+} Th17 cells being elevated in the synovium and peripheral blood in RA (Paulissen et al., 2015). IFNγ has been shown to be increased in sera from RA patients compared to healthy controls and to correlate with disease activity (Brzustewicz and Bryl, 2015). Interestingly, CXCL11 has previously been found to predict poor response to adalimumab therapy, although this was in patients that had not responded to conventional therapy (Badot et al., 2009). Differences may therefore be due to prior treatment or may highlight a marker specific to adalimumab, as the present study was not solely looking at response to adalimumab.

GIMAP5 and GIMAP7 were upregulated in future responders, both of which are GTPases belonging to the immuno-associated nucleotide (IAN) subfamily. GIMAP5 has been linked to type 1 diabetes and systemic lupus erythematosus (SLE) and plays a role in lymphocyte survival. There have been conflicting studies around the role of GIMAP5 on T cell apoptosis, with some studies suggesting a protective role (Patterson et al., 2018, Chen et al., 2015, Pino et al., 2009, Chadwick et al., 2010), and another finding that over-expression induces apoptosis (Dalberg et al., 2007). It may be that GIMAP5 has differing

roles depending on concentration, availability of other factors, or level of T cell activation (Dalberg et al., 2007, Chadwick et al., 2010). Although less well studied, GIMAP7 has been suggested to be tumour suppressive in multiple recent studies, potentially by increasing the anti-tumour immune response (Xi et al., 2020, Song et al., 2019, Megarbane et al., 2020, Meng et al., 2020, Zhang et al., 2020a).

B3GAT1 (CD57) is expressed by subsets of NK cells, $CD4^+$ and $CD8^+$ T cells. $CD57^+CD8^+$ T cells are replication deficient and thought to be senescent, however are still able to produce cytokines and have cytotoxic activity (Kared et al., 2016). $CD57^+CD8^+$ T cells have been found in higher proportions in PBMCs from RA patients compared to healthy controls and decrease following abatacept treatment (Wang et al., 1997, Scarsi et al., 2010). In NK cells, CD57 is thought to be a marker of terminal differentiation, being expressed in mature NK cells (Kared et al., 2016). This subset of $CD57^+$ NK cells was found to be present in joints of inflammatory arthritis patients (Dalbeth and Callan, 2002).

ZBED2, which was also upregulated in responders, is a relatively unknown transcription factor. It has been shown in one recent study to be involved in pancreatic cancer by repressing IFN response through negative interaction with interferon regulatory factor 1 (IRF1) (Somerville et al., 2020). IRF1-deficient CIA mice were shown to have less severe disease and blocking of IRF1 reduced IL18, a cytokine involved in RA pathogenesis, in RA synovial fibroblasts, suggesting a potential role for IRF1 in RA (Tada et al., 1997, Marotte et al., 2011).

Taken together, this suggests a link between genes related to inflammation and response in RA patients. Interestingly, some of these genes have been shown to

be anti-inflammatory, with CXCL11 inducing immune tolerance and ZBED2 inhibiting IFN response via IRF1, although this is not the case for all genes identified (Zohar et al., 2014, Somerville et al., 2020, Tada et al., 1997, Marotte et al., 2011). These genes are predominantly expressed by T cells and monocytes, so this signature may be reflective of the cell types present in the tissue.

Overall, genes downregulated in responders compared to non-responders highlight a fibroblast signature, potentially suggesting varying disease mechanisms that are driven by different cell types between responders and non-responders. These findings are broadly in agreement with what was found in a previous study by Lewis et al. (2019), which found that gene modules associated with $CD8^+$ T cells, mast cells, and TLR signalling were increased in responders and that a $CD55^+$ fibroblast module was decreased. Of all the response DEGs, DUOX2 was the only gene that was also associated with response in this dataset. This may be due to the differing timeframe used for response, with the Lewis et al. study looking at EULAR DAS28-CRP response at 6 months and the present study using DAS28-ESR response at 12 months. Interestingly, DUOX2 was also found to be significantly upregulated in the UMAP cluster C2, which was associated with pauci-immune pathotype and low US GS score.

DUOX2 drives hydrogen peroxide production, a reactive oxygen species (ROS) that causes oxidative stress, and has been found to be upregulated in response to TLR4 signalling in inflammatory bowel disease, potentially driving an abnormal response to the microbiome (Burgueno et al., 2021). Hydrogen peroxide is increased in RA and oxidative stress has been suggested to contribute to cartilage destruction and joint destruction, even though oxidative stress

generally reduces T cell function (Wruck et al., 2011, Mirshafiey and Mohsenzadegan, 2008).

The signature seen in non-responders may be reflective of the current targets of RA therapy, which are predominantly focused on controlling inflammation via the targeting of immune cells and inflammatory mediators. Given that non-responders have a fibroblast related gene expression signature, current research efforts into synovial fibroblasts may yield new therapeutic targets that prove to be more effective for patients who are not responding to current treatment options (Siebert et al., 2020, Croft et al., 2019, Montero-Melendez et al., 2020, Diller et al., 2019). This further highlights the requirement for patient stratification prior to commencement of treatment, to ensure that the predominant disease mechanisms in each patient is being targeted and to avoid delays in commencement of the most effective treatment. This should also reduce the cost of RA treatment overall, avoiding the use of ineffective treatment and controlling disease at an earlier stage in those that will not respond to current first line treatment.

Exploration of the average expression of genes involved in response compared to non-response was able to provide some validation of previously non-significant associations between response, pathotype and UMAP clusters. Addition of all RA patients (n=47), rather than only those with available response data (n=19), granted increased statistical power. This allowed for the identification of significantly increased expression of response genes in lymphoid pathotypes compared to pauci-immune and UMAP cluster 3 compared to 1 and 2, with the opposite also being true. This further highlights the potential clinical utility of the BEACON histology scoring system, which could allow for the prediction of

treatment response, although this still requires further validation to confirm. Furthermore, this supports the idea that patient stratification using synovial tissue will allow for more informed treatment decisions and allow for more personalised medicine in RA.

Overall, the analyses in this chapter highlight some of the challenges associated with bulk, whole tissue RNA sequencing in inflammatory arthritis, with the large amount of heterogeneity in these patients providing a challenge for traditional differential expression analyses. Further exploration of these hypotheses using other analysis methods may add to these data. More specific questions addressing individual cell types, or the use of single cell RNA sequencing may allow for the identification of changes that are not detectable at the whole tissue level.

6 MECHANISMS OF RESOLUTION

6.1 INTRODUCTION

This chapter will explore the final hypotheses that were set at the commencement of the study, exploring mechanisms of resolution by comparing early RA with resolving disease and both of these groups separately to normal synovium. This will allow for characterisation of early RA and resolving signatures. Exploration of mechanisms that differ between these groups may allow for the identification of genes or pathways that drive resolution versus persistence, allowing for future targeting of these to block persistence and drive resolution of inflammation.

Current therapies for RA are able to suppress inflammation and reduce joint damage but are not curative and rarely allow for drug-free remission. This suggests that there are mechanisms that continually drive the persistence of RA, or an absence of mechanisms that actively drive resolution, that are not being targeted by current therapeutics. It was hypothesised that by comparing the gene expression in synovium of inflammatory arthritis that resolves with that of early RA, this would allow for the identification of these mechanisms, potentially leading to new therapeutic targets.

6.2 SHORT DURATION RA SIGNATURE

To explore the mechanisms involved in short duration RA, synovium from these patients was compared to normal tissue without inflammation to build a sdRA gene expression signature. Figure 6.1 shows the 689 genes that were identified

as being differentially expressed (padj<0.05, absolute LFC>1). Of these genes, 131 were upregulated in the uninflamed controls and 558 were downregulated so associated with sdRA.

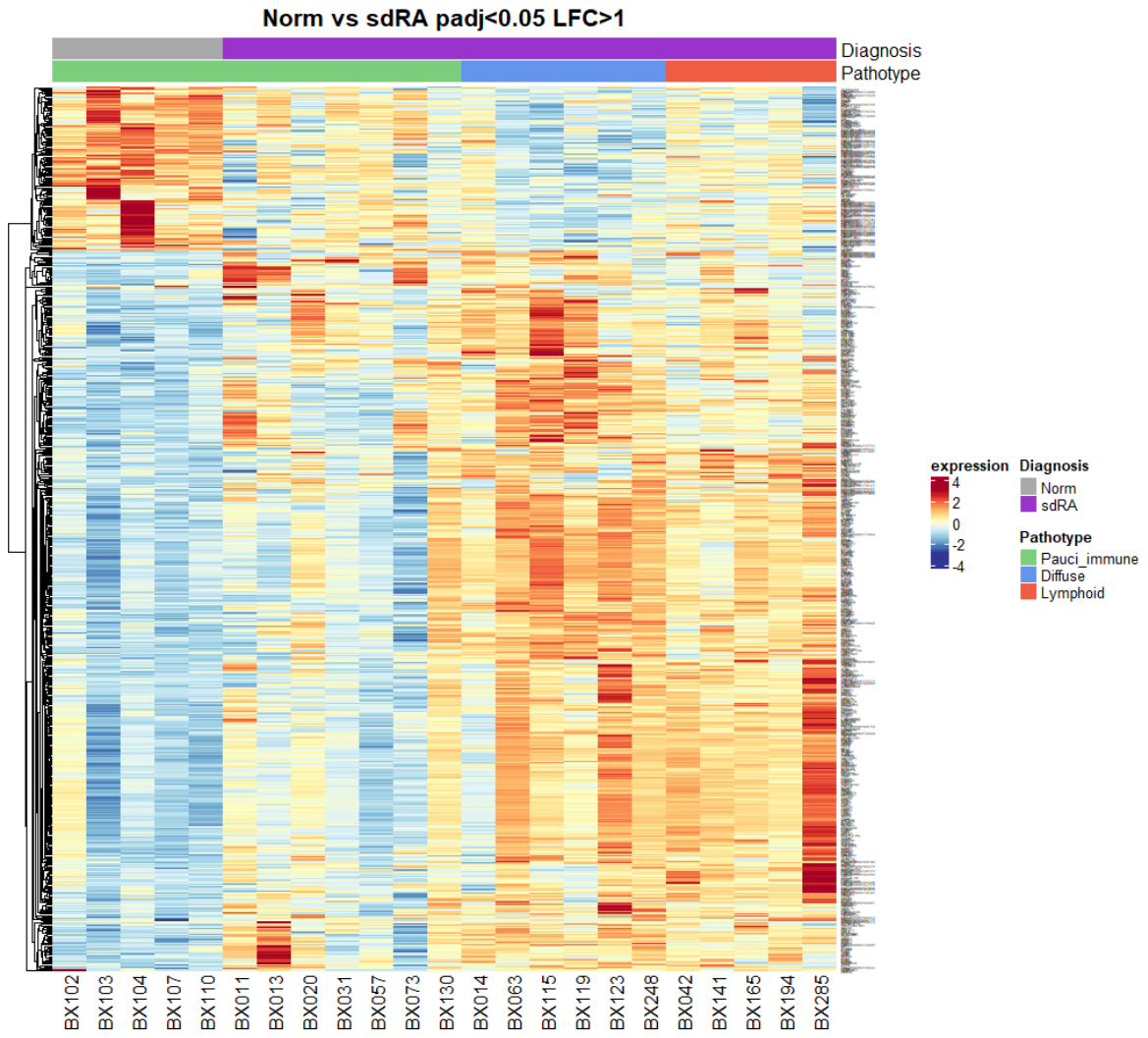


Figure 6.1: Differentially expressed genes between normal (uninflamed) control and sdRA. Bars along the top show clinical group and pathotype. Adjusted p<0.05, absolute LFC>1. Number of DEGs = 689.

As the number of DEGs was too large to explore in depth individually, exploration of the top 10 most significantly DEGs and GO pathway analysis on the full list of DEGs was undertaken. Figure 6.2 shows the GO biological processes that are associated with the DEGs that were downregulated in normal compared to sdRA.

This shows an inflammatory phenotype, as would be expected when comparing an uninflamed control with RA. The most significant process was T cell activation, followed by chemotaxis and chemokine pathways, showing activation and recruitment of immune cells, and suggesting the presence of a building or persistent activation of an immune response. There were pathways associated with both innate and adaptive immune responses, showing a full repertoire of immune activation in early RA. There were no pathways that were significantly associated with the genes upregulated in normal compared to sdRA samples. A list of all significant GO biological processes can be found in Appendix 9.5.3.

Figure 6.2: GO biological process analysis on normal vs sdRA DEGs. Top 10 processes are shown. Shows processes from genes upregulated in sdRA. DEG input from adjusted p<0.05, absolute LFC>1. Fold change shown on heatmap is the fold change of the gene between normal and sdRA. A full list of processes with associated genes can be found in Appendix 9.5.3.

6.2.1 Gene exploration

All of the top 10 most significantly DEGs were upregulated in sdRA compared to normal controls. IER3 (IEX-1) is involved in the regulation of Th1 and Th17 cell responses, with deficiency resulting in Th1 cell apoptosis and promotion of Th17

cell differentiation and survival (Zhi et al., 2012, Ustyugova et al., 2012). IER3 was previously found to protect from arthritis in mouse models and cultured cells, with knockout mice having more severe arthritis and knockdown being anti-apoptotic and increasing cytokine production in cultured RA synovial fibroblasts (Zhi et al., 2012, Morinobu et al., 2016). However, IER3 expression was actually found to be increased in RA synovial fibroblasts compared to OA in the same study (Morinobu et al., 2016).

DUSP2 induces apoptosis in response to oxidative stress and may decrease Th17 cell differentiation (Yin et al., 2003, Lu et al., 2015). However, the role of DUSP2 in RA is unclear, with one study finding that knockout mice were resistant to the development of an RA model and another finding that DUSP2 expression is pro-inflammatory (Jeffrey et al., 2006, Lu et al., 2015).

PIM1 expression in synovial fibroblasts has been found to promote proliferation, migration, and joint damage, and expression in $CD4^+$ T cells has been associated with Th1 cell differentiation (Ha et al., 2019, Aho et al., 2005). PIM1 has been found to be increased in $CD4^+$ T cells in RA compared to other early inflammatory arthritis and targeting PIM1 has been suggested to be therapeutic in RA (Anderson et al., 2019, Maney et al., 2021).

IL6, CXCL8, and CXCL2 are cytokines and chemokines that have established roles in RA via blockade of the IL6 receptor for the treatment of RA, CXCL8 being chemotactic and activating for neutrophils and being expressed by numerous immune cell types, and CXCL2 being produced predominantly by monocytes and also being chemotactic for neutrophils (Elemam et al., 2020, De Filippo et al., 2013).

Interestingly, SOCS3, which was also found to be differentially expressed, is anti-inflammatory and has been shown to downregulate IL6 responses, although expression of SOCS3 is upregulated in response to IL6, providing negative feedback for cytokine responses (Yin et al., 2015). In a mouse model for RA, reduction of SOCS3 resulted in more severe disease, with increased T cell and macrophage activation, as well as increased bone erosion (Wong et al., 2006).

LTBP2 encodes an extracellular matrix protein that regulates cell adhesion and is anti-angiogenic (Kan et al., 2015). LTBP2 has been found to inhibit maturation and increase production of inflammatory cytokines in adipocytes, which are reduced in number in RA compared to normal tissue (Srinivasa et al., 2021). A study by Nzeusseu Toukap et al. (2007) found that LTBP2 expression was reduced in SLE compared to RA and OA, although this could be due to an increased expression in RA, as suggested by the present study.

CR1 is a regulatory complement gene expressed in several immune cells that has previously been found to be reduced in RA B cells compared to healthy controls, with it inhibiting B cell activation and proliferation (Kremlitzka et al., 2013). However, CR1 was found to be increased in monocytes in RA compared to healthy controls during active disease (Hepburn et al., 2004). It may be that increased expression of CR1 in the present dataset is driven by expression in monocytes or other cells in the synovium that have not been previously described in relation to CR1 expression in RA. The final gene associated with sdRA was ENSG00000224114, which is a pseudogene.

6.3 RESOLVING SIGNATURE

Similarly, the gene expression signature associated with resolving disease was explored by comparing tissue from patients whose disease resolved with normal, uninflamed tissue (Figure 6.3). This resulted in fewer differentially expressed genes in total (n=116) than the comparison with sdRA (n=689). Again, more genes were downregulated (n=93) in normal tissues compared to resolvers than upregulated (n=23).

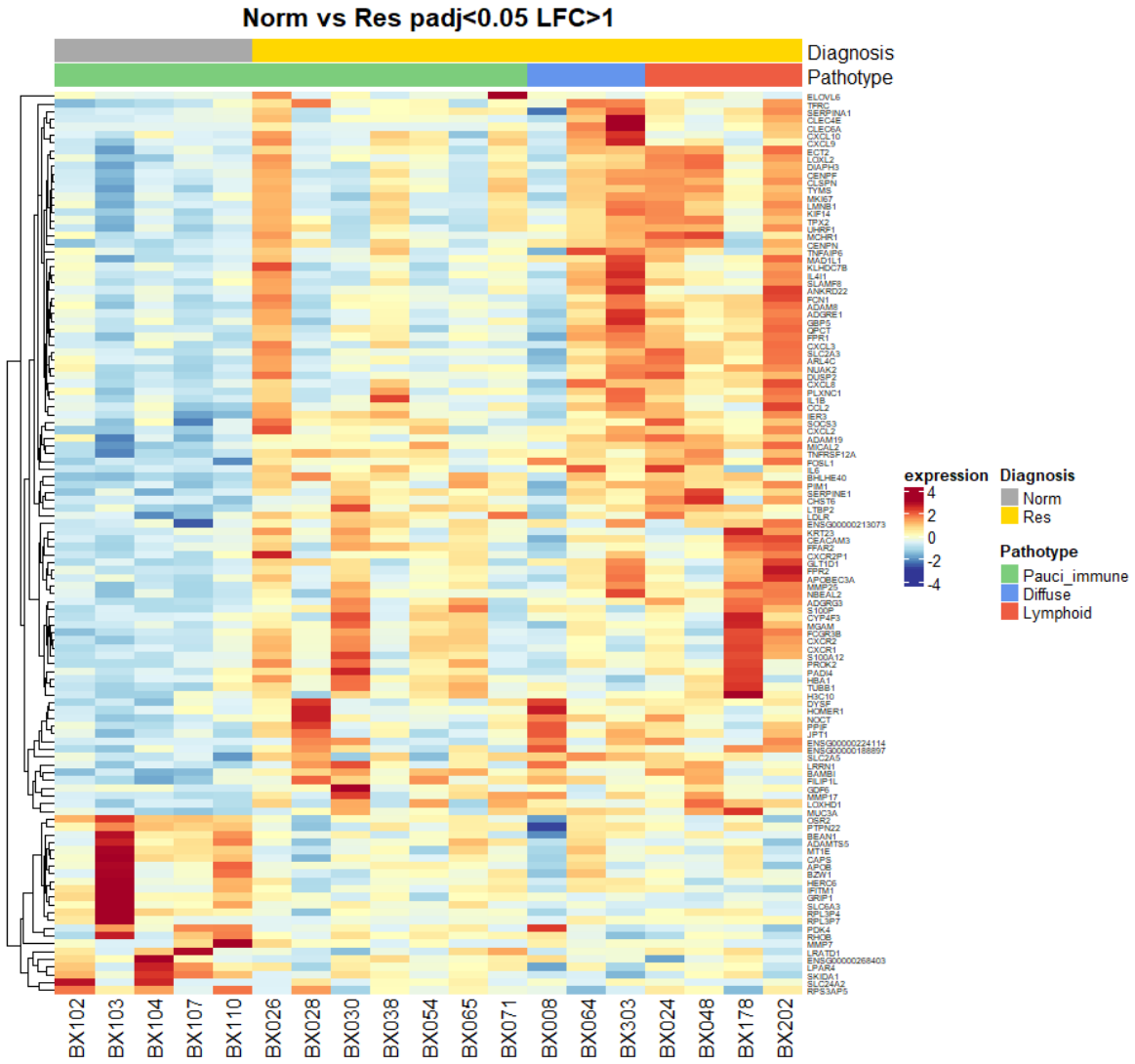


Figure 6.3: Differentially expressed genes between normal (uninflamed) control and resolving disease. Bars along the top show clinical group and pathotype. Adjusted p<0.05, absolute LFC>1. Number of DEGs = 116.

Exploration of the biological processes associated with the downregulated genes in normal compared to resolvers indicated involvement of a number of immune processes (Figure 6.4). However, the top pathways, including leukocyte chemotaxis, neutrophil activation, and neutrophil mediated immunity, were predominantly associated with innate immunity, with there being a relative lack of adaptive immune activation processes, although some adaptive immune

chemotaxis. This provides an interesting contrast to the sdRA signature, where the most significantly associated process was T cell activation. There were again no significant processes that were associated with the genes upregulated in normal compared to resolving samples. A list of all significant GO biological processes can be found in Appendix 9.5.3.

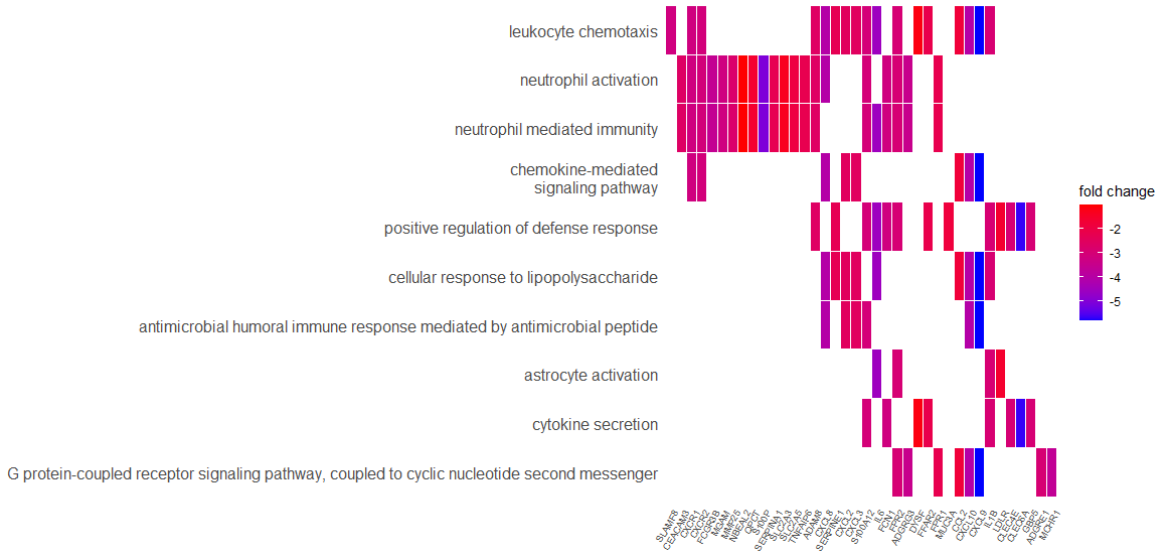


Figure 6.4: GO biological process analysis on normal vs Res DEGs. Top 10 processes are shown. Shows processes from genes upregulated in Res. DEG input from adjusted p<0.05, absolute LFC>1. Fold change shown on heatmap is the fold change of the gene between normal and Res. A full list of processes with associated genes can be found in Appendix 9.5.3.

6.3.1 Gene exploration

The top ten most significant DEGs were explored in the existing literature, eight of which were associated with resolving disease. The two genes upregulated in normal controls compared to resolving disease were RPL3P7, which is a pseudogene, and GRIP1. GRIP1 has been suggested to have a role in glucocorticoid signalling, being a cofactor of glucocorticoid receptor, although GRIP1 can also activate pro-inflammatory IFN-regulatory factors (Reily et al.,

2006, Flammer et al., 2010). Posttranslational modifications have been shown to have an impact on the role of GRIP1, with phosphorylation being suggested to increase its interaction with and subsequent activation of glucocorticoid receptors, highlighting a drawback of the exploration of gene expression (Dobrovolna et al., 2012).

Of the top DEGs associated with resolving disease, IER3, IL6, PIM1, and CXCL8 were also included in the top 10 most significant DEGs associated with sdRA. BHLHE40, CXCL9, LRRN1, and FCGR3B make up the rest of the top DEGs for resolving compared to normal tissue. BHLHE40 regulates a number of processes in several different immune cell types, including $CD4^+$ T cells, in which it regulates proliferation and increases pro-inflammatory cytokine production while reducing production of anti-inflammatory IL-10 (Cook et al., 2020). It has also been found to have a role in regulation of mitochondrial metabolism, in particular in tissue resident $CD8^+$ cells (Li et al., 2019, Cook et al., 2020).

CXCL9 is a chemokine that is produced by synovial fibroblasts and is chemotactic for activated T cells and plasma cells, having known roles in both RA and PsA (Farber, 1997, Tsubaki et al., 2005, Penkava et al., 2020). Furthermore, reducing CXCL9 has been found to inhibit the activation of synovial fibroblasts in RA, highlighting a role of CXCL9 in inflammatory arthritis (Meng and Qiu, 2020).

LRRN1 has been found to have a key role in neuronal development and to be expressed in cancer but a role in inflammation has not been described to date. In cancer, LRRN1 has been found to be anti-apoptotic, with knock down inducing apoptosis in gastric cancer and pancreatic ductal adenocarcinoma cells and

expression being linked to poor prognosis in some instances (Liu et al., 2019, Zhang et al., 2021b).

FCGR3B has been linked to risk of RA and other inflammatory disease, with low copy number and deletion increasing risk (Rahbari et al., 2017, McKinney and Merriman, 2012). This may be due to its ability to remove immune complexes without the pro-inflammatory formation of neutrophil extracellular traps (Chen et al., 2012). However, a recent study by Chen et al. (2020b) identified FCGR3B as being upregulated in RA and ulcerative colitis, with a decrease being seen following treatment in ulcerative colitis. Hence, the role of FCGR3B in inflammatory arthritis is unclear and requires further study.

6.4 Differential expression

Differential expression analysis was then undertaken directly comparing resolving disease to early RA (sdRA). There was only one significant gene with padj<0.05, so genes with an adjusted p<0.2 and absolute LFC>1 were explored. This resulted in 15 DEGs, with six being upregulated and nine downregulated in resolvers (Figure 6.5). One of these genes was the pseudogene MTND1P2, which was predominantly expressed in a single sample so likely lacks biological relevance.

In order to explore pathways associated with DEGs, genes with an unadjusted p<0.05 and absolute LFC>1 that had associated Entrez IDs were used as input to GO biological process analysis (Figure 6.6). Figure 6.7 shows the biological processes that were associated with upregulated and downregulated genes. The top processes downregulated in resolvers compared to sdRA included positive regulation of leukocyte activation, positive regulation of cell activation, regulation

of lymphocyte activation, and regulation of humoral immune response. This suggests more prominent persistent immune activation occurring in sdRA than resolving disease.

On the other hand, processes upregulated in resolvers were mostly involved in amine transport and fatty acid metabolism, with the top three processes being regulation of amine transport, amine transport, and fatty acid derivative metabolic process. This may suggest differential metabolic processes occurring between these two clinical groups.

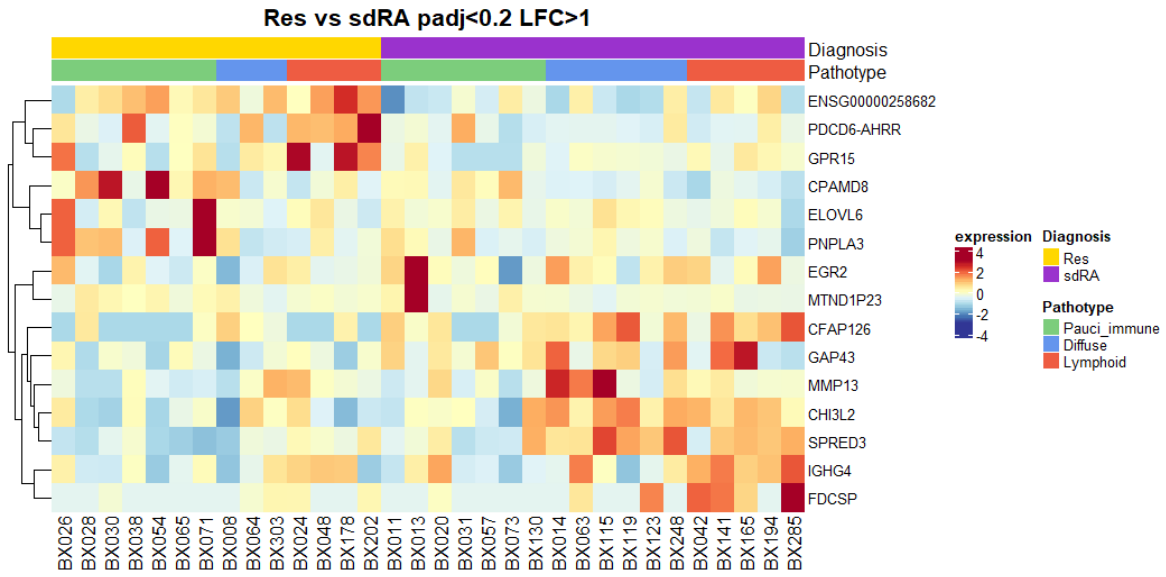


Figure 6.5: Differentially expressed genes between resolving inflammatory arthritis and sdRA. Bars along the top show clinical group and pathotype. Adjusted p<0.2, absolute LFC>1.

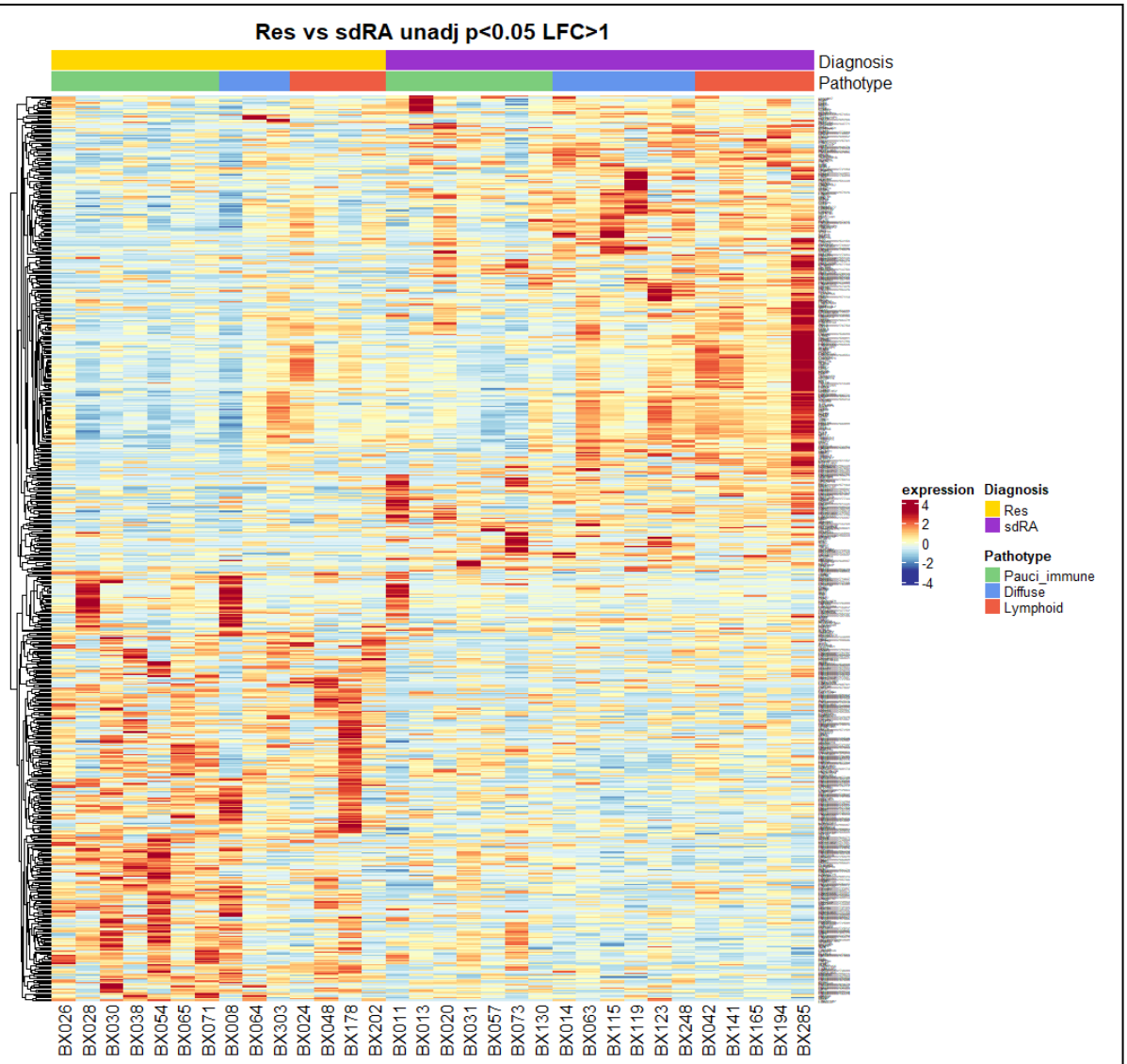


Figure 6.6: Differentially expressed genes between resolving inflammatory arthritis and sdRA. Bars along the top show clinical group and pathotype. Unadjusted p<0.05, absolute LFC>1. Number of DEGs = 665.

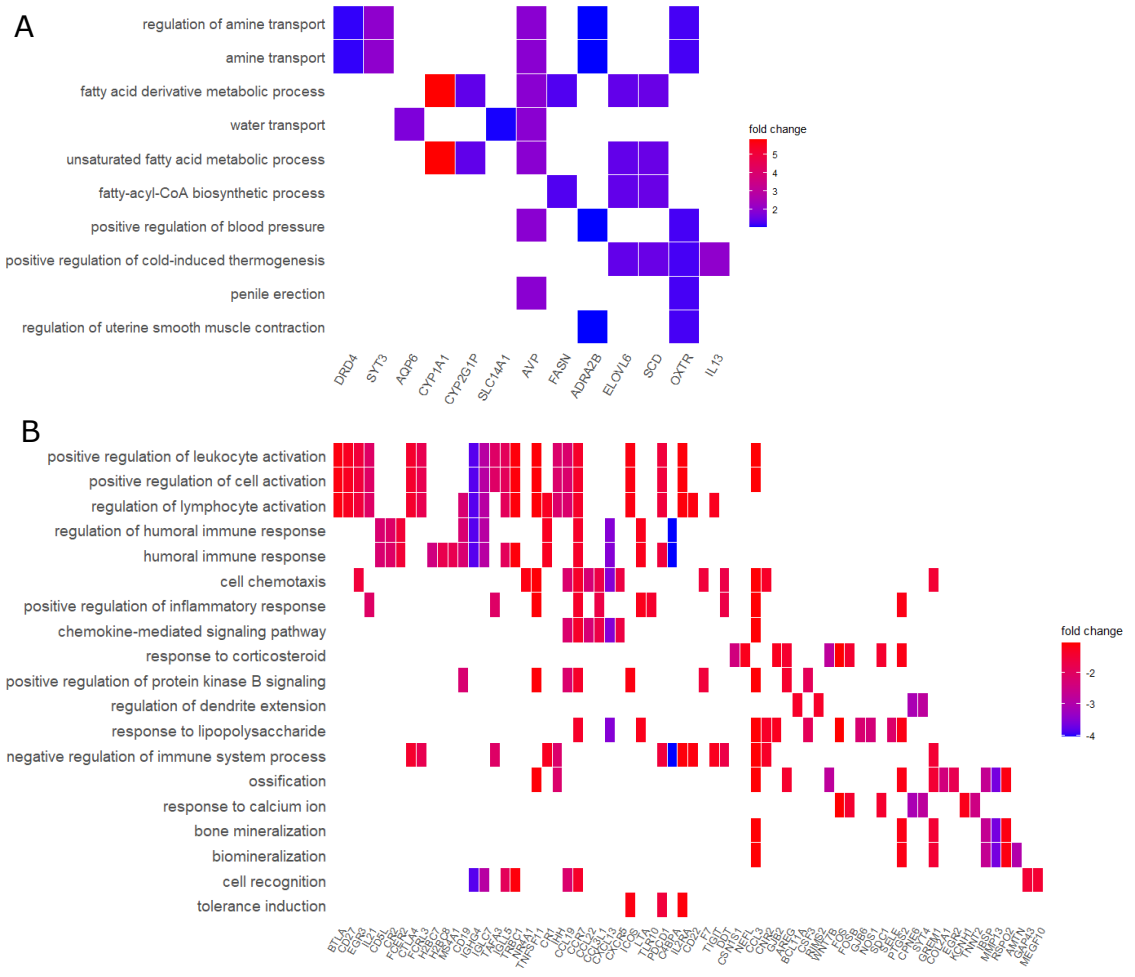


Figure 6.7: GO biological process analysis on resolving vs sdRA DEGs. All significant processes are shown (p<0.05). Shows processes from upregulated (A) and downregulated (B) genes. DEG input from unadjusted p<0.05, absolute LFC>1. Fold change shown on heatmap is the fold change of the gene between resolving and sdRA.

6.4.1 Gene exploration

Genes with padj<0.2 and absolute LFC>1 were explored further (n=15). Of these, nine were downregulated (MTND1P23, FDCSP, MMP13, CFAP126, CHI3L2, GAP43, SPRED3, IGHG4, and EGR2) and six were upregulated (ENSG00000258682, ELOVL6, PDCD6-AHRR, PNPLA3, CPAMD8, and GPR15) in resolvers. A literature search was undertaken to find similarities and differences

between these genes and to highlight areas where further exploration within the dataset may aid interpretation.

MMP13, a matrix metalloproteinase known to be involved in destruction of the joint in RA, was found to be downregulated in resolvers compared to sdRA. This finding helped to provide confidence in the results, as MMP13 has been extensively linked to RA pathogenesis (Wernicke et al., 1996, Moore et al., 2000, Konttinen et al., 1999, Burrage et al., 2006). Furthermore, a recent study by Liao et al. identified MMP13 as a marker for RA, although this was exclusive to fibroblasts and compared to OA, rather than other inflammatory arthritides (Liao et al., 2021).

FDCSP, also identified as being downregulated in resolvers compared to sdRA, is expressed by follicular dendritic cells and is thought to regulate IgA production and germinal centre responses (Marshall et al., 2002, Al-Alwan et al., 2007, Hou et al., 2014). Interestingly, FDCSP was found to be increased in samples with a lymphoid pathotype compared to both diffuse and pauci-immune ($p<0.05$ and $p=0.01$, respectively), supporting a link between FDCSP and lymphocyte aggregation (Figure 6.8A). FDCSP expression also appeared to be higher in the lympho-myeloid pathotype in the PEAC RNA sequencing dataset (Lewis et al., 2019), although statistics were unavailable for this (Figure 6.8B).

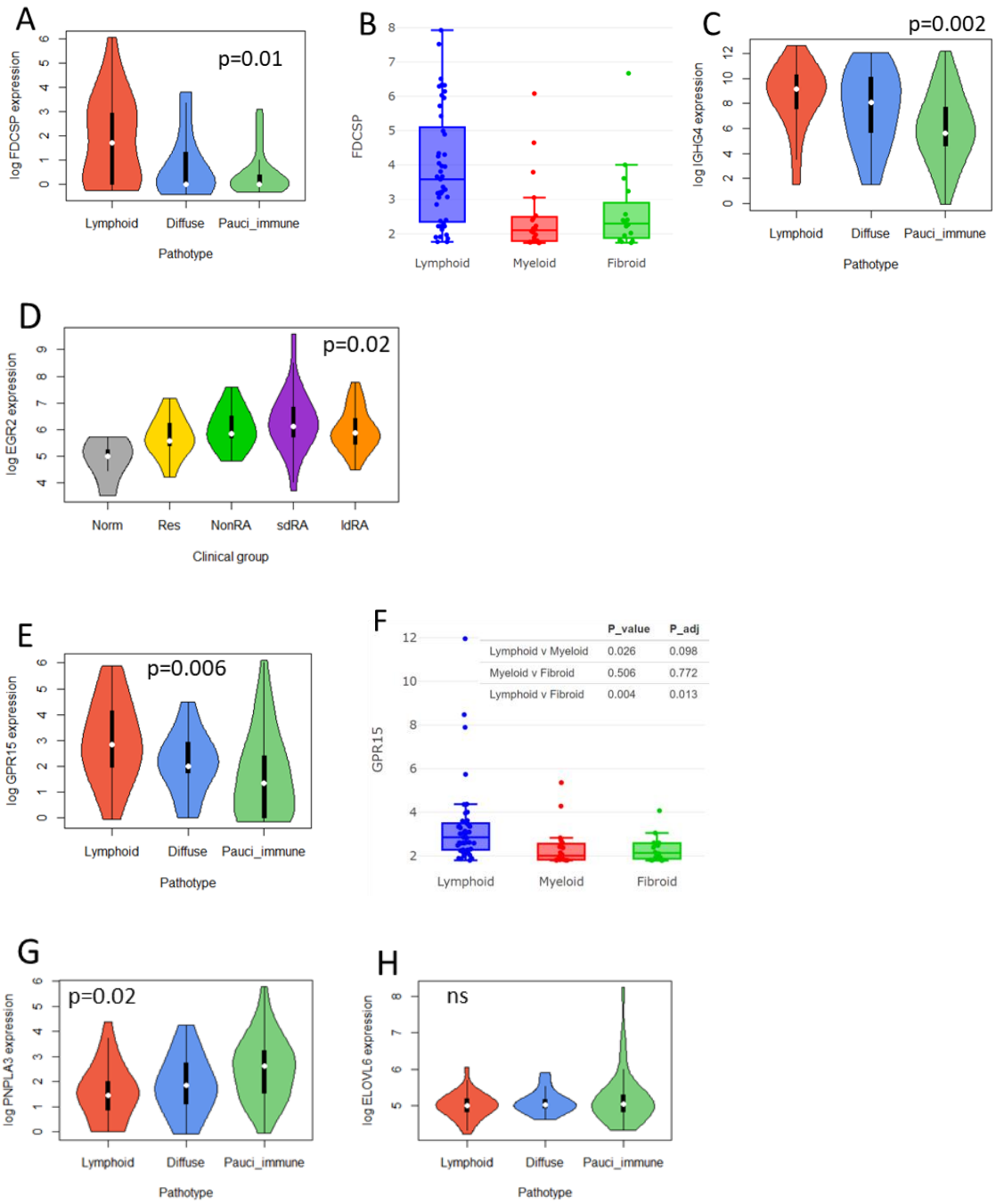


Figure 6.8: Gene exploration across pathotype and clinical group. Genes explored were differentially expressed between resolving and sdRA (unadj p<0.05, absolute LFC>1) and identified as potentially being linked to pathotype or clinical group following a literature search. Graphs B and F are copied directly from the PEAC publically available dataset (Lewis et al., 2019).

IGHG4 was also identified as being downregulated in resolving disease.

Increased serum IgG4 has been shown to be associated with higher disease

activity and poor response to methotrexate and leflunomide in RA patients, with serum levels also correlating with the presence of $IgG4^+$ plasma cells in the synovium (Chen et al., 2014). IGHG4 expression was higher in lymphoid pathotypes than pauci-immune ($p=0.002$) but no different from diffuse tissues (Figure 6.8C).

EGR2 can inhibit T cell activation and regulates Th17 differentiation, making it broadly anti-inflammatory (Safford et al., 2005, Miao et al., 2013). Contrary to the findings in the present study of EGR2 being decreased in resolvers compared to early RA, EGR2 has previously been found to be hypermethylated in RA synovial fibroblasts, resulting in reduced expression (Park et al., 2013). However, this was in cultured synovial fibroblasts alone, which have been found to have an altered phenotype compared to fibroblasts in the tissue, was from late stage disease, and was in comparison to OA and healthy controls, rather than other inflammatory arthritis (Wei et al., 2020, Casnici et al., 2014). Upon exploration of EGR2 expression across the clinical groups in the BEACON cohort, EGR2 expression was generally higher in early inflammatory arthritis when compared to normal tissue, although only sdRA and NonRA had significantly higher expression than normal controls ($p<0.02$ and $p<0.05$, respectively) (Figure 6.8D). One explanation could be that EGR2 is expressed in early RA but is switched off later in the disease course but it was not identified as being differentially expressed between sdRA and ldRA.

GAP43 has been found to be increased following nerve injury and in a collagen antibody-induced arthritis mouse model (Su et al., 2015b). It is associated with sprouting in neurons and may have a role in chronic pain that is not managed by anti-inflammatory treatment in RA (Goncalves Dos Santos et al., 2020).

Increased expression of GAP43 may suggest more nerve damage in sdRA when compared to resolving disease.

CHI3L2, SPRED3, and CFAP126 have not previously been associated with RA, although SPRED3 is expressed in the liver where it has been found to inhibit cellular proliferation (King et al., 2006). CHI3L2 has been found to be upregulated in OA and a related protein, CHI3L1 has been suggested to be an autoantigen in RA (Kzhyshkowska et al., 2007, Mazur et al., 2021).

Upon exploration of the genes that were upregulated in resolvers, GPR15 was identified as being differentially expressed. This is potentially in contrast to a previous study that identified GPR15 as being upregulated in RA, although this was on late-stage disease that had undergone treatment and was in comparison to uninflamed controls (Cartwright et al., 2014). In the same study, GPR15 was found to be expressed by monocytes and macrophages but not T or B cells in RA synovium but studies in other areas have found GPR15 expression in T cell subsets, including Treg cells from inflamed gut (Farzan et al., 1997, Fischer et al., 2016, Adamczyk et al., 2017). Unfortunately, there was no expression in the AMP phase I RA dataset (Zhang et al., 2019) for exploration of expression of GPR15 in specific cell types but expression was higher in tissues with a lymphoid pathotype than those with pauci-immune in our dataset ($p=0.004$, Figure 6.8E) and higher in lympho-myeloid than pauci-immune fibroid in the PEAC RNA sequencing dataset (Lewis et al., 2019) ($p=0.01$, Figure 6.8F), suggesting an association with lymphocyte infiltration and aggregation. Therefore, this might be indicative of increased expression of GPR15 in lymphocytes, potentially Treg cells, in resolvers compared to normal controls, although this would require more direct validation to explore further.

CPAMD8, PDCD6-AHRR and ENSG00000258682 were also upregulated in resolving disease. CPAMD8 has a potential role in innate immunity, is involved in glaucoma, and is associated with susceptibility for multiple sclerosis but has not been associated with inflammatory arthritis to date (Baranzini et al., 2009, Li et al., 2004, Wiggs, 2020, Bonet-Fernandez et al., 2020). CPAMD8 was also upregulated in the UMAP cluster C2, which was predominantly associated with pauci-immune pathotype. PDCD6-AHRR and ENSG00000258682 are both lncRNAs with unknown functions that have not yet been studied in any depth.

PNPLA3, also upregulated in Res, is able to inhibit adipose lipolysis and is thought to be involved in triacylglycerol (TAG) synthesis, meaning it broadly drives lipid synthesis (Yang and Mottillo, 2020, Yang et al., 2019). Furthermore, PNPLA3 has been found to have a role in lipid droplet formation (Chamoun et al., 2013). ELOVL6, a fatty acid elongase, was also found to be upregulated in resolvers compared to sdRA. Taken together, this suggests a difference in lipid metabolism between resolving inflammatory arthritis and early RA. Although PNPLA3 was upregulated in the pauci-immune pathotype compared to lymphoid ($p<0.02$, Figure 6.8G), there was no difference compared to diffuse, and there was no significant association between pathotype and ELOVL6 (Figure 6.8H). Therefore, the level and organisation of the lymphocytic infiltrate does not necessarily correlate with the difference in lipid metabolism.

6.5 Exploring metabolism

As transcription of genes involved in metabolic processes was identified as being potentially different between resolving disease and sdRA, further exploration of metabolic genes was undertaken. Genes involved in specific metabolic pathways

were curated by Dr Valentina Pucino and used for the exploration of average gene expression across different metabolic processes (Figure 6.9) (Pucino et al., 2019). Samples containing muscle were removed due to the inability to correct using the muscle score in this analysis and the distinct metabolic differences between muscle and synovium. There were no significant differences between resolving disease and sdRA for any of the metabolic processes tested in this way. Glucose transporters were higher in Res than sdRA ($p=0.01$) but significance was lost following Benjamini Hochberg correction for multiple comparisons. Even so, there was a trend towards increased lipid metabolism, including fatty acid synthesis and elongation, glucose metabolism, in particular increased glucose transporters, and amino acid metabolism in resolving disease.

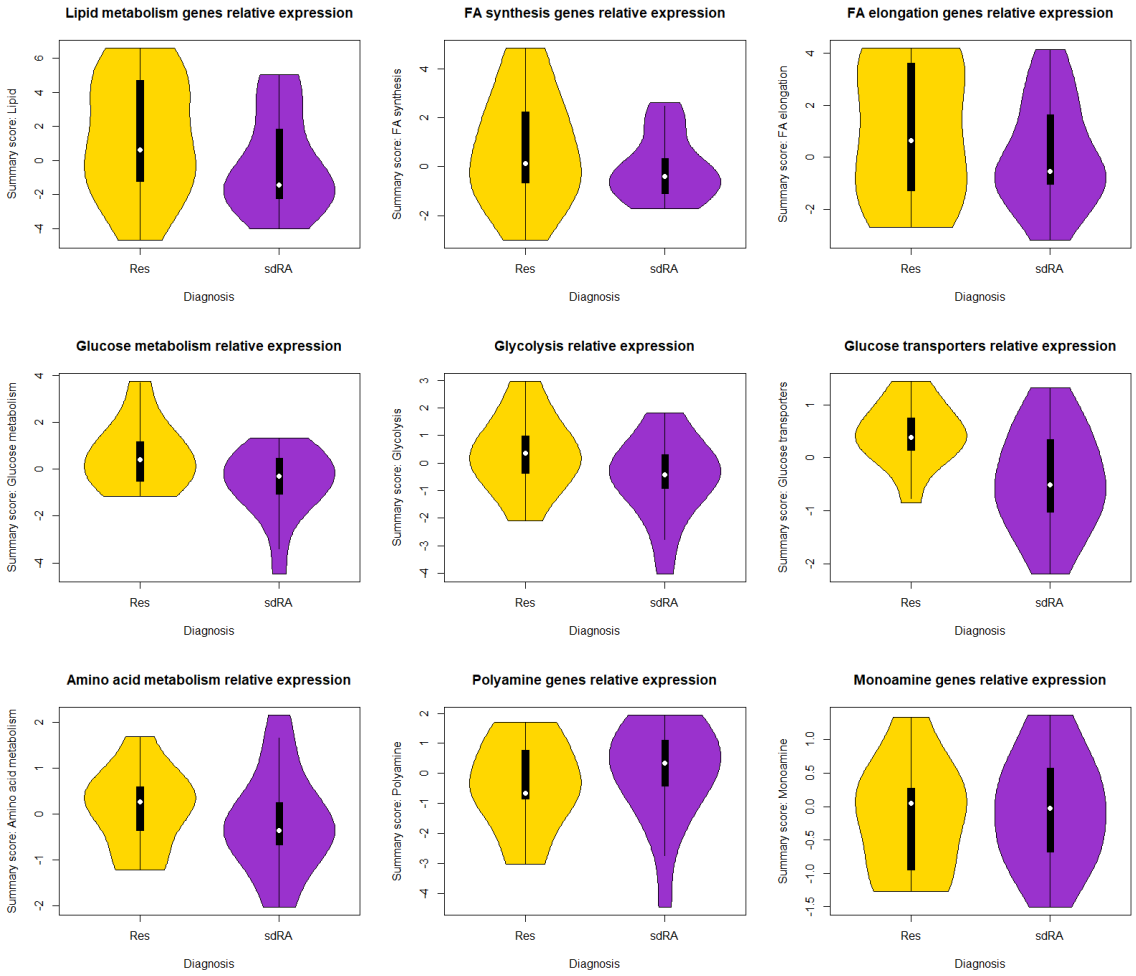


Figure 6.9: Relative expression for genes involved in metabolic pathways in Res and sdRA. Gene lists manually curated by Dr Valentina Pucino and can be found in Appendix 9.7. Relative expression calculated using technique developed by Wagle et al. (2018). Samples containing muscle were removed for this analysis.

Heatmaps of the genes involved in these metabolic processes were then explored, to explore in more detail the metabolic gene expression between resolving disease and sdRA. Figure 6.10 shows the expression of genes involved in fatty acid synthesis and fatty acid elongation and Figure 6.11 shows genes involved in glycolysis and glucose transporters. Hierarchical clustering of samples based on the expression of these genes showed that the majority of resolving samples clustered together based on FA synthesis and glucose transporter

genes, with less clear clustering by clinical group based on FA elongation and glycolysis genes. These heatmaps also highlight the influence of pathotype, with samples with the same pathotype often clustering together, particularly in the heatmaps based on FA synthesis and elongation.

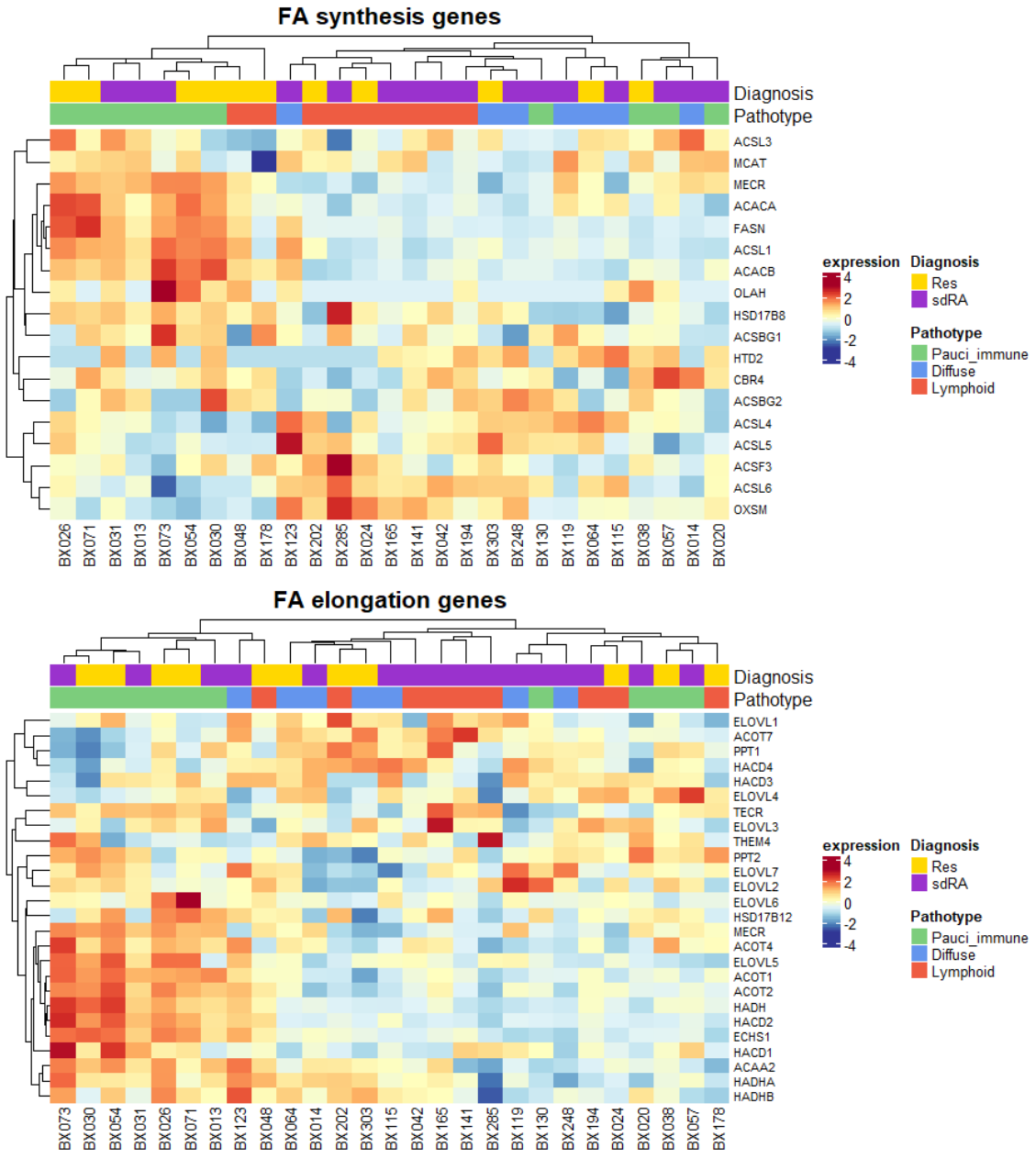


Figure 6.10: Heatmaps showing expression of genes involved in FA synthesis and FA elongation in Res and sdRA. Bars along the top show clinical group and pathotype. Samples were clustered using hierarchical clustering. Gene lists manually curated by Dr Valentina Pucino and can be found in Appendix 9.7. Samples containing muscle were removed for this analysis.

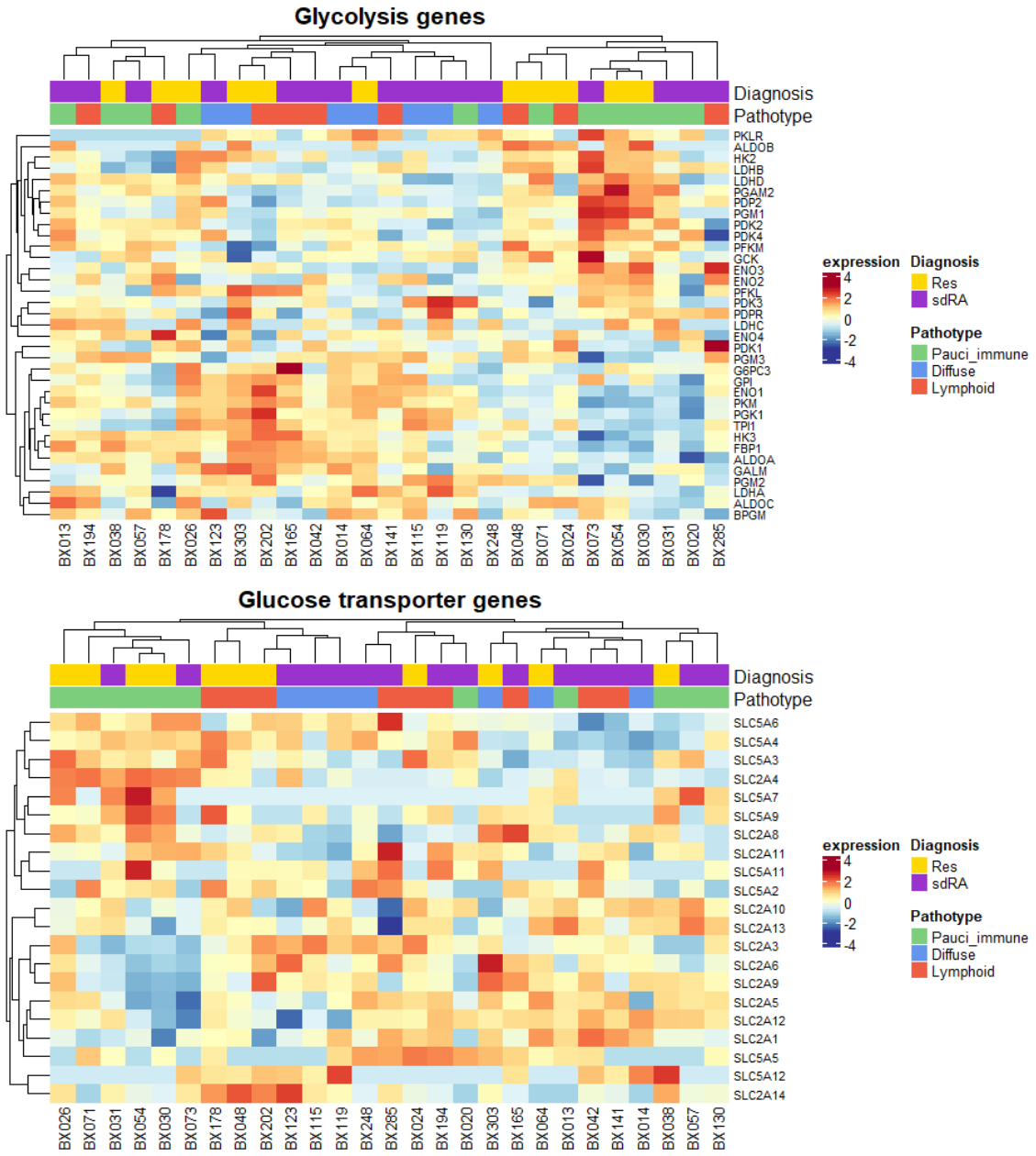


Figure 6.11: Heatmaps showing expression of genes involved in glycolysis and glucose transporters in Res and sdRA. Bars along the top show clinical group and pathotype. Samples were clustered using hierarchical clustering. Gene lists manually curated by Dr Valentina Pucino and can be found in Appendix 9.7. Samples containing muscle were removed for this analysis.

6.6 DISCUSSION

There were fewer DEGs identified when comparing resolvers to normal than sdRA to normal (n=116 and 689, respectively), which may suggest that patients that are destined to resolve have gene expression that is closer to normal synovium than those that have early RA. However, this may also be as a result of the fewer numbers of resolving patients in the cohort (Res n=14, sdRA n=18) or due to the mixed nature of the resolving group containing patients with many different diagnoses, rather than a single disease as is the case in the sdRA group, introducing more noise to the comparison.

Interestingly, in both resolving and sdRA exploration compared to normal, there were no processes significantly associated with genes downregulated in diseased tissue, despite there being a large number of downregulated genes in the sdRA comparison. There were only 23 genes downregulated in resolving tissue compared to normal, so it is unsurprising that there are no associated processes, but the comparison to sdRA resulted in 131 downregulated genes associated with a more normal signature. This may be due to bias in research, with disease-associated processes being better studied and documented than processes associated with normal tissue homeostasis.

Exploration of the genes involved in the sdRA signature resulted in a number of genes that are known to have a key role in RA pathogenesis, thus increasing confidence in the results. Among these are cytokines that are already being targeted by RA therapeutics, such as IL6, which was one of the genes most significantly associated with sdRA compared to normal, and matrix metalloproteinases, including MMP13 that has a well-documented role in RA

pathogenesis (Moore et al., 2000, Konttinen et al., 1999, Singh et al., 2013, Burrage et al., 2006, Wernicke et al., 1996).

Upon exploration of the top 10 most significant genes and pathways associated with sdRA or resolving disease when compared to normal, there were a number of pro- and anti-inflammatory genes in both signatures, although a relatively more prominent T cell signature was observed in sdRA. There were a number of cytokines and chemokines associated with the resolving signature, such as IL6 and CXCL9, that are chemotactic and activating for both innate and adaptive immune cells, although GO processes involving lymphocyte activation were less evident, potentially suggesting that recruitment of lymphocytes into the synovium occurs in both resolving disease and sdRA but that there is a stronger activation of these cells occurring in sdRA than resolving tissue.

Exploration of the pathways associated with downregulated DEGs (unadjusted $p<0.05$, LFC>1) in resolving disease compared to sdRA again revealed a potentially adaptive immune response signature, with regulation of lymphocyte activation and humoral immune response both being identified as significant processes associated with sdRA. When taken alongside the relative absence of adaptive immune activation seen in the resolving signature when compared to normal, this suggests a less active adaptive immune response in resolving disease. This is despite the presence of similar levels of immune cell infiltration when measured by histology, with there being no difference in the prevalence of pauci-immune, diffuse, or lymphoid pathotypes in resolving compared to sdRA (Chapter 3.4).

The presence of an innate immune response signal in resolving disease could be due to the upregulation of resolution pathways, which involves the recruitment of anti-inflammatory macrophages that phagocytose apoptotic immune cells, in particular neutrophils, and mediate tissue repair (Panigrahy et al., 2021). However, the presence of neutrophil activation processes in the resolving disease signature would suggest that this is not a complete explanation of the differences in immune activation, as neutrophils are typically pro-inflammatory and undergo apoptosis during resolution of inflammation (Panigrahy et al., 2021). Furthermore, a number of broadly pro-inflammatory genes, including IL6, were found to be increased in resolving disease compared to normal tissue. Although the resolving patients are predominantly in an improving phase, the biopsy joint was still inflamed at the time of biopsy, so it is expected that there would still be some pro-inflammatory signals.

Neutrophil chemotaxis and activation pathways were present in both the resolving and sdRA signatures, despite neutrophils being largely absent from synovial tissue, with them only being found in 15% of RA synovial tissue sections in a study by Orange et al. (2020). However, neutrophils are greatly increased in RA synovial fluid, and are likely recruited via the synovium to the synovial fluid where they produce large amounts of pro-inflammatory cytokines and chemokines as well as reactive oxygen species and proteases responsible for joint destruction, thus driving pathogenesis (Thieblemont et al., 2016).

Whether or not the adaptive immune response is ever fully activated in resolving inflammatory arthritis is another question that arises from this analysis. It may be that adaptive immunity was activated at the initiation of disease but that at the time of biopsy activation processes had already been downregulated.

Alternatively, it may be that in resolving disease a full adaptive immune response is not mounted and instead inflammation is predominantly driven by innate immunity. Although the presence of lymphocytes within the resolving tissue may support the first hypothesis, it is not possible to answer this definitively from this dataset as exploration of the tissue at multiple time points over the course of resolution would be required.

A lipid metabolism signature was also found upon exploration of GO biological processes associated with differentially expressed genes (unadjusted $p<0.05$, LFC>1), with fatty acid derivative metabolic process, unsaturated fatty acid metabolic process, and fatty-acyl-CoA biosynthetic process all being associated with genes upregulated in resolving patients. Lipid metabolism has been demonstrated to have association with inflammation, with pro-inflammatory immune cells, such as M1 macrophages and effector T cells, having increased fatty acid synthesis and anti-inflammatory cells, such as M2 macrophages and Treg cells, having increased fatty acid oxidation (Boucher et al., 2021, O'Neill et al., 2016, Weyand et al., 2020). Furthermore, T cells in RA have been found to have increased expression of genes involved in FA synthesis, including FASN, and SCD, compared to normal controls, both of which were found to be upregulated in Res compared to sdRA in the current study (Wu et al., 2020). This finding of increased lipid synthesis in resolving inflammatory arthritis, where one might expect to find more anti-inflammatory cellular subtypes, is therefore broadly contrary to existing literature on lipid metabolism in inflammatory disease, although with the caveat that the existing literature describes expression in immune cell types, which may not be responsible for the signature in resolving disease.

Another potential source of lipid metabolism differences could be due to the presence of lipid droplets. The role of lipid droplets in inflammation is an area of increasing interest and the novel finding that PNPLA3, which is involved in lipid droplet formation, is increased in resolving patients compared to sdRA suggests an interesting potential link between lipid droplets and resolution of inflammatory arthritis (Chamoun et al., 2013). This is again unexpected, as lipid droplets increase in number during inflammation in many immune cell types, including macrophages and activated lymphocytes (Bozza et al., 2009). T cells in RA have been found to have increased presence of lipid droplets, which is thought to enable the rapid formation of cell membrane required during cell proliferation and allow for the hypermobility seen in RA T cells (Weyand et al., 2020, Wu et al., 2020).

However, as healthy synovial tissue contains adipocytes, an increase in lipid synthesis and lipid droplets in resolving disease may be independent of expression in immune cells and instead could be a sign of increasing adipocyte activity, potentially indicating a reversion towards normal tissue. To date there has not been systematic study into the presence or function of adipocytes in the synovium so it is currently unknown whether restoration of adipocytes is required for resolution of disease.

Alternatively, lipid synthesis may be increased in other cell types that have not been well studied in resolving disease, as existing literature predominantly focuses on the study of leukocytes in acute and persistent inflammation. For example, lipid droplets have been shown to be protective against oxidative stress and lipotoxicity, as well as other causes of cellular stress, by sequestering and exporting or neutralising toxic lipids, proteins, and reactive oxygen species

(Pressly et al., 2022). It could be that in other cell types that have been less well studied in this regard, for example fibroblasts, an increase in lipid droplets in resolving patients makes them more resilient to these stressors, enabling clearance and resolution of inflammation.

Further exploration of the cell types responsible for the increase in lipid synthesis would help to clarify the role of lipid metabolism in resolving inflammatory arthritis. In order to address this question, a number of approaches are currently ongoing, including spatial transcriptomics using Visium, which allows for resolution of gene expression at the level of small groups (10-50) of cells, and nanostring GeoMx, which allows for the exploration of gene expression in immune pathways within regions of interest of around 100 cells (Stahl et al., 2016, Merritt et al., 2020). Another approach that is being explored is single-cell metabolic regulome profiling (scMEP), which allows for the spatial quantification rate-limiting and important metabolic regulators, including transcription factors, enzymes, signalling molecules, and transporters, to give a spatial overview of the activity of metabolic processes (Hartmann et al., 2021).

Overall, the differential expression and GO biological process analyses suggest differential immune responses and metabolic processes between sdRA and resolving disease, with resolvers having increased lipid metabolism and a predominantly innate immune response signature and sdRA having a more active adaptive immune response. Further exploration and validation of the differences in immune response and metabolism between early RA and resolving inflammatory arthritis may allow for the identification of targets that could drive early RA away from persistence and towards resolution, potentially aiming for greater levels of drug free remission or even cure in RA patients.

7 DISCUSSION

The clinical heterogeneity present in RA and other inflammatory arthropathies has been widely documented yet still provides a significant challenge for the diagnosis and management of these conditions. Many patients are initially classified as having UA before later developing a persistent inflammatory arthritis and a significant proportion of patients do not respond to first line treatments. This study has furthered our understanding of this heterogeneity, exploring the similarities and differences between different patient groups, across diagnosis, disease severity, response to treatment, and persistence versus resolution of inflammatory disease on both a histological and transcriptomic level.

Firstly, tissue morphology was explored using simple, readily available H&E staining, to address the hypothesis that histological pathotypes would stratify early inflammatory arthritis patients and correlate with clinical variables. Overall, this hypothesis was accepted, as pathotype correlated with local and systemic measures of inflammation, including CRP and US GS, even though it was not able to separate patients by clinical group.

Furthermore, pathotype appears to be associated with response to treatment in RA. This was suggested by a trend based on EULAR DAS28-ESR response, although there were insufficient numbers to test this statistically, and by later exploration of the average expression of DEGs associated with response and non-response in RA. This exploration was done based on standard of care approaches to treatment as opposed to algorithms imposed in a clinical trial, but it would be interesting to explore whether different therapies work better on

specific pathotypes, as has been previously suggested based on other pathotype definitions using more complex staining (Nerviani et al., 2020, Dennis et al., 2014). To test the clinical utility of the BEACON pathotype, validation with greater numbers, exploration in external cohorts, and calculation of the specificity and sensitivity of this approach would be required, followed by a clinical trial to test association between pathotype and treatment response for different therapies prior to clinical use. However, if proven effective at predicting treatment response, particularly for specific therapies, this could improve on the current trial and error approach to DMARD selection and instead aid more informed treatment decisions, using simple staining that is already used in clinical settings. This could improve response rates, particularly to first-line therapy, and drive down the unnecessary costs, side effects, and increased disease burden associated with ineffective treatment. Investigation into association with treatment response in the other patient groups would also be worthwhile, to test utility in other inflammatory arthritis.

The utility of the BEACON scoring was explored in the AMP RA Phase II cohort, which showed some overlapping clinical associations with the BEACON cohort, including associations with CRP, a marker of systemic inflammation. However, a number of the associations were not replicated in the AMP cohort. This is likely due to the different compositions of these cohorts, with the BEACON cohort consisting of patients with mixed diagnoses and disease activities and the AMP cohort including patients with RA at different disease stages but all with high disease activity, some of whom had undergone DMARD treatment. It would therefore be beneficial to test the BEACON scoring system on another treatment-

naïve early inflammatory arthritis cohort, as the scoring system was developed for use in these cohorts.

Overall, The BEACON scoring system is an improvement on previous scoring systems due to its simplicity in both the scoring and in the staining required for pathotype derivation but also because it was developed for use in all early inflammatory arthritis, rather than RA alone. However, it still comes with the challenges associated with manual scoring systems, namely that it is time intensive, has the potential to be influenced by human error or bias, and only produces semi-quantitative results. AI-assisted scoring is currently in development to attempt to overcome these issues, allowing for automatic and quantitative investigation of the synovium.

The heterogeneity seen in early inflammatory arthritis provided a challenge when exploring diagnostic biomarkers for early RA, with no significantly DEGs being identified when comparing sdRA to resolving and nonRA transcriptomes. On top of general heterogeneity seen in early inflammatory arthritis, this is a noisy comparison due to the presence of multiple different diagnoses within both the resolving and nonRA groups. Filtering of these groups further to include single diagnoses in each comparator group could reduce the noise, however would also significantly reduce numbers within the comparisons and would reduce utility of the results as biomarkers of sdRA in mixed early inflammatory arthritis cohorts.

Differential expression analysis was also unable to yield any meaningful results when looking for mechanisms that differ between short duration and longer duration RA, despite the exploration of less noisy groups. This again may be due to the large amount of heterogeneity seen within RA or it could be that there are

no detectable differences at the gene expression level, with any differences being at the protein level. Correcting for pathotype in the differential expression analysis could be one solution to overcome this, to remove variation broadly based on tissue composition.

The hypotheses that differentially expressed genes or signatures in short duration RA patients will distinguish them from other patient groups and that there are genes or pathways that differ between short duration and long duration RA that can be identified using whole tissue RNA sequencing were therefore rejected in this study. This highlights the need either for larger patient cohorts, which may enable the detection of more subtle differences masked by the heterogeneity, or for more specific exploration, as bulk RNA sequencing may miss changes occurring in specific cell types or tissue areas due to the mixed nature of the data. Integrating single cell and bulk RNA sequencing data could be useful in this regard, as single cell allows for the detection of changes within cellular subtypes, while bulk sequencing is more sensitive for the detection of low-level gene expression.

Other analysis methods that take into account gene correlations or networks may enable the reduction of some of the noise, which could achieve more meaningful results. Success with a machine learning technique that accounted for gene co-expression was demonstrated in a previous study by Yeo et al. (2016), which identified the cytokines CXCL4 and CXCL7 as being associated with early RA when compared to established RA and resolving disease. WGCNA is a similar technique that is being explored, which enables the identification of gene networks that can be explored across patient groups.

Despite working with relatively low sample numbers, the exploration of DEGs involved in EULAR DAS28-ESR response compared to non-response at 12 months did yield significant genes and allowed for the identification of an immune-related signature associated with responders and a fibroblast signature in non-responders, resulting in the acceptance of the hypothesis that biomarkers for good or poor outcomes in RA can be identified in gene expression data.

This was broadly in agreement with a previous study by Lewis et al. (2019), although there was only a single response DEG that was also found to be associated with response in their dataset. This may be due to the difference in timeframe, with the Lewis et al. study looking at response at 6 months and the current study looking at response at 12 months. This means that patients will have undergone different treatments, with most patients likely to have only received methotrexate at 6 months, while some in the BEACON cohort will have moved onto biologics by 12 months. Furthermore, the Lewis et al. study excluded patients who had received biologics, confirming this difference in treatment approaches between these studies. The non-overlapping differentially expressed genes may, therefore, be due to the differences in treatment mechanisms, with methotrexate targeting several processes, most notably cellular proliferation and apoptosis, and biologic therapies having more targeted mechanisms of action, often inhibiting a specific cytokine or pathway (Suzuki et al., 2018). Exploration of which treatments were received in both cohorts may help to deconvolute these differences and potentially associate differentially expressed genes with specific therapeutics, which may allow for patient stratification based on response to treatment, however there were not enough patients in this study to explore these questions with sufficient statistical power.

Nevertheless, the ability to identify a signature between responders and non-responders in only 18 RA samples that is broadly in agreement with signatures found in a similar previous study suggests that it is a strong signal that may be easily identified in patient cohorts and therefore would be worth further exploration and validation.

The final hypothesis explored was that mechanisms that drive resolution versus persistence can be identified using whole tissue RNA sequencing. Differential expression followed by GO biological process analyses was able to find differences in immune activation and metabolism between resolving and sdRA patients, thus confirming the alternative hypothesis. Differences in immune activation were perhaps anticipated, as one might expect pro-inflammatory pathways to be higher in sdRA and pro-resolution processes to be found in resolving patients. However, the differences in lipid metabolism were not expected. Lipid synthesis is generally associated with more pro-inflammatory processes, being increased in M1 macrophages and activated effector T cells, so the increase of lipid synthesis in resolving patients was surprising. In normal synovial tissue there is a large proportion of adipocytes so it may be that during resolution of inflammatory arthritis, adipocytes again increase in number, which could explain the increase in lipid synthesis. Alternatively, expression in fibroblasts or other cells that do not have as well characterised metabolism could be driving these differences. It is not possible to confirm this from the currently available data but this is an area of ongoing research.

A potential role for lipid droplets is supported by the presence of PNPLA3 as a DEG increased in resolving disease, as this has been found to have a role in lipid droplet formation (Chamoun et al., 2013). However, lipid droplets have also been

found to be associated with inflammation, being increased in activated immune cells, such as T cells and macrophages, in response to pro-inflammatory signals, including LPS and some cytokines (Bozza et al., 2009).

Pauci-immune samples had higher PNPLA3 expression compared to the other pathotypes and, from exploration of the PCA gene contributions, also appeared to be associated with genes predominantly expressed in adipocytes, including ADIPOQ, PLIN1, and PLIN4. Interestingly, PNPLA3, as well as PLIN1, PLIN4, PLIN5, and ADIPOQ were also upregulated in UMAP C2, which predominantly consisted of pauci-immune samples, compared to the other two clusters. These data may suggest that pauci-immune samples have more adipocytes present than the other pathotypes. This does not explain why there is a difference in lipid metabolism between resolving disease and sdRA, as there was no difference in the pathotypes between these groups and other differentially expressed genes involved in lipid metabolism, such as ELOVL6, were not associated with any particular pathotype. This may suggest that the presence of adipocytes is not solely responsible for this signature, although direct quantification of adipocytes would be required to confirm this. It may be that there is a combination of cell types contributing to this signature, with different signals and cellular interactions driving resolution processes and altered metabolism across the synovium.

Further investigation into this area is required to elucidate what is driving the metabolic differences between resolving disease and sdRA, whether lipid droplets are increased in resolving disease, which cell types are associated with this signature, and what the mechanistic role is in resolution.

7.1 Future Work

To continue and build upon the work completed in this study, there are a number of techniques that could be utilised and further areas of study to explore. Firstly, work needs to be competed on the Visiopharm APPs. This will involve completion of validation of the lymphocyte and plasma cell APPs before the whole pipeline can be run on the BEACON cohort. Results from this pipeline can then be compared to the manual BEACON scoring and clinical variables to explore the utility of these APPs. Validation on the AMP external cohort will then be sought to ensure reproducibility on samples outside of the BEACON cohort.

As adipocyte-related genes drive a large amount of variation in the RNA sequencing dataset, expanding the BEACON histology scoring to include an adipocyte score or building an AI-assisted APP to detect adipocytes may enable for patient stratification based on this, which could have links to different clinical variables than the current grading. Furthermore, this would allow for the comparison of levels of adipocytes in resolving tissue compared to early RA, which could help to explain the difference in lipid metabolism between the two groups. Testing for the presence of lipid droplets using staining, such as nile red that stains intracellular lipid droplets, alongside cellular markers could also enable further exploration of the source of the lipid metabolism differences and enable investigation into the cell types that are driving this.

Integrating the bulk RNA sequencing dataset with single cell RNA sequencing data via the exploration of key gene or gene network expression across cell types or for use as a map for CIBERSORTx, which allows for computational deconvolution of cell types, may allow for identification of the cell types

expressing these signatures (Newman et al., 2019). This will also give clues as to which cell types to prioritise exploration of in future validation studies.

Alternatively, spatial transcriptomic techniques, such as 10x Visium or Nanostring GeoMx®, may give insights into which tissue areas or cell types are driving the expression of gene signatures identified in this study (Stahl et al., 2016, Merritt et al., 2020). This could be particularly useful for exploring differences in immune processes, such as the difference in adaptive immune activation between resolvers and sdRA, by allowing for the exploration of gene expression specifically within lymphocyte aggregates, which are present in both patient groups despite the transcriptional differences identified.

However, transcriptomic signatures are not always translated into differences in protein levels due to variations in protein stability via modification and degradation, as well as translational regulation, so validation at the protein level is important following transcriptomic-based studies. This could be achieved by using immunohistochemistry or immunofluorescence staining for proteins of interest. However, to investigate metabolic processes spatially and at the level of proteins, use of a technique developed by Dr Felix Hartmann called single-cell metabolic regulome profiling (scMEP) is being explored (Hartmann et al., 2021). This allows for the spatial profiling of metabolic profiles via the use of multiplexed ion beam imaging by time of flight (MIBI-TOF) (Keren et al., 2019) for detection of key enzymes, transcription factors, transporters, and other important factors involved in metabolic processes, such as glycolysis, fatty acid metabolism, and amino acid metabolism. These are detected alongside markers for identification of cellular subtypes, to enable identification of the cell types utilising different metabolic processes. This will allow for validation of the

metabolic signature found in resolving disease in another dataset and at the protein level.

8 REFERENCES

Human Protein Atlas available from http://www.proteinatlas.org.

Skeletal muscle gene list available from https://www.proteinatlas.org/humanproteome/tissue/skeletal+muscle.

Abbasi, M., Mousavi, M. J., Jamalzehi, S., Alimohammadi, R., Bezvan, M. H., Mohammadi, H. & Aslani, S. 2019. Strategies toward rheumatoid arthritis therapy; the old and the new. *J Cell Physiol,* 234**,** 10018-10031.

Adamczyk, A., Gageik, D., Frede, A., Pastille, E., Hansen, W., Rueffer, A., Buer, J., Buning, J., Langhorst, J. & Westendorf, A. M. 2017. Differential expression of GPR15 on T cells during ulcerative colitis. *JCI Insight,* 2.

Aho, T. L., Lund, R. J., Ylikoski, E. K., Matikainen, S., Lahesmaa, R. & Koskinen, P. J. 2005. Expression of human pim family genes is selectively up-regulated by cytokines promoting T helper type 1, but not T helper type 2, cell differentiation. *Immunology,* 116**,** 82-8.

Al-Alwan, M., Du, Q., Hou, S., Nashed, B., Fan, Y., Yang, X. & Marshall, A. J. 2007. Follicular dendritic cell secreted protein (FDC-SP) regulates germinal center and antibody responses. *J Immunol,* 178**,** 7859-67.

Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., Bingham, C. O., 3rd, Birnbaum, N. S., Burmester, G. R., Bykerk, V. P., Cohen, M. D., Combe, B., Costenbader, K. H., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J. M., Hobbs, K., Huizinga, T. W., Kavanaugh, A., Kay, J., Kvien, T. K., Laing, T., Mease, P., Menard, H. A., Moreland, L. W., Naden, R. L., Pincus, T., Smolen, J. S., Stanislawska-Biernat, E., Symmons, D., Tak, P. P., Upchurch, K. S., Vencovsky, J., Wolfe, F. & Hawker, G. 2010. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum,* 62**,** 2569-81.

Alivernini, S., Macdonald, L., Elmesmari, A., Finlay, S., Tolusso, B., Gigante, M. R., Petricca, L., Di Mario, C., Bui, L., Perniola, S., Attar, M., Gessi, M., Fedele, A. L., Chilaka, S., Somma, D., Sansom, S. N., Filer, A., Mcsharry, C., Millar, N. L., Kirschner, K., Nerviani, A., Lewis, M. J., Pitzalis, C., Clark, A. R., Ferraccioli, G., Udalova, I., Buckley, C. D., Gremese, E., Mcinnes, I. B., Otto, T. D. & Kurowska-Stolarska, M. 2020. Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat Med,* 26**,** 1295-1306.

Alpizar-Rodriguez, D., Lesker, T. R., Gronow, A., Gilbert, B., Raemy, E., Lamacchia, C., Gabay, C., Finckh, A. & Strowig, T. 2019. Prevotella copri in individuals at risk for rheumatoid arthritis. *Ann Rheum Dis,* 78**,** 590-593.

Anderson, A. E., Maney, N. J., Nair, N., Lendrem, D. W., Skelton, A. J., Diboll, J., Brown, P. M., Smith, G. R., Carmody, R. J., Barton, A., Isaacs, J. D. & Pratt, A. G. 2019. Expression of STAT3-regulated genes in circulating CD4+ T cells discriminates rheumatoid arthritis independently of clinical parameters in early arthritis. *Rheumatology (Oxford),* 58**,** 1250-1258.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.*

Arias De La Rosa, I., Escudero-Contreras, A., Rodriguez-Cuenca, S., Ruiz-Ponce, M., Jimenez-Gomez, Y., Ruiz-Limon, P., Perez-Sanchez, C., Abalos-Aguilera, M. C., Cecchi, I., Ortega, R., Calvo, J., Guzman-Ruiz, R., Malagon, M. M., Collantes-Estevez, E., Vidal-Puig, A., Lopez-Pedrera, C. & Barbarroja, N. 2018. Defective glucose and lipid metabolism in rheumatoid arthritis is determined by chronic inflammation in metabolic tissues. *J Intern Med,* 284**,** 61-77.

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, M. H., Luthra, H. S. & Et Al. 1988a. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum,* 31**,** 315-24.

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, M. H., Luthra, H. S., Medsger, T. A., Mitchell, D. M., Neustadt, D. H., Pinals, R. S., Schaller, J. G., Sharp, J. T., Wilder, R. L. & Hunder, G. G. 1988b. The American-Rheumatism-Association 1987 Revised Criteria for the Classification of Rheumatoid-Arthritis. *Arthritis and Rheumatism,* 31**,** 315-324.

Badot, V., Galant, C., Nzeusseu Toukap, A., Theate, I., Maudoux, A. L., Van Den Eynde, B. J., Durez, P., Houssiau, F. A. & Lauwerys, B. R. 2009. Gene expression profiling in the synovium identifies a predictive signature of absence of response to adalimumab therapy in rheumatoid arthritis. *Arthritis Res Ther,* 11**,** R57.

Baeten, D., Demetter, P., Cuvelier, C., Van Den Bosch, F., Kruithof, E., Van Damme, N., Verbruggen, G., Mielants, H., Veys, E. M. & De Keyser, F. 2000. Comparative study of the synovial histology in rheumatoid arthritis, spondyloarthropathy, and osteoarthritis: influence of disease duration and activity. *Ann Rheum Dis,* 59**,** 945-53.

Baranzini, S. E., Wang, J., Gibson, R. A., Galwey, N., Naegelin, Y., Barkhof, F., Radue, E. W., Lindberg, R. L., Uitdehaag, B. M., Johnson, M. R., Angelakopoulou, A., Hall, L., Richardson, J. C., Prinjha, R. K., Gass, A., Geurts, J. J., Kragt, J., Sombekke, M., Vrenken, H., Qualley, P., Lincoln, R. R., Gomez, R., Caillier, S. J., George, M. F., Mousavi, H., Guerrero, R., Okuda, D. T., Cree, B. A., Green, A. J., Waubant, E., Goodin, D. S., Pelletier, D., Matthews, P. M., Hauser, S. L., Kappos, L., Polman, C. H. & Oksenberg, J. R. 2009. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet,* 18**,** 767-78.

Bartok, B. & Firestein, G. S. 2010. Fibroblast-like synoviocytes: key effector cells in rheumatoid arthritis. *Immunol Rev,* 233**,** 233-55.

Bartsch, J. W. & Schlomann, U. 2013. Chapter 249 - ADAM8/MS2/CD156a. *In:* RAWLINGS, N. D. & SALVESEN, G. (eds.) *Handbook of Proteolytic Enzymes (Third Edition).* Academic Press.

Benito-Garcia, E., Feskanich, D., Hu, F. B., Mandl, L. A. & Karlson, E. W. 2007. Protein, iron, and meat consumption and risk for rheumatoid arthritis: a prospective cohort study. *Arthritis Res Ther,* 9**,** R16.

Bergstra, S. A. & Allaart, C. F. 2018. What is the optimal target for treat-to-target strategies in rheumatoid arthritis? *Curr Opin Rheumatol,* 30**,** 282-287.

Bluett, J., Sergeant, J. C., Macgregor, A. J., Chipping, J. R., Marshall, T., Symmons, D. P. M. & Verstappen, S. M. M. 2018. Risk factors for oral methotrexate failure in patients with inflammatory polyarthritis: results from a UK prospective cohort study. *Arthritis Res Ther,* 20**,** 50.

Bonet-Fernandez, J. M., Aroca-Aguilar, J. D., Corton, M., Ramirez, A. I., Alexandre-Moreno, S., Garcia-Anton, M. T., Salazar, J. J., Ferre-Fernandez, J. J., Atienzar-Aroca, R., Villaverde, C., Iancu, I., Tamayo, A., Mendez-Hernandez, C. D., Morales-Fernandez, L., Rojas, B., Ayuso, C., Coca-Prados, M., Martinez-De-La-Casa, J. M., Garcia-Feijoo, J. & Escribano, J. 2020. CPAMD8 loss-of-function underlies non-dominant congenital glaucoma with variable anterior segment dysgenesis and abnormal extracellular matrix. *Hum Genet,* 139**,** 1209-1231.

Bosello, S., Fedele, A. L., Peluso, G., Gremese, E., Tolusso, B. & Ferraccioli, G. 2011. Very early rheumatoid arthritis is the major predictor of major outcomes: clinical ACR remission and radiographic non-progression. *Annals of the Rheumatic Diseases,* 70**,** 1292-1295.

Boucher, D. M., Vijithakumar, V. & Ouimet, M. 2021. Lipid Droplets as Regulators of Metabolism and Immunity. *Immunometabolism,* 3**,** e210021.

Boulland, M. L., Marquet, J., Molinier-Frenkel, V., Moller, P., Guiter, C., Lasoudris, F., Copie-Bergman, C., Baia, M., Gaulard, P., Leroy, K. & Castellano, F. 2007. Human IL4I1 is a secreted L-phenylalanine oxidase expressed by mature dendritic cells that inhibits T-lymphocyte proliferation. *Blood,* 110**,** 220-7.

Boyle, D. L., Rosengren, S., Bugbee, W., Kavanaugh, A. & Firestein, G. S. 2003. Quantitative biomarker analysis of synovial gene expression by real-time PCR. *Arthritis Res Ther,* 5**,** R352-60.

Bozza, P. T., Magalhaes, K. G. & Weller, P. F. 2009. Leukocyte lipid bodies - Biogenesis and functions in inflammation. *Biochim Biophys Acta,* 1791**,** 540-51.

Bradley, J. R. 2008. TNF-mediated inflammatory disease. *J Pathol,* 214**,** 149-60.

Brzustewicz, E. & Bryl, E. 2015. The role of cytokines in the pathogenesis of rheumatoid arthritis--Practical and potential application of cytokines as biomarkers and targets of personalized therapy. *Cytokine,* 76**,** 527-536.

Buch, M. H. 2018. Defining refractory rheumatoid arthritis. *Ann Rheum Dis,* 77**,** 966-969.

Bucht, A., Larsson, P., Weisbrot, L., Thorne, C., Pisa, P., Smedegard, G., Keystone, E. C. & Gronberg, A. 1996. Expression of interferon-gamma (IFN-gamma), IL-10, IL-12 and transforming growth factor-beta (TGF-beta) mRNA in synovial fluid cells from patients in the early and late phases of rheumatoid arthritis (RA). *Clin Exp Immunol,* 103**,** 357-67.

Buckley, C. D., Pilling, D., Lord, J. M., Akbar, A. N., Scheel-Toellner, D. & Salmon, M. 2001. Fibroblasts regulate the switch from acute resolving to chronic persistent inflammation. *Trends Immunol,* 22**,** 199-204.

Buckwalter, J. A. & Lappin, D. R. 2000. The disproportionate impact of chronic arthralgia and arthritis among women. *Clin Orthop Relat Res***,** 159-68.

Burgueno, J. F., Fritsch, J., Gonzalez, E. E., Landau, K. S., Santander, A. M., Fernandez, I., Hazime, H., Davies, J. M., Santaolalla, R., Phillips, M. C., Diaz, S., Dheer, R., Brito, N., Pignac-Kobinger, J., Fernandez, E., Conner, G. E. & Abreu, M. T. 2021. Epithelial TLR4 Signaling Activates DUOX2 to

Induce Microbiota-Driven Tumorigenesis. *Gastroenterology,* 160**,** 797-808 e6.
Burrage, P. S., Mix, K. S. & Brinckerhoff, C. E. 2006. Matrix metalloproteinases: role in arthritis. *Front Biosci,* 11**,** 529-43.
Cader, M. Z., Filer, A., Hazlehurst, J., De Pablo, P., Buckley, C. D. & Raza, K. 2011. Performance of the 2010 ACR/EULAR criteria for rheumatoid arthritis: comparison with 1987 ACR criteria in a very early synovitis cohort. *Annals of the Rheumatic Diseases,* 70**,** 949-955.
Canete, J. D., Celis, R., Moll, C., Izquierdo, E., Marsal, S., Sanmarti, R., Palacin, A., Lora, D., De La Cruz, J. & Pablos, J. L. 2009. Clinical significance of synovial lymphoid neogenesis and its reversal after anti-tumour necrosis factor alpha therapy in rheumatoid arthritis. *Ann Rheum Dis,* 68**,** 751-6.
Cao, Y.-N., Yue, S.-S., Wang, A.-Y., Xu, L., Hu, Y.-T., Qiao, X., Wu, T.-Y., Ye, M., Wu, Y.-C. & Qi, R. 2022. Antrodia cinnamomea and its compound dehydroeburicoic acid attenuate nonalcoholic fatty liver disease by upregulating ALDH2 activity. *Journal of Ethnopharmacology,* 292**,** 115146.
Cartwright, A., Schmutz, C., Askari, A., Kuiper, J. H. & Middleton, J. 2014. Orphan receptor GPR15/BOB is up-regulated in rheumatoid arthritis. *Cytokine,* 67**,** 53-9.
Casnici, C., Lattuada, D., Tonna, N., Crotta, K., Storini, C., Bianco, F., Truzzi, M. C., Corradini, C. & Marelli, O. 2014. Optimized "in vitro" culture conditions for human rheumatoid arthritis synovial fibroblasts. *Mediators Inflamm,* 2014**,** 702057.
Cathcart, M. K. & Bhattacharjee, A. 2014. Monoamine oxidase A (MAO-A): a signature marker of alternatively activated monocytes/macrophages. *Inflamm Cell Signal,* 1.
Chadwick, N., Zeef, L., Portillo, V., Boros, J., Hoyle, S., Van Doesburg, J. C. & Buckle, A. M. 2010. Notch protection against apoptosis in T-ALL cells mediated by GIMAP5. *Blood Cells Mol Dis,* 45**,** 201-9.
Chamoun, Z., Vacca, F., Parton, R. G. & Gruenberg, J. 2013. PNPLA3/adiponutrin functions in lipid droplet formation. *Biol Cell,* 105**,** 219-233.
Chen, J., Wright, K., Davis, J. M., Jeraldo, P., Marietta, E. V., Murray, J., Nelson, H., Matteson, E. L. & Taneja, V. 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med,* 8**,** 43.
Chen, K., Nishi, H., Travers, R., Tsuboi, N., Martinod, K., Wagner, D. D., Stan, R., Croce, K. & Mayadas, T. N. 2012. Endocytosis of soluble immune complexes leads to their clearance by FcgammaRIIIB but induces neutrophil extracellular traps via FcgammaRIIA in vivo. *Blood,* 120**,** 4421-31.
Chen, L. F., Mo, Y. Q., Ma, J. D., Luo, L., Zheng, D. H. & Dai, L. 2014. Elevated serum IgG4 defines specific clinical phenotype of rheumatoid arthritis. *Mediators Inflamm,* 2014**,** 635293.
Chen, P., Gao, Y., Ouyang, S., Wei, L., Zhou, M., You, H. & Wang, Y. 2020a. A Prognostic Model Based on Immune-Related Long Non-Coding RNAs for Patients With Cervical Cancer. *Front Pharmacol,* 11**,** 585255.
Chen, W., Zhu, G., Hao, L., Wu, M., Ci, H. & Li, Y.-P. 2013. C/EBPα regulates osteoclast lineage commitment. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 7294-7299.

Chen, X. L., Serrano, D., Mayhue, M., Hoebe, K., Ilangumaran, S. & Ramanathan, S. 2015. GIMAP5 Deficiency Is Associated with Increased AKT Activity in T Lymphocytes. *PLoS One,* 10**,** e0139019.

Chen, Y., Li, H., Lai, L., Feng, Q. & Shen, J. 2020b. Identification of Common Differentially Expressed Genes and Potential Therapeutic Targets in Ulcerative Colitis and Rheumatoid Arthritis. *Front Genet,* 11**,** 572194.

Choi, I. Y., Karpus, O. N., Turner, J. D., Hardie, D., Marshall, J. L., De Hair, M. J. H., Maijer, K. I., Tak, P. P., Raza, K., Hamann, J., Buckley, C. D., Gerlag, D. M. & Filer, A. 2017. Stromal cell markers are differentially expressed in the synovial tissue of patients with early arthritis. *Plos One,* 12.

Coates, L. C. & Helliwell, P. S. 2017. Psoriatic arthritis: state of the art review. *Clin Med (Lond),* 17**,** 65-70.

Coates, L. C., Moverley, A. R., Mcparland, L., Brown, S., Navarro-Coy, N., O'dwyer, J. L., Meads, D. M., Emery, P., Conaghan, P. G. & Helliwell, P. S. 2015. Effect of tight control of inflammation in early psoriatic arthritis (TICOPA): a UK multicentre, open-label, randomised controlled trial. *Lancet,* 386**,** 2489-98.

Cook, M. E., Jarjour, N. N., Lin, C. C. & Edelson, B. T. 2020. Transcription Factor Bhlhe40 in Immunity and Autoimmunity. *Trends Immunol,* 41**,** 1023-1036.

Cousin, C., Aubatin, A., Le Gouvello, S., Apetoh, L., Castellano, F. & Molinier-Frenkel, V. 2015. The immunosuppressive enzyme IL4I1 promotes FoxP3(+) regulatory T lymphocyte differentiation. *Eur J Immunol,* 45**,** 1772-82.

Croft, A. P., Campos, J., Jansen, K., Turner, J. D., Marshall, J., Attar, M., Savary, L., Wehmeyer, C., Naylor, A. J., Kemble, S., Begum, J., Durholz, K., Perlman, H., Barone, F., Mcgettrick, H. M., Fearon, D. T., Wei, K., Raychaudhuri, S., Korsunsky, I., Brenner, M. B., Coles, M., Sansom, S. N., Filer, A. & Buckley, C. D. 2019. Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature,* 570**,** 246-251.

Dalberg, U., Markholst, H. & Hornum, L. 2007. Both Gimap5 and the diabetogenic BBDP allele of Gimap5 induce apoptosis in T cells. *Int Immunol,* 19**,** 447-53.

Dalbeth, N. & Callan, M. F. 2002. A subset of natural killer cells is greatly expanded within inflamed joints. *Arthritis Rheum,* 46**,** 1763-72.

Damgaard, D., Senolt, L. & Nielsen, C. H. 2016. Increased levels of peptidylarginine deiminase 2 in synovial fluid from anti-CCP-positive rheumatoid arthritis patients: Association with disease activity and inflammatory markers. *Rheumatology (Oxford),* 55**,** 918-27.

De Filippo, K., Dudeck, A., Hasenberg, M., Nye, E., Van Rooijen, N., Hartmann, K., Gunzer, M., Roers, A. & Hogg, N. 2013. Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. *Blood,* 121**,** 4930-7.

De La Rosa Rodriguez, M. A. & Kersten, S. 2020. Regulation of lipid droplet homeostasis by hypoxia inducible lipid droplet associated HILPDA. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids,* 1865**,** 158738.

Deane, K. D., Demoruelle, M. K., Kelmenson, L. B., Kuhn, K. A., Norris, J. M. & Holers, V. M. 2017. Genetic and environmental risk factors for rheumatoid arthritis. *Best Pract Res Clin Rheumatol,* 31**,** 3-18.

Dennis, G., Jr., Holweg, C. T., Kummerfeld, S. K., Choy, D. F., Setiadi, A. F., Hackney, J. A., Haverty, P. M., Gilbert, H., Lin, W. Y., Diehl, L., Fischer, S., Song, A., Musselman, D., Klearman, M., Gabay, C., Kavanaugh, A., Endres, J., Fox, D. A., Martin, F. & Townsend, M. J. 2014. Synovial phenotypes in rheumatoid arthritis correlate with response to biologic therapeutics. *Arthritis Res Ther,* 16**,** R90.

Derksen, V., Huizinga, T. W. J. & Van Der Woude, D. 2017. The role of autoantibodies in the pathophysiology of rheumatoid arthritis. *Semin Immunopathol,* 39**,** 437-446.

Di Stefano, P., Damiano, L., Cabodi, S., Aramu, S., Tordella, L., Praduroux, A., Piva, R., Cavallo, F., Forni, G., Silengo, L., Tarone, G., Turco, E. & Defilippi, P. 2007. p140Cap protein suppresses tumour cell properties, regulating Csk and Src kinase activity. *The EMBO journal,* 26**,** 2843-2855.

Diller, M., Hasseli, R., Hulser, M. L., Aykara, I., Frommer, K., Rehart, S., Muller-Ladner, U. & Neumann, E. 2019. Targeting Activated Synovial Fibroblasts in Rheumatoid Arthritis by Peficitinib. *Front Immunol,* 10**,** 541.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

Dobrovolna, J., Chinenov, Y., Kennedy, M. A., Liu, B. & Rogatsky, I. 2012. Glucocorticoid-dependent phosphorylation of the transcriptional coregulator GRIP1. *Mol Cell Biol,* 32**,** 730-9.

Dolhain, R. J., Ter Haar, N. T., De Kuiper, R., Nieuwenhuis, I. G., Zwinderman, A. H., Breedveld, F. C. & Miltenburg, A. M. 1998. Distribution of T cells and signs of T-cell activation in the rheumatoid joint: implications for semiquantitative comparative histology. *Br J Rheumatol,* 37**,** 324-30.

Dörner, T. & Lipsky, P. E. 2009. CHAPTER 8B - The Role of B Cells in Rheumatoid Arthritis A2 - Hochberg, Marc C. *In:* SILMAN, A. J., SMOLEN, J. S., WEINBLATT, M. E. & WEISMAN, M. H. (eds.) *Rheumatoid Arthritis.* Philadelphia: Mosby.

Dougados, M., Soubrier, M., Antunez, A., Balint, P., Balsa, A., Buch, M. H., Casado, G., Detert, J., El-Zorkany, B., Emery, P., Hajjaj-Hassouni, N., Harigai, M., Luo, S. F., Kurucz, R., Maciel, G., Mola, E. M., Montecucco, C. M., Mcinnes, I., Radner, H., Smolen, J. S., Song, Y. W., Vonkeman, H. E., Winthrop, K. & Kay, J. 2014. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). *Annals of the Rheumatic Diseases,* 73**,** 62-68.

Duan, C. 2016. Hypoxia-inducible factor 3 biology: complexities and emerging themes. *Am J Physiol Cell Physiol,* 310**,** C260-9.

Ebihara, T., Azuma, M., Oshiumi, H., Kasamatsu, J., Iwabuchi, K., Matsumoto, K., Saito, H., Taniguchi, T., Matsumoto, M. & Seya, T. 2010. Identification of a polyI:C-inducible membrane protein that participates in dendritic cell-mediated natural killer cell activation. *J Exp Med,* 207**,** 2675-87.

Ebrahimiadib, N., Berijani, S., Ghahari, M. & Pahlaviani, F. G. 2021. Ankylosing Spondylitis. *J Ophthalmic Vis Res,* 16**,** 462-469.

Elemam, N. M., Hannawi, S. & Maghazachi, A. A. 2020. Role of Chemokines and Chemokine Receptors in Rheumatoid Arthritis. *Immunotargets Ther,* 9**,** 43-56.

Eshed, I., Feist, E., Althoff, C. E., Hamm, B., Konen, E., Burmester, G. R., Backhaus, M. & Hermann, K. G. 2009. Tenosynovitis of the flexor tendons of the hand detected by MRI: an early indicator of rheumatoid arthritis. *Rheumatology (Oxford),* 48**,** 887-91.

Fagerberg, L., Hallstrom, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C. A., Skogs, M., Takanen, J. O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J. M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., Von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F. & Uhlen, M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics,* 13**,** 397-406.

Farber, J. M. 1997. Mig and IP-10: CXC chemokines that target lymphocytes. *J Leukoc Biol,* 61**,** 246-57.

Farzan, M., Choe, H., Martin, K., Marcon, L., Hofmann, W., Karlsson, G., Sun, Y., Barrett, P., Marchand, N., Sullivan, N., Gerard, N., Gerard, C. & Sodroski, J. 1997. Two orphan seven-transmembrane segment receptors which are expressed in CD4-positive cells support simian immunodeficiency virus infection. *J Exp Med,* 186**,** 405-11.

Fautrel, B. 2018. Therapeutic strategy for rheumatoid arthritis patients who have achieved remission. *Joint Bone Spine*.

Fearon, U., Griosios, K., Fraser, A., Reece, R., Emery, P., Jones, P. F. & Veale, D. J. 2003. Angiopoietins, growth factors, and vascular morphology in early arthritis. *J Rheumatol,* 30**,** 260-8.

Filer, A., De Pablo, P., Allen, G., Nightingale, P., Jordan, A., Jobanputra, P., Bowman, S., Buckley, C. D. & Raza, K. 2011. Utility of ultrasound joint counts in the prediction of rheumatoid arthritis in patients with very early synovitis. *Ann Rheum Dis,* 70**,** 500-7.

Firestein, G. S. 2003. Evolving concepts of rheumatoid arthritis. *Nature,* 423**,** 356-361.

Fischer, A., Zundler, S., Atreya, R., Rath, T., Voskens, C., Hirschmann, S., Lopez-Posadas, R., Watson, A., Becker, C., Schuler, G., Neufert, C., Atreya, I. & Neurath, M. F. 2016. Differential effects of alpha4beta7 and GPR15 on homing of effector and regulatory T cells from patients with UC to the inflamed gut in vivo. *Gut,* 65**,** 1642-64.

Flammer, J. R., Dobrovolna, J., Kennedy, M. A., Chinenov, Y., Glass, C. K., Ivashkiv, L. B. & Rogatsky, I. 2010. The type I interferon signaling pathway is a target for glucocorticoid inhibition. *Mol Cell Biol,* 30**,** 4564-74.

Freeman, G. J., Long, A. J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., Fitz, L. J., Malenkovich, N., Okazaki, T., Byrne, M. C., Horton, H. F., Fouser, L., Carter, L., Ling, V., Bowman, M. R., Carreno, B. M., Collins, M., Wood, C. R. & Honjo, T. 2000. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med,* 192**,** 1027-34.

Gabriel, S. E. & Michaud, K. 2009. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther,* 11**,** 229.

Gavrilă, B. I., Ciofu, C. & Stoica, V. 2016. Biomarkers in Rheumatoid Arthritis, what is new? *Journal of Medicine and Life,* 9**,** 144-148.

Goeb, V., Aegerter, P., Parmar, R., Fardellone, P., Vittecoq, O., Conaghan, P. G., Emery, P., Le Loet, X. & Ponchel, F. 2013. Progression to rheumatoid arthritis in early inflammatory arthritis is associated with low IL-7 serum levels. *Annals of the Rheumatic Diseases,* 72**,** 1032-1036.

Goncalves Dos Santos, G., Jimenez-Andrade, J. M., Woller, S. A., Munoz-Islas, E., Ramirez-Rosas, M. B., Ohashi, N., Ferreira Catroli, G., Fujita, Y., Yaksh, T. L. & Corr, M. 2020. The neuropathic phenotype of the K/BxN transgenic mouse with spontaneous arthritis: pain, nerve sprouting and joint remodeling. *Sci Rep,* 10**,** 15596.

Gorman, C. L. & Cope, A. P. 2008. Immune-mediated pathways in chronic inflammatory arthritis. *Best Pract Res Clin Rheumatol,* 22**,** 221-38.

Gossec, L., Smolen, J. S., Ramiro, S., De Wit, M., Cutolo, M., Dougados, M., Emery, P., Landewe, R., Oliver, S., Aletaha, D., Betteridge, N., Braun, J., Burmester, G., Canete, J. D., Damjanov, N., Fitzgerald, O., Haglund, E., Helliwell, P., Kvien, T. K., Lories, R., Luger, T., Maccarone, M., Marzo-Ortega, H., Mcgonagle, D., Mcinnes, I. B., Olivieri, I., Pavelka, K., Schett, G., Sieper, J., Van Den Bosch, F., Veale, D. J., Wollenhaupt, J., Zink, A. & Van Der Heijde, D. 2016. European League Against Rheumatism (EULAR) recommendations for the management of psoriatic arthritis with pharmacological therapies: 2015 update. *Ann Rheum Dis,* 75**,** 499-510.

Grabowski, P. S., Wright, P. K., Van 'T Hof, R. J., Helfrich, M. H., Ohshima, H. & Ralston, S. H. 1997. Immunolocalization of inducible nitric oxide synthase in synovium and cartilage in rheumatoid arthritis and osteoarthritis. *Br J Rheumatol,* 36**,** 651-5.

Gremese, E., Salaffi, F., Bosello, S. L., Ciapetti, A., Bobbio-Pallavicini, F., Caporali, R. & Ferraccioli, G. 2013. Very early rheumatoid arthritis as a predictor of remission: a multicentre real life prospective study. *Annals of the Rheumatic Diseases,* 72**,** 858-862.

Gu, Z., Eils, R. & Schlesner, M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics,* 32**,** 2847-9.

Guo, Q., Wang, Y., Xu, D., Nossent, J., Pavlos, N. J. & Xu, J. 2018. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Res,* 6**,** 15.

Ha, Y. J., Choi, Y. S., Han, D. W., Kang, E. H., Yoo, I. S., Kim, J. H., Kang, S. W., Lee, E. Y., Song, Y. W. & Lee, Y. J. 2019. PIM-1 kinase is a novel regulator of proinflammatory cytokine-mediated responses in rheumatoid arthritis fibroblast-like synoviocytes. *Rheumatology (Oxford),* 58**,** 154-164.

Hara, S., Hamada, J., Kobayashi, C., Kondo, Y. & Imura, N. 2001. Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in human kidney: suppression of HIF-mediated gene expression by HIF-3alpha. *Biochem Biophys Res Commun,* 287**,** 808-13.

Harre, U., Georgess, D., Bang, H., Bozec, A., Axmann, R., Ossipova, E., Jakobsson, P. J., Baum, W., Nimmerjahn, F., Szarka, E., Sarmay, G., Krumbholz, G., Neumann, E., Toes, R., Scherer, H. U., Catrina, A. I., Klareskog, L., Jurdic, P. & Schett, G. 2012. Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin. *J Clin Invest,* 122**,** 1791-802.

Hartmann, F. J., Mrdjen, D., Mccaffrey, E., Glass, D. R., Greenwald, N. F., Bharadwaj, A., Khair, Z., Verberk, S. G. S., Baranski, A., Baskar, R., Graf, W., Van Valen, D., Van Den Bossche, J., Angelo, M. & Bendall, S. C. 2021.

Single-cell metabolic profiling of human cytotoxic T cells. *Nat Biotechnol,* 39**,** 186-197.
Hazes, J. M. & Luime, J. J. 2011. The epidemiology of early inflammatory arthritis. *Nat Rev Rheumatol,* 7**,** 381-90.
He, L., Liu, X., Yang, J., Li, W., Liu, S., Liu, X., Yang, Z., Ren, J., Wang, Y., Shan, L., Guan, C., Pei, F., Lei, L., Zhang, Y., Yi, X., Yang, X., Liang, J., Liu, R., Sun, L. & Shang, Y. 2018. Imbalance of the reciprocally inhibitory loop between the ubiquitin-specific protease USP43 and EGFR/PI3K/AKT drives breast carcinogenesis. *Cell Res,* 28**,** 934-951.
Hebbali, A. 2020. olsrr: Tools for Building OLS Regression Models. *R package version 0.5.3. https://CRAN.R-project.org/package=olsrr.*
Heidari, B. 2011. Rheumatoid Arthritis: Early diagnosis and treatment outcomes. *Caspian J Intern Med,* 2**,** 161-70.
Hepburn, A. L., Mason, J. C. & Davies, K. A. 2004. Expression of Fcgamma and complement receptors on peripheral blood monocytes in systemic lupus erythematosus and rheumatoid arthritis. *Rheumatology (Oxford),* 43**,** 547-54.
Hofsteen, P., Robitaille, A. M., Strash, N., Palpant, N., Moon, R. T., Pabon, L. & Murry, C. E. 2018. ALPK2 Promotes Cardiogenesis in Zebrafish and Human Pluripotent Stem Cells. *iScience,* 2**,** 88-100.
Hogan, V. E., Holweg, C. T., Choy, D. F., Kummerfeld, S. K., Hackney, J. A., Teng, Y. K., Townsend, M. J. & Van Laar, J. M. 2012. Pretreatment synovial transcriptional profile is associated with early and late clinical response in rheumatoid arthritis patients treated with rituximab. *Ann Rheum Dis,* 71**,** 1888-94.
Horta-Baas, G., Romero-Figueroa, M. D. S., Montiel-Jarquin, A. J., Pizano-Zarate, M. L., Garcia-Mena, J. & Ramirez-Duran, N. 2017. Intestinal Dysbiosis and Rheumatoid Arthritis: A Link between Gut Microbiota and the Pathogenesis of Rheumatoid Arthritis. *J Immunol Res,* 2017**,** 4835189.
Hou, S., Landego, I., Jayachandran, N., Miller, A., Gibson, I. W., Ambrose, C. & Marshall, A. J. 2014. Follicular dendritic cell secreted protein FDC-SP controls IgA production. *Mucosal Immunol,* 7**,** 948-57.
Hozumi, Y., Wang, R., Yin, C. & Wei, G.-W. 2021. UMAP-assisted K-means clustering of large-scale SARS-CoV-2 mutation datasets. *Computers in biology and medicine,* 131**,** 104264-104264.
Hu, Y., Sparks, J. A., Malspeis, S., Costenbader, K. H., Hu, F. B., Karlson, E. W. & Lu, B. 2017. Long-term dietary quality and risk of developing rheumatoid arthritis in women. *Ann Rheum Dis,* 76**,** 1357-1364.
Humby, F., Lewis, M., Ramamoorthi, N., Hackney, J. A., Barnes, M. R., Bombardieri, M., Setiadi, A. F., Kelly, S., Bene, F., Dicicco, M., Riahi, S., Rocher, V., Ng, N., Lazarou, I., Hands, R., Van Der Heijde, D., Landewe, R. B. M., Van Der Helm-Van Mil, A., Cauli, A., Mcinnes, I., Buckley, C. D., Choy, E. H., Taylor, P. C., Townsend, M. J. & Pitzalis, C. 2019a. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann Rheum Dis*.
Humby, F., Lewis, M., Ramamoorthi, N., Hackney, J. A., Barnes, M. R., Bombardieri, M., Setiadi, A. F., Kelly, S., Bene, F., Dicicco, M., Riahi, S., Rocher, V., Ng, N., Lazarou, I., Hands, R., Van Der Heijde, D., Landewe, R. B. M., Van Der Helm-Van Mil, A., Cauli, A., Mcinnes, I., Buckley, C. D.,

Choy, E. H., Taylor, P. C., Townsend, M. J. & Pitzalis, C. 2019b. Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann Rheum Dis,* 78**,** 761-772.

Ibrahim, F., Lorente-Canovas, B., Dore, C. J., Bosworth, A., Ma, M. H., Galloway, J. B., Cope, A. P., Pande, I., Walker, D. & Scott, D. L. 2017. Optimizing treatment with tumour necrosis factor inhibitors in rheumatoid arthritis-a proof of principle and exploratory trial: is dose tapering practical in good responders? *Rheumatology (Oxford),* 56**,** 2004-2014.

Janke, J., Engeli, S., Gorzelniak, K. & Sharma, A. M. 2001. Extraction of total RNA from adipocytes. *Horm Metab Res,* 33**,** 213-5.

Jeffrey, K. L., Brummer, T., Rolph, M. S., Liu, S. M., Callejas, N. A., Grumont, R. J., Gillieron, C., Mackay, F., Grey, S., Camps, M., Rommel, C., Gerondakis, S. D. & Mackay, C. R. 2006. Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nat Immunol,* 7**,** 274-83.

Juarez, M., Bang, H., Hammar, F., Reimer, U., Dyke, B., Sahbudin, I., Buckley, C. D., Fisher, B., Filer, A. & Raza, K. 2016. Identification of novel antiacetylated vimentin antibodies in patients with early inflammatory arthritis. *Ann Rheum Dis,* 75**,** 1099-107.

Jubair, W. K., Hendrickson, J. D., Severs, E. L., Schulz, H. M., Adhikari, S., Ir, D., Pagan, J. D., Anthony, R. M., Robertson, C. E., Frank, D. N., Banda, N. K. & Kuhn, K. A. 2018. Modulation of Inflammatory Arthritis in Mice by Gut Microbiota Through Mucosal Inflammation and Autoantibody Generation. *Arthritis Rheumatol,* 70**,** 1220-1233.

Just, S. A., Humby, F., Lindegaard, H., Meric De Bellefon, L., Durez, P., Vieira-Sousa, E., Teixeira, R., Stoenoiu, M., Werlinrud, J., Rosmark, S., Larsen, P. V., Pratt, A., Choy, E., Gendi, N., Buch, M. H., Edwards, C. J., Taylor, P. C., Mcinnes, I. B., Fonseca, J. E., Pitzalis, C. & Filer, A. 2018. Patient-reported outcomes and safety in patients undergoing synovial biopsy: comparison of ultrasound-guided needle biopsy, ultrasound-guided portal and forceps and arthroscopic-guided synovial biopsy techniques in five centres across Europe. *RMD Open,* 4**,** e000799.

Kan, R., Shuen, W. H., Lung, H. L., Cheung, A. K., Dai, W., Kwong, D. L., Ng, W. T., Lee, A. W., Yau, C. C., Ngan, R. K., Tung, S. Y. & Lung, M. L. 2015. NF-kappaB p65 Subunit Is Modulated by Latent Transforming Growth Factor-beta Binding Protein 2 (LTBP2) in Nasopharyngeal Carcinoma HONE1 and HK1 Cells. *PLoS One,* 10**,** e0127239.

Kang, K. Y., Woo, J. W. & Park, S. H. 2014. S100A8/A9 as a biomarker for synovial inflammation and joint damage in patients with rheumatoid arthritis. *Korean J Intern Med,* 29**,** 12-9.

Kared, H., Martelli, S., Ng, T. P., Pender, S. L. & Larbi, A. 2016. CD57 in human natural killer cells and T-lymphocytes. *Cancer Immunol Immunother,* 65**,** 441-52.

Kassambara, A. & Mundt, F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. *R package version 1.0.7. https://CRAN.R-project.org/package=factoextra*.

Kelly, S., Humby, F., Filer, A., Ng, N., Di Cicco, M., Hands, R. E., Rocher, V., Bombardieri, M., D'agostino, M. A., Mcinnes, I. B., Buckley, C. D., Taylor, P. C. & Pitzalis, C. 2015. Ultrasound-guided synovial biopsy: a safe, well-

tolerated and reliable technique for obtaining high-quality synovial tissue from both large and small joints in early arthritis patients. *Ann Rheum Dis,* 74**,** 611-7.

Kennedy, T. D., Plater-Zyberk, C., Partridge, T. A., Woodrow, D. F. & Maini, R. N. 1988. Representative sample of rheumatoid synovium: a morphometric study. *J Clin Pathol,* 41**,** 841-6.

Keren, L., Bosse, M., Thompson, S., Risom, T., Vijayaragavan, K., Mccaffrey, E., Marquez, D., Angoshtari, R., Greenwald, N. F., Fienberg, H., Wang, J., Kambham, N., Kirkwood, D., Nolan, G., Montine, T. J., Galli, S. J., West, R., Bendall, S. C. & Angelo, M. 2019. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci Adv,* 5**,** eaax5851.

King, J. A., Corcoran, N. M., D'abaco, G. M., Straffon, A. F., Smith, C. T., Poon, C. L., Buchert, M., I, S., Hall, N. E., Lock, P. & Hovens, C. M. 2006. Eve-3: a liver enriched suppressor of Ras/MAPK signaling. *J Hepatol,* 44**,** 758-67.

Kingsley, G. H., Kowalczyk, A., Taylor, H., Ibrahim, F., Packham, J. C., Mchugh, N. J., Mulherin, D. M., Kitas, G. D., Chakravarty, K., Tom, B. D., O'keeffe, A. G., Maddison, P. J. & Scott, D. L. 2012. A randomized placebo-controlled trial of methotrexate in psoriatic arthritis. *Rheumatology (Oxford),* 51**,** 1368-77.

Koda, T., Namba, A., Kinoshita, M., Nakatsuji, Y., Sugimoto, T., Sakakibara, K., Tada, S., Shimizu, M., Yamashita, K., Takata, K., Ishikura, T., Murata, S., Beppu, S., Kumanogoh, A., Mochizuki, H. & Okuno, T. 2020. Sema4A is implicated in the acceleration of Th17 cell-mediated neuroinflammation in the effector phase. *J Neuroinflammation,* 17**,** 82.

Konopka, T. 2020. umap: Uniform Manifold Approximation and Projection. *R package version 0.2.7.0. https://CRAN.R-project.org/package=umap.*

Konttinen, Y. T., Salo, T., Hanemaaijer, R., Valleala, H., Sorsa, T., Sutinen, M., Ceponis, A., Xu, J. W., Santavirta, S., Teronen, O. & Lopez-Otin, C. 1999. Collagenase-3 (MMP-13) and its activators in rheumatoid arthritis: localization in the pannus-hard tissue junction and inhibition by alendronate. *Matrix Biol,* 18**,** 401-12.

Kraan, M. C., Haringman, J. J., Post, W. J., Versendaal, J., Breedveld, F. C. & Tak, P. P. 1999. Immunohistological analysis of synovial tissue for differential diagnosis in early arthritis. *Rheumatology (Oxford),* 38**,** 1074-80.

Krabben, A., Huizinga, T. W. & Van Der Helm-Van Mil, A. H. 2012. Undifferentiated arthritis characteristics and outcomes when applying the 2010 and 1987 criteria for rheumatoid arthritis. *Ann Rheum Dis,* 71**,** 238-41.

Kremlitzka, M., Polgar, A., Fulop, L., Kiss, E., Poor, G. & Erdei, A. 2013. Complement receptor type 1 (CR1, CD35) is a potent inhibitor of B-cell functions in rheumatoid arthritis patients. *Int Immunol,* 25**,** 25-33.

Krenn, V., Morawietz, L., Burmester, G. R., Kinne, R. W., Mueller-Ladner, U., Muller, B. & Haupl, T. 2006. Synovitis score: discrimination between chronic low-grade and high-grade synovitis. *Histopathology,* 49**,** 358-64.

Kuijper, T. M., Folmer, R., Stolk, E. A., Hazes, J. M. W. & Luime, J. J. 2017. Doctors' preferences in de-escalating DMARDs in rheumatoid arthritis: a discrete choice experiment. *Arthritis Res Ther,* 19**,** 78.

Kwak, D. H., Lee, J.-H., Kim, T., Ahn, H. S., Cho, W.-K., Ha, H., Hwang, Y.-H. & Ma, J. Y. 2012. Aristolochia manshuriensis Kom inhibits adipocyte differentiation by regulation of ERK1/2 and Akt pathway. *PloS one,* 7**,** e49530-e49530.

Kzhyshkowska, J., Gratchev, A. & Goerdt, S. 2007. Human chitinases and chitinase-like proteins as indicators for inflammation and cancer. *Biomark Insights,* 2**,** 128-46.

Lai, C. H., Chen, A. T., Burns, A. B., Sriram, K., Luo, Y., Tang, X., Branciamore, S., O'meally, D., Chang, S. L., Huang, P. H., Shyy, J. Y., Chien, S., Rockne, R. C. & Chen, Z. B. 2021. RAMP2-AS1 Regulates Endothelial Homeostasis and Aging. *Front Cell Dev Biol,* 9**,** 635307.

Langfelder, P. & Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

Lee, H.-Y., Lee, G.-H., Yoon, Y., Hoang, T.-H. & Chae, H.-J. 2022a. IBF-R Regulates IRE1α Post-Translational Modifications and ER Stress in High-Fat Diet-Induced Obese Mice. *Nutrients,* 14**,** 217.

Lee, S.-Y., Lee, C.-M., Ma, B., Kamle, S., Elias, J. A., Zhou, Y. & Lee, C. G. 2022b. Targeting Chitinase 1 and Chitinase 3-Like 1 as Novel Therapeutic Strategy of Pulmonary Fibrosis. *Frontiers in pharmacology,* 13**,** 826471-826471.

Lewis, M. J., Barnes, M. R., Blighe, K., Goldmann, K., Rana, S., Hackney, J. A., Ramamoorthi, N., John, C. R., Watson, D. S., Kummerfeld, S. K., Hands, R., Riahi, S., Rocher-Ros, V., Rivellese, F., Humby, F., Kelly, S., Bombardieri, M., Ng, N., Dicicco, M., Van Der Heijde, D., Landewe, R., Van Der Helm-Van Mil, A., Cauli, A., Mcinnes, I. B., Buckley, C. D., Choy, E., Taylor, P. C., Townsend, M. J. & Pitzalis, C. 2019. Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep,* 28**,** 2455-2470 e5.

Li, C., Liu, M., Liu, K., Li, M., Liu, Y., Li, T., Wei, Y., Long, Y., He, W., Shi, X., Li, Y. & Zhang, H. 2021. BATF2 balances the T cell-mediated immune response of CADM with an anti-MDA5 autoantibody. *Biochemical and Biophysical Research Communications,* 551**,** 155-160.

Li, C., Zhu, B., Son, Y. M., Wang, Z., Jiang, L., Xiang, M., Ye, Z., Beckermann, K. E., Wu, Y., Jenkins, J. W., Siska, P. J., Vincent, B. G., Prakash, Y. S., Peikert, T., Edelson, B. T., Taneja, R., Kaplan, M. H., Rathmell, J. C., Dong, H., Hitosugi, T. & Sun, J. 2019. The Transcription Factor Bhlhe40 Programs Mitochondrial Regulation of Resident CD8(+) T Cell Fitness and Functionality. *Immunity,* 51**,** 491-507 e7.

Li, Z. F., Wu, X. H. & Engvall, E. 2004. Identification and characterization of CPAMD8, a novel member of the complement 3/alpha2-macroglobulin family with a C-terminal Kazal domain. *Genomics,* 83**,** 1083-93.

Liao, L., Liang, K., Lan, L., Wang, J. & Guo, J. 2021. Marker Genes Change of Synovial Fibroblasts in Rheumatoid Arthritis Patients. *Biomed Res Int,* 2021**,** 5544264.

Liao, Y., Smyth, G. K. & Shi, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-30.

Lindskog, C., Linne, J., Fagerberg, L., Hallstrom, B. M., Sundberg, C. J., Lindholm, M., Huss, M., Kampf, C., Choi, H., Liem, D. A., Ping, P., Varemo, L., Mardinoglu, A., Nielsen, J., Larsson, E., Ponten, F. & Uhlen, M. 2015.

The human cardiac and skeletal muscle proteomes defined by transcriptomics and antibody-based profiling. *BMC Genomics,* 16**,** 475.
Listing, J., Gerhold, K. & Zink, A. 2013. The risk of infections associated with rheumatoid arthritis, with its comorbidity and treatment. *Rheumatology (Oxford),* 52**,** 53-61.
Liu, B., Zhang, Y., Fan, Y., Wang, S., Li, Z., Deng, M., Li, C., Wang, J., Ma, R., Wang, X., Wang, Y., Xu, L., Hou, K., Che, X., Liu, Y. & Qu, X. 2019. Leucine-rich repeat neuronal protein-1 suppresses apoptosis of gastric cancer cells through regulation of Fas/FasL. *Cancer Sci,* 110**,** 2145-2155.
Liu, H., Wei, S. P., Zhi, L. Q., Liu, L. P., Cao, T. P., Wang, S. Z., Chen, Q. P. & Liu, D. 2018. Synovial GATA1 mediates rheumatoid arthritis progression via transcriptional activation of NOS2 signaling. *Microbiol Immunol,* 62**,** 594-606.
Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.
Lu, D., Liu, L., Ji, X., Gao, Y., Chen, X., Liu, Y., Liu, Y., Zhao, X., Li, Y., Li, Y., Jin, Y., Zhang, Y., Mcnutt, M. A. & Yin, Y. 2015. The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation. *Nat Immunol,* 16**,** 1263-73.
Lugrin, J. & Martinon, F. 2018. The AIM2 inflammasome: Sensor of pathogens and cellular perturbations. *Immunol Rev,* 281**,** 99-114.
Luime, J. J., Colin, E. M., Hazes, J. M. W. & Lubberts, E. 2010. Does anti-mutated citrullinated vimentin have additional value as a serological marker in the diagnostic and prognostic investigation of patients with rheumatoid arthritis? A systematic review. *Annals of the Rheumatic Diseases,* 69**,** 337-344.
Maeda, Y., Kurakawa, T., Umemoto, E., Motooka, D., Ito, Y., Gotoh, K., Hirota, K., Matsushita, M., Furuta, Y., Narazaki, M., Sakaguchi, N., Kayama, H., Nakamura, S., Iida, T., Saeki, Y., Kumanogoh, A., Sakaguchi, S. & Takeda, K. 2016. Dysbiosis Contributes to Arthritis Development via Activation of Autoreactive T Cells in the Intestine. *Arthritis Rheumatol,* 68**,** 2646-2661.
Makki, M. S. & Haqqi, T. M. 2017. Histone deacetylase inhibitor vorinostat (SAHA, MK0683) perturb miR-9-MCPIP1 axis to block IL-1β-induced IL-6 expression in human OA chondrocytes. *Connective tissue research,* 58**,** 64-75.
Maksymowych, W. P., Naides, S. J., Bykerk, V., Siminovitch, K. A., Van Schaardenburg, D., Boers, M., Landewe, R., Van Der Heijde, D., Tak, P. P., Genovese, M. C., Weinblatt, M. E., Keystone, E. C., Zhukov, O. S., Abolhosn, R. W., Popov, J. M., Britsemmer, K., Van Kuijk, A. W. & Marotta, A. 2014. Serum 14-3-3 eta is a Novel Marker that Complements Current Serological Measurements to Enhance Detection of Patients with Rheumatoid Arthritis. *Journal of Rheumatology,* 41**,** 2104-2113.
Maney, N. J., Lemos, H., Barron-Millar, B., Carey, C., Herron, I., Anderson, A. E., Mellor, A. L., Isaacs, J. D. & Pratt, A. G. 2021. Pim Kinases as Therapeutic Targets in Early Rheumatoid Arthritis. *Arthritis Rheumatol,* 73**,** 1820-1830.
Manzo, A., Paoletti, S., Carulli, M., Blades, M. C., Barone, F., Yanni, G., Fitzgerald, O., Bresnihan, B., Caporali, R., Montecucco, C., Uguccioni, M. & Pitzalis, C. 2005. Systematic microanatomical analysis of CXCL13 and

CCL21 in situ production and progressive lymphoid organization in rheumatoid synovitis. *Eur J Immunol,* 35**,** 1347-59.
Marotte, H., Tsou, P. S., Rabquer, B. J., Pinney, A. J., Fedorova, T., Lalwani, N. & Koch, A. E. 2011. Blocking of interferon regulatory factor 1 reduces tumor necrosis factor alpha-induced interleukin-18 bioactivity in rheumatoid arthritis synovial fibroblasts by induction of interleukin-18 binding protein a: role of the nuclear interferon regulatory factor 1-NF-kappaB-c-jun complex. *Arthritis Rheum,* 63**,** 3253-62.
Marshall, A. J., Du, Q., Draves, K. E., Shikishima, Y., Hayglass, K. T. & Clark, E. A. 2002. FDC-SP, a novel secreted protein expressed by follicular dendritic cells. *J Immunol,* 169**,** 2381-9.
Martin, M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17.
Maynard, M. A., Evans, A. J., Hosomi, T., Hara, S., Jewett, M. A. & Ohh, M. 2005. Human HIF-3alpha4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. *FASEB J,* 19**,** 1396-406.
Mazur, M., Zielinska, A., Grzybowski, M. M., Olczak, J. & Fichna, J. 2021. Chitinases and Chitinase-Like Proteins as Therapeutic Targets in Inflammatory Diseases, with a Special Focus on Inflammatory Bowel Diseases. *Int J Mol Sci,* 22.
Mc Ardle, A., Flatley, B., Pennington, S. R. & Fitzgerald, O. 2015. Early biomarkers of joint damage in rheumatoid and psoriatic arthritis. *Arthritis Res Ther,* 17**,** 141.
Mckinney, C. & Merriman, T. R. 2012. Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes. *Hum Mol Genet,* 21**,** 2370-6.
Megarbane, A., Piquemal, D., Rebillat, A. S., Stora, S., Pierrat, F., Bruno, R., Noguier, F., Mircher, C., Ravel, A., Vilaire-Meunier, M., Durand, S. & Lefranc, G. 2020. Transcriptomic study in women with trisomy 21 identifies a possible role of the GTPases of the immunity-associated proteins (GIMAP) in the protection of breast cancer. *Sci Rep,* 10**,** 9447.
Meng, Q. & Qiu, B. 2020. Exosomal MicroRNA-320a Derived From Mesenchymal Stem Cells Regulates Rheumatoid Arthritis Fibroblast-Like Synoviocyte Activation by Suppressing CXCL9 Expression. *Front Physiol,* 11**,** 441.
Meng, Z., Ren, D., Zhang, K., Zhao, J., Jin, X. & Wu, H. 2020. Using ESTIMATE algorithm to establish an 8-mRNA signature prognosis prediction system and identify immunocyte infiltration-related genes in Pancreatic adenocarcinoma. *Aging (Albany NY),* 12**,** 5048-5070.
Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., Mckay-Fleisch, J., Nguyen, K., Norgaard, Z., Sorg, K., Sprague, I., Warren, C., Warren, S., Webster, P. J., Zhou, Z., Zollinger, D. R., Dunaway, D. L., Mills, G. B. & Beechem, J. M. 2020. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol,* 38**,** 586-599.
Metzemaekers, M., Vanheule, V., Janssens, R., Struyf, S. & Proost, P. 2017. Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. *Front Immunol,* 8**,** 1970.
Miao, T., Raymond, M., Bhullar, P., Ghaffari, E., Symonds, A. L., Meier, U. C., Giovannoni, G., Li, S. & Wang, P. 2013. Early growth response gene-2

controls IL-17 expression and Th17 differentiation by negatively regulating Batf. *J Immunol,* 190**,** 58-65.
Mirshafiey, A. & Mohsenzadegan, M. 2008. The role of reactive oxygen species in immunopathogenesis of rheumatoid arthritis. *Iran J Allergy Asthma Immunol,* 7**,** 195-202.
Moelants, E. A., Mortier, A., Van Damme, J. & Proost, P. 2013. Regulation of TNF-alpha with a focus on rheumatoid arthritis. *Immunol Cell Biol,* 91**,** 393-401.
Montero-Melendez, T., Nagano, A., Chelala, C., Filer, A., Buckley, C. D. & Perretti, M. 2020. Therapeutic senescence via GPCR activation in synovial fibroblasts facilitates resolution of arthritis. *Nat Commun,* 11**,** 745.
Monti, S., Montecucco, C., Bugatti, S. & Caporali, R. 2015. Rheumatoid arthritis treatment: the earlier the better to prevent joint damage. *RMD Open,* 1**,** e000057.
Moore, B. A., Aznavoorian, S., Engler, J. A. & Windsor, L. J. 2000. Induction of collagenase-3 (MMP-13) in rheumatoid arthritis synovial fibroblasts. *Biochim Biophys Acta,* 1502**,** 307-18.
Morinobu, A., Tanaka, S., Nishimura, K., Takahashi, S., Kageyama, G., Miura, Y., Kurosaka, M., Saegusa, J. & Kumagai, S. 2016. Expression and Functions of Immediate Early Response Gene X-1 (IEX-1) in Rheumatoid Arthritis Synovial Fibroblasts. *PLoS One,* 11**,** e0164350.
Munday, J., Kerr, S., Ni, J., Cornish, A. L., Zhang, J. Q., Nicoll, G., Floyd, H., Mattei, M. G., Moore, P., Liu, D. & Crocker, P. R. 2001. Identification, characterization and leucocyte expression of Siglec-10, a novel human sialic acid-binding receptor. *Biochem J,* 355**,** 489-97.
Murata, K., Yoshitomi, H., Tanida, S., Ishikawa, M., Nishitani, K., Ito, H. & Nakamura, T. 2010. Plasma and synovial fluid microRNAs as potential biomarkers of rheumatoid arthritis and osteoarthritis. *Arthritis Research & Therapy,* 12.
Myngbay, A., Bexeitov, Y., Adilbayeva, A., Assylbekov, Z., Yevstratenko, B. P., Aitzhanova, R. M., Matkarimov, B., Adarichev, V. A. & Kunz, J. 2019. CTHRC1: A New Candidate Biomarker for Improved Rheumatoid Arthritis Diagnosis. *Front Immunol,* 10**,** 1353.
Nagy, G., Koncz, A., Telarico, T., Fernandez, D., Ersek, B., Buzas, E. & Perl, A. 2010. Central role of nitric oxide in the pathogenesis of rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Res Ther,* 12**,** 210.
Nerviani, A., Di Cicco, M., Mahto, A., Lliso-Ribera, G., Rivellese, F., Thorborn, G., Hands, R., Bellan, M., Mauro, D., Boutet, M. A., Giorli, G., Lewis, M., Kelly, S., Bombardieri, M., Humby, F. & Pitzalis, C. 2020. A Pauci-Immune Synovial Pathotype Predicts Inadequate Response to TNFalpha-Blockade in Rheumatoid Arthritis Patients. *Front Immunol,* 11**,** 845.
Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., Khodadoust, M. S., Esfahani, M. S., Luca, B. A., Steiner, D., Diehn, M. & Alizadeh, A. A. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol,* 37**,** 773-782.
Noack, M. & Miossec, P. 2017. Selected cytokine pathways in rheumatoid arthritis. *Semin Immunopathol,* 39**,** 365-383.
Nzeusseu Toukap, A., Galant, C., Theate, I., Maudoux, A. L., Lories, R. J., Houssiau, F. A. & Lauwerys, B. R. 2007. Identification of distinct gene

expression profiles in the synovium of patients with systemic lupus erythematosus. *Arthritis Rheum,* 56**,** 1579-88.
O'neill, L. A., Kishton, R. J. & Rathmell, J. 2016. A guide to immunometabolism for immunologists. *Nat Rev Immunol,* 16**,** 553-65.
Ocampo, D. V. & Gladman, D. 2019. Psoriatic arthritis. *F1000Res,* 8.
Ometto, F., Friso, L., Astorri, D., Botsios, C., Raffeiner, B., Punzi, L. & Doria, A. 2017. Calprotectin in rheumatic diseases. *Exp Biol Med (Maywood),* 242**,** 859-873.
Orange, D. E., Agius, P., Dicarlo, E. F., Robine, N., Geiger, H., Szymonifka, J., Mcnamara, M., Cummings, R., Andersen, K. M., Mirza, S., Figgie, M., Ivashkiv, L. B., Pernis, A. B., Jiang, C. S., Frank, M. O., Darnell, R. B., Lingampali, N., Robinson, W. H., Gravallese, E., Accelerating Medicines Partnership in Rheumatoid, A., Lupus, N., Bykerk, V. P., Goodman, S. M. & Donlin, L. T. 2018. Identification of Three Rheumatoid Arthritis Disease Subtypes by Machine Learning Integration of Synovial Histologic Features and RNA Sequencing Data. *Arthritis Rheumatol,* 70**,** 690-701.
Orange, D. E., Blachere, N. E., Dicarlo, E. F., Mirza, S., Pannellini, T., Jiang, C. S., Frank, M. O., Parveen, S., Figgie, M. P., Gravallese, E. M., Bykerk, V. P., Orbai, A. M., Mackie, S. L. & Goodman, S. M. 2020. Rheumatoid Arthritis Morning Stiffness Is Associated With Synovial Fibrin and Neutrophils. *Arthritis Rheumatol,* 72**,** 557-564.
Pan, L., Ding, W., Li, J., Gan, K., Shen, Y., Xu, J. & Zheng, M. 2021. Aldehyde dehydrogenase 2 alleviates monosodium iodoacetate-induced oxidative stress, inflammation and apoptosis in chondrocytes via inhibiting aquaporin 4 expression. *Biomedical engineering online,* 20**,** 80-80.
Panigrahy, D., Gilligan, M. M., Serhan, C. N. & Kashfi, K. 2021. Resolution of inflammation: An organizing principle in biology and medicine. *Pharmacol Ther,* 227**,** 107879.
Park, S. H., Kim, S. K., Choe, J. Y., Moon, Y., An, S., Park, M. J. & Kim, D. S. 2013. Hypermethylation of EBF3 and IRX1 genes in synovial fibroblasts of patients with rheumatoid arthritis. *Mol Cells,* 35**,** 298-304.
Patterson, A. R., Endale, M., Lampe, K., Aksoylar, H. I., Flagg, A., Woodgett, J. R., Hildeman, D., Jordan, M. B., Singh, H., Kucuk, Z., Bleesing, J. & Hoebe, K. 2018. Gimap5-dependent inactivation of GSK3beta is required for CD4(+) T cell homeostasis and prevention of immune pathology. *Nat Commun,* 9**,** 430.
Pattison, D. J., Harrison, R. A. & Symmons, D. P. 2004a. The role of diet in susceptibility to rheumatoid arthritis: a systematic review. *J Rheumatol,* 31**,** 1310-9.
Pattison, D. J., Symmons, D. P., Lunt, M., Welch, A., Luben, R., Bingham, S. A., Khaw, K. T., Day, N. E. & Silman, A. J. 2004b. Dietary risk factors for the development of inflammatory polyarthritis: evidence for a role of high level of red meat consumption. *Arthritis Rheum,* 50**,** 3804-12.
Paulissen, S. M., Van Hamburg, J. P., Dankers, W. & Lubberts, E. 2015. The role and modulation of CCR6+ Th17 cell populations in rheumatoid arthritis. *Cytokine,* 74**,** 43-53.
Penkava, F., Velasco-Herrera, M. D. C., Young, M. D., Yager, N., Nwosu, L. N., Pratt, A. G., Lara, A. L., Guzzo, C., Maroof, A., Mamanova, L., Cole, S., Efremova, M., Simone, D., Filer, A., Brown, C. C., Croxford, A. L., Isaacs, J. D., Teichmann, S., Bowness, P., Behjati, S. & Hussein Al-Mossawi, M.

2020. Single-cell sequencing reveals clonal expansions of pro-inflammatory synovial CD8 T cells expressing tissue-homing receptors in psoriatic arthritis. *Nat Commun,* 11**,** 4767.
Perkins, D. J., St Clair, E. W., Misukonis, M. A. & Weinberg, J. B. 1998. Reduction of NOS2 overexpression in rheumatoid arthritis patients treated with anti-tumor necrosis factor alpha monoclonal antibody (cA2). *Arthritis Rheum,* 41**,** 2205-10.
Pino, S. C., O'sullivan-Murphy, B., Lidstone, E. A., Yang, C., Lipson, K. L., Jurczyk, A., Diiorio, P., Brehm, M. A., Mordes, J. P., Greiner, D. L., Rossini, A. A. & Bortell, R. 2009. CHOP mediates endoplasmic reticulum stress-induced apoptosis in Gimap5-deficient T cells. *PLoS One,* 4**,** e5468.
Pitzalis, C., Kelly, S. & Humby, F. 2013. New learnings on the pathophysiology of RA from synovial biopsies. *Curr Opin Rheumatol,* 25**,** 334-44.
Pressly, J. D., Gurumani, M. Z., Santos, J. T. V., Fornoni, A., Merscher, S. & Al-Ali, H. 2022. Adaptive and maladaptive roles of lipid droplets in health and disease. *American Journal of Physiology-Cell Physiology,* 322**,** C468-C481.
Pucino, V., Certo, M., Bulusu, V., Cucchi, D., Goldmann, K., Pontarini, E., Haas, R., Smith, J., Headland, S. E., Blighe, K., Ruscica, M., Humby, F., Lewis, M. J., Kamphorst, J. J., Bombardieri, M., Pitzalis, C. & Mauro, C. 2019. Lactate Buildup at the Site of Chronic Inflammation Promotes Disease by Inducing CD4(+) T Cell Metabolic Rewiring. *Cell Metab,* 30**,** 1055-1074 e8.
R Core Team 2021. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.*
Radner, H. & Aletaha, D. 2015. Anti-TNF in rheumatoid arthritis: an overview. *Wien Med Wochenschr,* 165**,** 3-9.
Rahbari, R., Zuccherato, L. W., Tischler, G., Chihota, B., Ozturk, H., Saleem, S., Tarazona-Santos, E., Machado, L. R. & Hollox, E. J. 2017. Understanding the Genomic Structure of Copy-Number Variation of the Low-Affinity Fcgamma Receptor Region Allows Confirmation of the Association of FCGR3B Deletion with Rheumatoid Arthritis. *Hum Mutat,* 38**,** 390-399.
Rahmani, M., Chegini, H., Najafizadeh, S. R., Azimi, M., Habibollahi, P. & Shakiba, M. 2010. Detection of bone erosion in early rheumatoid arthritis: ultrasonography and conventional radiography versus non-contrast magnetic resonance imaging. *Clinical Rheumatology,* 29**,** 883-891.
Rao, D. A., Gurish, M. F., Marshall, J. L., Slowikowski, K., Fonseka, C. Y., Liu, Y., Donlin, L. T., Henderson, L. A., Wei, K., Mizoguchi, F., Teslovich, N. C., Weinblatt, M. E., Massarotti, E. M., Coblyn, J. S., Helfgott, S. M., Lee, Y. C., Todd, D. J., Bykerk, V. P., Goodman, S. M., Pernis, A. B., Ivashkiv, L. B., Karlson, E. W., Nigrovic, P. A., Filer, A., Buckley, C. D., Lederer, J. A., Raychaudhuri, S. & Brenner, M. B. 2017. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature,* 542**,** 110-114.
Raza, K., Falciani, F., Curnow, S. J., Ross, E. J., Lee, C. Y., Akbar, A. N., Lord, J. M., Gordon, C., Buckley, C. D. & Salmon, M. 2005a. Early rheumatoid arthritis is characterized by a distinct and transient synovial fluid cytokine profile of T cell and stromal cell origin. *Arthritis Res Ther,* 7**,** R784-95.
Raza, K., Falciani, F., Curnow, S. J., Ross, E. J., Lee, C. Y., Akbar, A. N., Lord, J. M., Gordon, C., Buckley, C. D. & Salmon, M. 2005b. Early rheumatoid arthritis is characterized by a distinct and transient synovial fluid cytokine

profile of T cell and stromal cell origin. *Arthritis Research & Therapy,* 7**,** R784-R795.
Reece, R. J., Canete, J. D., Parsons, W. J., Emery, P. & Veale, D. J. 1999. Distinct vascular patterns of early synovitis in psoriatic, reactive, and rheumatoid arthritis. *Arthritis Rheum,* 42**,** 1481-4.
Reily, M. M., Pantoja, C., Hu, X., Chinenov, Y. & Rogatsky, I. 2006. The GRIP1:IRF3 interaction as a target for glucocorticoid receptor-mediated immunosuppression. *EMBO J,* 25**,** 108-17.
Renaudineau, Y., Jamin, C., Saraux, A. & Youinou, P. 2005. Rheumatoid factor on a daily basis. *Autoimmunity,* 38**,** 11-6.
Rogier, C., Hayer, S. & Van Der Helm-Van Mil, A. 2020. Not only synovitis but also tenosynovitis needs to be considered: why it is time to update textbook images of rheumatoid arthritis. *Ann Rheum Dis,* 79**,** 546-547.
Rooney, M., Condell, D., Quinlan, W., Daly, L., Whelan, A., Feighery, C. & Bresnihan, B. 1988. Analysis of the histologic variation of synovitis in rheumatoid arthritis. *Arthritis Rheum,* 31**,** 956-63.
Rstudio Team 2021. RStudio: Integrated Development Environment for R. *RStudio, PBC, Boston, MA URL http://www.rstudio.com/*.
Safford, M., Collins, S., Lutz, M. A., Allen, A., Huang, C. T., Kowalski, J., Blackford, A., Horton, M. R., Drake, C., Schwartz, R. H. & Powell, J. D. 2005. Egr-2 and Egr-3 are negative regulators of T cell activation. *Nat Immunol,* 6**,** 472-80.
Sahbudin, I., Pickup, L., Nightingale, P., Allen, G., Cader, Z., Singh, R., De Pablo, P., Buckley, C. D., Raza, K. & Filer, A. 2018. The role of ultrasound-defined tenosynovitis and synovitis in the prediction of rheumatoid arthritis development. *Rheumatology (Oxford),* 57**,** 1243-1252.
Salemme, V., Angelini, C., Chapelle, J., Centonze, G., Natalini, D., Morellato, A., Taverna, D., Turco, E., Ala, U. & Defilippi, P. 2021. The p140Cap adaptor protein as a molecular hub to block cancer aggressiveness. *Cellular and molecular life sciences : CMLS,* 78**,** 1355-1367.
Sanjabi, S., Oh, S. A. & Li, M. O. 2017. Regulation of the Immune Response by TGF-beta: From Conception to Autoimmunity and Infection. *Cold Spring Harb Perspect Biol,* 9.
Scarsi, M., Ziglioli, T. & Airo, P. 2010. Decreased circulating CD28-negative T cells in patients with rheumatoid arthritis treated with abatacept are correlated with clinical response. *J Rheumatol,* 37**,** 911-6.
Schneider, M. & Kruger, K. 2013. Rheumatoid arthritis--early diagnosis and disease management. *Dtsch Arztebl Int,* 110**,** 477-84.
Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H. C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., Mcgrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J. M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C. S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., Eichler, E. E. & Church, D. M. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res,* 27**,** 849-864.

Schonfeldova, B., Zec, K. & Udalova, I. A. 2022. Synovial single-cell heterogeneity, zonation and interactions: a patchwork of effectors in arthritis. *Rheumatology (Oxford, England),* 61**,** 913-925.
Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. & Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology,* 7**,** 3.
Selmi, C. & Gershwin, M. E. 2014. Diagnosis and classification of reactive arthritis. *Autoimmun Rev,* 13**,** 546-9.
Seror, R., Boudaoud, S., Pavy, S., Nocturne, G., Schaeverbeke, T., Saraux, A., Chanson, P., Gottenberg, J. E., Devauchelle-Pensec, V., Tobon, G. J., Mariette, X. & Miceli-Richard, C. 2016. Increased Dickkopf-1 in Recent-onset Rheumatoid Arthritis is a New Biomarker of Structural Severity. Data from the ESPOIR Cohort. *Sci Rep,* 6**,** 18421.
Shenoy, A. R., Wellington, D. A., Kumar, P., Kassa, H., Booth, C. J., Cresswell, P. & Macmicking, J. D. 2012. GBP5 promotes NLRP3 inflammasome assembly and immunity in mammals. *Science,* 336**,** 481-5.
Shi, J., Van Steenbergen, H. W., Van Nies, J. A., Levarht, E. W., Huizinga, T. W., Van Der Helm-Van Mil, A. H., Toes, R. E. & Trouw, L. A. 2015. The specificity of anti-carbamylated protein antibodies for rheumatoid arthritis in a setting of early arthritis. *Arthritis Res Ther,* 17**,** 339.
Sidhu, N., Wouters, F., Niemantsverdriet, E. & Van Der Helm-Van Mil, A. H. M. 2021. MRI detected synovitis of the small joints predicts rheumatoid arthritis development in large joint undifferentiated inflammatory arthritis. *Rheumatology (Oxford)*.
Siebert, S., Pratt, A. G., Stocken, D. D., Morton, M., Cranston, A., Cole, M., Frame, S., Buckley, C. D., Ng, W. F., Filer, A., Mcinnes, I. B. & Isaacs, J. D. 2020. Targeting the rheumatoid arthritis synovial fibroblast via cyclin dependent kinase inhibition: An early phase trial. *Medicine (Baltimore),* 99**,** e20458.
Singh, A., Rajasekaran, N., Hartenstein, B., Szabowski, S., Gajda, M., Angel, P., Brauer, R. & Illges, H. 2013. Collagenase-3 (MMP-13) deficiency protects C57BL/6 mice from antibody-induced arthritis. *Arthritis Res Ther,* 15**,** R222.
Smith, M. D. 2011. The normal synovium. *Open Rheumatol J,* 5**,** 100-6.
Smolen, J. S., Aletaha, D. & Mcinnes, I. B. 2016a. Rheumatoid arthritis. *Lancet,* 388**,** 2023-2038.
Smolen, J. S., Breedveld, F. C., Burmester, G. R., Bykerk, V., Dougados, M., Emery, P., Kvien, T. K., Navarro-Compan, M. V., Oliver, S., Schoels, M., Scholte-Voshaar, M., Stamm, T., Stoffer, M., Takeuchi, T., Aletaha, D., Andreu, J. L., Aringer, M., Bergman, M., Betteridge, N., Bijlsma, H., Burkhardt, H., Cardiel, M., Combe, B., Durez, P., Fonseca, J. E., Gibofsky, A., Gomez-Reino, J. J., Graninger, W., Hannonen, P., Haraoui, B., Kouloumas, M., Landewe, R., Martin-Mola, E., Nash, P., Ostergaard, M., Ostor, A., Richards, P., Sokka-Isler, T., Thorne, C., Tzioufas, A. G., Van Vollenhoven, R., De Wit, M. & Van Der Heijde, D. 2016b. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis,* 75**,** 3-15.
Somerville, T. D. D., Xu, Y., Wu, X. S., Maia-Silva, D., Hur, S. K., De Almeida, L. M. N., Preall, J. B., Koo, P. K. & Vakoc, C. R. 2020. ZBED2 is an antagonist

of interferon regulatory factor 1 and modifies cell identity in pancreatic cancer. *Proc Natl Acad Sci U S A,* 117**,** 11471-11482.
Song, Y., Pan, Y. & Liu, J. 2019. The relevance between the immune response-related gene module and clinical traits in head and neck squamous cell carcinoma. *Cancer Manag Res,* 11**,** 7455-7472.
Srinivasa, S., Garcia-Martin, R., Torriani, M., Fitch, K. V., Carlson, A. R., Kahn, C. R. & Grinspoon, S. K. 2021. Altered pattern of circulating miRNAs in HIV lipodystrophy perturbs key adipose differentiation and inflammation pathways. *JCI Insight,* 6.
Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, A., Ponten, F., Costea, P. I., Sahlen, P., Mulder, J., Bergmann, O., Lundeberg, J. & Frisen, J. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science,* 353**,** 78-82.
Stanford, S. M. & Bottini, N. 2014. PTPN22: the archetypal non-HLA autoimmunity gene. *Nature Reviews Rheumatology,* 10**,** 602-611.
Stockman, A., Tait, B. D., Wolfe, R., Brand, C. A., Rowley, M. J., Varney, M. D., Buchbinder, R. & Muirden, K. D. 2006. Clinical, laboratory and genetic markers associated with erosions and remission in patients with early inflammatory arthritis: a prospective cohort study. *Rheumatology International,* 26**,** 500-509.
Stoilov, I., Krueger, W., Mankowski, D., Guernsey, L., Kaur, A., Glynn, J. & Thrall, R. S. 2006. The cytochromes P450 (CYP) response to allergic inflammation of the lung. *Arch Biochem Biophys,* 456**,** 30-8.
Su, C. T., Huang, J. W., Chiang, C. K., Lawrence, E. C., Levine, K. L., Dabovic, B., Jung, C., Davis, E. C., Madan-Khetarpal, S. & Urban, Z. 2015a. Latent transforming growth factor binding protein 4 regulates transforming growth factor beta receptor stability. *Hum Mol Genet,* 24**,** 4024-36.
Su, J., Gao, T., Shi, T., Xiang, Q., Xu, X., Wiesenfeld-Hallin, Z., Hokfelt, T. & Svensson, C. I. 2015b. Phenotypic changes in dorsal root ganglion and spinal cord in the collagen antibody-induced arthritis mouse model. *J Comp Neurol,* 523**,** 1505-28.
Susman, M. W., Karuna, E. P., Kunz, R. C., Gujral, T. S., Cantu, A. V., Choi, S. S., Jong, B. Y., Okada, K., Scales, M. K., Hum, J., Hu, L. S., Kirschner, M. W., Nishinakamura, R., Yamada, S., Laird, D. J., Jao, L. E., Gygi, S. P., Greenberg, M. E. & Ho, H. H. 2017. Kinesin superfamily protein Kif26b links Wnt5a-Ror signaling to the control of cell and tissue behaviors in vertebrates. *Elife,* 6.
Suzuki, K., Yoshida, K., Ueha, T., Kaneshiro, K., Nakai, A., Hashimoto, N., Uchida, K., Hashimoto, T., Kawasaki, Y., Shibanuma, N., Nakagawa, N., Sakai, Y. & Hashiramoto, A. 2018. Methotrexate upregulates circadian transcriptional factors PAR bZIP to induce apoptosis on rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther,* 20**,** 55.
Szumilas, K., Szumilas, P., Sluczanowska-Glabowska, S., Zgutka, K. & Pawlik, A. 2020. Role of Adiponectin in the Pathogenesis of Rheumatoid Arthritis. *Int J Mol Sci,* 21.
Tada, Y., Ho, A., Matsuyama, T. & Mak, T. W. 1997. Reduced incidence and severity of antigen-induced autoimmune diseases in mice lacking interferon regulatory factor-1. *J Exp Med,* 185**,** 231-8.

Tak, P. P., Smeets, T. J., Daha, M. R., Kluin, P. M., Meijers, K. A., Brand, R., Meinders, A. E. & Breedveld, F. C. 1997. Analysis of the synovial cell infiltrate in early rheumatoid synovial tissue in relation to local disease activity. *Arthritis Rheum,* 40**,** 217-25.

Taylor, W., Gladman, D., Helliwell, P., Marchesoni, A., Mease, P., Mielants, H. & Group, C. S. 2006. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. *Arthritis Rheum,* 54**,** 2665-73.

Thesseling, F. A., Hutter, M. C., Wiek, C., Kowalski, J. P., Rettie, A. E. & Girhard, M. 2020. Novel insights into oxidation of fatty acids and fatty alcohols by cytochrome P450 monooxygenase CYP4B1. *Arch Biochem Biophys,* 679**,** 108216.

Thieblemont, N., Wright, H. L., Edwards, S. W. & Witko-Sarsat, V. 2016. Human neutrophils in auto-immunity. *Semin Immunol,* 28**,** 159-73.

Thompson, S. J., Thompson, S. E. M. & Cazier, J.-B. 2019. CaStLeS (Compute and Storage for the Life Sciences): a collection of compute and storage resources for supporting research at the University of Birmingham.

Tracy, A., Buckley, C. D. & Raza, K. 2017. Pre-symptomatic autoimmunity in rheumatoid arthritis: when does the disease start? *Semin Immunopathol,* 39**,** 423-435.

Tseng, S., Reddi, A. H. & Di Cesare, P. E. 2009. Cartilage Oligomeric Matrix Protein (COMP): A Biomarker of Arthritis. *Biomark Insights,* 4**,** 33-44.

Tsubaki, T., Takegawa, S., Hanamoto, H., Arita, N., Kamogawa, J., Yamamoto, H., Takubo, N., Nakata, S., Yamada, K., Yamamoto, S., Yoshie, O. & Nose, M. 2005. Accumulation of plasma cells expressing CXCR3 in the synovial sublining regions of early rheumatoid arthritis in association with production of Mig/CXCL9 by synovial fibroblasts. *Clin Exp Immunol,* 141**,** 363-71.

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., Von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., Von Heijne, G., Nielsen, J. & Ponten, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.

Ustyugova, I. V., Zhi, L. & Wu, M. X. 2012. Reciprocal regulation of the survival and apoptosis of Th17 and Th1 cells in the colon. *Inflamm Bowel Dis,* 18**,** 333-43.

Van Aken, J., Van Bilsen, J. H., Allaart, C. F., Huizinga, T. W. & Breedveld, F. C. 2003. The Leiden Early Arthritis Clinic. *Clin Exp Rheumatol,* 21**,** S100-5.

Van Baarsen, L. G., Wijbrandts, C. A., Timmer, T. C., Van Der Pouw Kraan, T. C., Tak, P. P. & Verweij, C. L. 2010. Synovial tissue heterogeneity in rheumatoid arthritis in relation to disease activity and biomarkers in peripheral blood. *Arthritis Rheum,* 62**,** 1602-7.

Van Der Helm-Van Mil, A. H. M., Le Cessie, S., Van Dongen, H., Breedveld, F. C., Toes, R. E. M. & Huizinga, T. W. J. 2007. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis - How to guide individual treatment decisions. *Arthritis and Rheumatism,* 56**,** 433-440.

Van Der Pouw Kraan, T. C., Van Gaalen, F. A., Huizinga, T. W., Pieterman, E., Breedveld, F. C. & Verweij, C. L. 2003. Discovery of distinctive gene expression profiles in rheumatoid synovium using cDNA microarray technology: evidence for the existence of multiple pathways of tissue destruction and repair. *Genes Immun,* 4**,** 187-96.

Van Gestel, A. M., Prevoo, M. L., Van 'T Hof, M. A., Van Rijswijk, M. H., Van De Putte, L. B. & Van Riel, P. L. 1996. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum,* 39**,** 34-40.

Van Nies, J. A., Tsonaka, R., Gaujoux-Viala, C., Fautrel, B. & Van Der Helm-Van Mil, A. H. 2015. Evaluating relationships between symptom duration and persistence of rheumatoid arthritis: does a window of opportunity exist? Results on the Leiden early arthritis clinic and ESPOIR cohorts. *Ann Rheum Dis,* 74**,** 806-12.

Veale, D. J. & Fearon, U. 2018. The pathogenesis of psoriatic arthritis. *Lancet,* 391**,** 2273-2284.

Versteeg, G. A., Steunebrink, L. M. M., Vonkeman, H. E., Ten Klooster, P. M., Van Der Bijl, A. E. & Van De Laar, M. 2018. Long-term disease and patient-reported outcomes of a continuous treat-to-target approach in patients with early rheumatoid arthritis in daily clinical practice. *Clin Rheumatol*.

Wagle, M. C., Kirouac, D., Klijn, C., Liu, B., Mahajan, S., Junttila, M., Moffat, J., Merchant, M., Huw, L., Wongchenko, M., Okrah, K., Srinivasan, S., Mounir, Z., Sumiyoshi, T., Haverty, P. M., Yauch, R. L., Yan, Y., Kabbarah, O., Hampton, G., Amler, L., Ramanujan, S., Lackner, M. R. & Huang, S. A. 2018. A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. *NPJ Precis Oncol,* 2**,** 7.

Wang, E. C., Lawson, T. M., Vedhara, K., Moss, P. A., Lehner, P. J. & Borysiewicz, L. K. 1997. CD8high+ (CD57+) T cells in patients with rheumatoid arthritis. *Arthritis Rheum,* 40**,** 237-48.

Wang, W., Wang, L., Gulko, P. S. & Zhu, J. 2019a. Computational deconvolution of synovial tissue cellular composition: presence of adipocytes in synovial tissue decreased during arthritis pathogenesis and progression. *Physiol Genomics,* 51**,** 241-253.

Wang, W., Zhu, X., Du, X., Xu, A., Yuan, X., Zhan, Y., Liu, M. & Wang, S. 2019b. MiR-150 promotes angiogensis and proliferation of endothelial progenitor cells in deep venous thrombosis by targeting SRCIN1. *Microvascular Research,* 123**,** 35-41.

Waters, K. M., Sontag, R. L. & Weber, T. J. 2013. Hepatic leukemia factor promotes resistance to cell death: implications for therapeutics and chronotherapy. *Toxicol Appl Pharmacol,* 268**,** 141-8.

Wei, K., Korsunsky, I., Marshall, J. L., Gao, A., Watts, G. F. M., Major, T., Croft, A. P., Watts, J., Blazar, P. E., Lange, J. K., Thornhill, T. S., Filer, A., Raza, K., Donlin, L. T., Accelerating Medicines Partnership Rheumatoid, A., Systemic Lupus Erythematosus, C., Siebel, C. W., Buckley, C. D., Raychaudhuri, S. & Brenner, M. B. 2020. Notch signalling drives synovial fibroblast identity and arthritis pathology. *Nature,* 582**,** 259-264.

Wells, P. M., Adebayo, A. S., Bowyer, R. C. E., Freidin, M. B., Finckh, A., Strowig, T., Lesker, T. R., Alpizar-Rodriguez, D., Gilbert, B., Kirkham, B., Cope, A. P., Steves, C. J. & Williams, F. M. K. 2020. Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol,* 2**,** e418-e427.

Wernicke, D., Seyfert, C., Hinzmann, B. & Gromnica-Ihle, E. 1996. Cloning of collagenase 3 from the synovial membrane and its expression in rheumatoid arthritis and osteoarthritis. *J Rheumatol,* 23**,** 590-5.

Weyand, C. M., Wu, B. & Goronzy, J. J. 2020. The metabolic signature of T cells in rheumatoid arthritis. *Curr Opin Rheumatol,* 32**,** 159-167.

Wiggs, J. L. 2020. CPAMD8, a New Gene for Anterior Segment Dysgenesis and Childhood Glaucoma. *Ophthalmology,* 127**,** 767-768.

Wolins, N. E., Quaynor, B. K., Skinner, J. R., Schoenfish, M. J., Tzekov, A. & Bickel, P. E. 2005. S3-12, Adipophilin, and TIP47 package lipid in adipocytes. *J Biol Chem,* 280**,** 19146-55.

Wong, P. K., Egan, P. J., Croker, B. A., O'donnell, K., Sims, N. A., Drake, S., Kiu, H., Mcmanus, E. J., Alexander, W. S., Roberts, A. W. & Wicks, I. P. 2006. SOCS-3 negatively regulates innate and adaptive immune mechanisms in acute IL-1-dependent inflammatory arthritis. *J Clin Invest,* 116**,** 1571-81.

Wruck, C. J., Fragoulis, A., Gurzynski, A., Brandenburg, L. O., Kan, Y. W., Chan, K., Hassenpflug, J., Freitag-Wolf, S., Varoga, D., Lippross, S. & Pufe, T. 2011. Role of oxidative stress in rheumatoid arthritis: insights from the Nrf2-knockout mice. *Ann Rheum Dis,* 70**,** 844-50.

Wu, B., Goronzy, J. J. & Weyand, C. M. 2020. Metabolic Fitness of T Cells in Autoimmune Disease. *Immunometabolism,* 2.

Wu, H. J., Ivanov, Ii, Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D. R., Benoist, C. & Mathis, D. 2010. Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity,* 32**,** 815-27.

Xi, Y., Jing, Z., Haihong, L., Yizhen, J., Weili, G. & Shuwen, H. 2020. Analysis of T lymphocyte-related biomarkers in pancreatic cancer. *Pancreatology,* 20**,** 1502-1510.

Xu, Z., Ke, T., Zhang, Y., Fu, C. & He, W. 2020. Agonism of GPR120 prevented IL-1β-induced reduction of extracellular matrix through SOX-9. *Aging,* 12**,** 12074-12085.

Yang, A. & Mottillo, E. P. 2020. Adipocyte lipolysis: from molecular mechanisms of regulation to disease and therapeutics. *Biochem J,* 477**,** 985-1008.

Yang, A., Mottillo, E. P., Mladenovic-Lucas, L., Zhou, L. & Granneman, J. G. 2019. Dynamic interactions of ABHD5 with PNPLA3 regulate triacylglycerol metabolism in brown adipocytes. *Nat Metab,* 1**,** 560-569.

Yang, X. Y., Lin, K., Ni, S. M., Wang, J. M., Tian, Q. Q., Chen, H. J., Brown, M. A., Zheng, K. D., Zhai, W. T., Sun, L., Jin, S. W. & Wang, J. G. 2017. Serum connective tissue growth factor is a highly discriminatory biomarker for the diagnosis of rheumatoid arthritis. *Arthritis Research & Therapy,* 19.

Yeo, L., Adlard, N., Biehl, M., Juarez, M., Smallie, T., Snow, M., Buckley, C. D., Raza, K., Filer, A. & Scheel-Toellner, D. 2016. Expression of chemokines CXCL4 and CXCL7 by synovial macrophages defines an early stage of rheumatoid arthritis. *Ann Rheum Dis,* 75**,** 763-71.

Yin, Y., Liu, W. & Dai, Y. 2015. SOCS3 and its role in associated diseases. *Hum Immunol,* 76**,** 775-80.

Yin, Y., Liu, Y. X., Jin, Y. J., Hall, E. J. & Barrett, J. C. 2003. PAC1 phosphatase is a transcription target of p53 in signalling apoptosis and growth suppression. *Nature,* 422**,** 527-31.
Yokoyama-Kokuryo, W., Yamazaki, H., Takeuchi, T., Amano, K., Kikuchi, J., Kondo, T., Nakamura, S., Sakai, R., Hirano, F., Nanki, T., Koike, R. & Harigai, M. 2020. Identification of molecules associated with response to abatacept in patients with rheumatoid arthritis. *Arthritis research & therapy,* 22**,** 46-46.
You, Y., Zhou, C., Li, D., Cao, Z. L., Shen, W., Li, W. Z., Zhang, S., Hu, B. & Shen, X. 2016. Sorting nexin 10 acting as a novel regulator of macrophage polarization mediates inflammatory response in experimental mouse colitis. *Sci Rep,* 6**,** 20630.
Yu, G., Wang, L. G., Han, Y. & He, Q. Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS,* 16**,** 284-7.
Yu, N. Y., Hallstrom, B. M., Fagerberg, L., Ponten, F., Kawaji, H., Carninci, P., Forrest, A. R., Fantom, C., Hayashizaki, Y., Uhlen, M. & Daub, C. O. 2015. Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic Acids Res,* 43**,** 6787-98.
Yuan, F. L., Li, X., Lu, W. G., Sun, J. M., Jiang, D. L. & Xu, R. S. 2013. Epidermal growth factor receptor (EGFR) as a therapeutic target in rheumatoid arthritis. *Clin Rheumatol,* 32**,** 289-92.
Zaiss, M. M., Joyce Wu, H. J., Mauro, D., Schett, G. & Ciccia, F. 2021. The gut-joint axis in rheumatoid arthritis. *Nat Rev Rheumatol,* 17**,** 224-237.
Zhang, F., Luo, K., Rong, Z., Wang, Z., Luo, F., Zhang, Z., Sun, D., Dong, S., Xu, J. & Dai, F. 2021a. Author Correction: Periostin Upregulates Wnt/beta-Catenin Signaling to Promote the Osteogenesis of CTLA4-Modified Human Bone Marrow-Mesenchymal Stem Cells. *Sci Rep,* 11**,** 9598.
Zhang, F., Wei, K., Slowikowski, K., Fonseka, C. Y., Rao, D. A., Kelly, S., Goodman, S. M., Tabechian, D., Hughes, L. B., Salomon-Escoto, K., Watts, G. F. M., Jonsson, A. H., Rangel-Moreno, J., Meednu, N., Rozo, C., Apruzzese, W., Eisenhaure, T. M., Lieb, D. J., Boyle, D. L., Mandelin, A. M., 2nd, Accelerating Medicines Partnership Rheumatoid, A., Systemic Lupus Erythematosus, C., Boyce, B. F., Dicarlo, E., Gravallese, E. M., Gregersen, P. K., Moreland, L., Firestein, G. S., Hacohen, N., Nusbaum, C., Lederer, J. A., Perlman, H., Pitzalis, C., Filer, A., Holers, V. M., Bykerk, V. P., Donlin, L. T., Anolik, J. H., Brenner, M. B. & Raychaudhuri, S. 2019. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol,* 20**,** 928-942.
Zhang, H. G., Wang, Y., Xie, J. F., Liang, X., Liu, D., Yang, P., Hsu, H. C., Ray, R. B. & Mountz, J. D. 2001. Regulation of tumor necrosis factor alpha-mediated apoptosis of rheumatoid arthritis synovial fibroblasts by the protein kinase Akt. *Arthritis Rheum,* 44**,** 1555-67.
Zhang, J., Wang, L., Xu, X., Li, X., Guan, W., Meng, T. & Xu, G. 2020a. Transcriptome-Based Network Analysis Unveils Eight Immune-Related Genes as Molecular Signatures in the Immunomodulatory Subtype of Triple-Negative Breast Cancer. *Front Oncol,* 10**,** 1787.

Zhang, P., Lu, X., Tao, K., Shi, L., Li, W., Wang, G. & Wu, K. 2015a. Siglec-10 is associated with survival and natural killer cell dysfunction in hepatocellular carcinoma. *J Surg Res,* 194**,** 107-13.

Zhang, W., Dai, L., Li, X., Li, Y., Hung Yap, M. K., Liu, L. & Deng, H. 2020b. SARI prevents ocular angiogenesis and inflammation in mice. *Journal of cellular and molecular medicine,* 24**,** 4341-4349.

Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., Lan, Z., Chen, B., Li, Y., Zhong, H., Xie, H., Jie, Z., Chen, W., Tang, S., Xu, X., Wang, X., Cai, X., Liu, S., Xia, Y., Li, J., Qiao, X., Al-Aama, J. Y., Chen, H., Wang, L., Wu, Q. J., Zhang, F., Zheng, W., Li, Y., Zhang, M., Luo, G., Xue, W., Xiao, L., Li, J., Chen, W., Xu, X., Yin, Y., Yang, H., Wang, J., Kristiansen, K., Liu, L., Li, T., Huang, Q., Li, Y. & Wang, J. 2015b. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med,* 21**,** 895-905.

Zhang, Y., Liu, Q., Yang, S. & Liao, Q. 2021b. Knockdown of LRRN1 inhibits malignant phenotypes through the regulation of HIF-1alpha/Notch pathway in pancreatic ductal adenocarcinoma. *Mol Ther Oncolytics,* 23**,** 51-64.

Zheng, Y. M., Fisher, M. B., Yokotani, N., Fujii-Kuriyama, Y. & Rettie, A. E. 1998. Identification of a meander region proline residue critical for heme binding to cytochrome P450: implications for the catalytic function of human CYP4B1. *Biochemistry,* 37**,** 12847-51.

Zhi, L., Ustyugova, I. V., Chen, X., Zhang, Q. & Wu, M. X. 2012. Enhanced Th17 differentiation and aggravated arthritis in IEX-1-deficient mice by mitochondrial reactive oxygen species-mediated signaling. *J Immunol,* 189**,** 1639-47.

Zhou, C., Wang, Y., Peng, J., Li, C., Liu, P. & Shen, X. 2017. SNX10 Plays a Critical Role in MMP9 Secretion via JNK-p38-ERK Signaling Pathway. *J Cell Biochem,* 118**,** 4664-4671.

Zohar, Y., Wildbaum, G., Novak, R., Salzman, A. L., Thelen, M., Alon, R., Barsheshet, Y., Karp, C. L. & Karin, N. 2014. CXCL11-dependent induction of FOXP3-negative regulatory T cells suppresses autoimmune encephalomyelitis. *J Clin Invest,* 124**,** 2009-22.

9 APPENDIX

9.1 PATIENT CHARACTERISTICS

Table 9.1: Patient characteristics for the BEACON cohort. Samples from this cohort used for the BEACON scoring in Chapter 3. ACPA = Anti-citrullinated protein antibodies, RF = Rheumatoid Factor, SJC28 = Swollen Joint Count 28, TJC28 = Tender Joint Count 28, CRP = C Reactive Protein, ESR = Erythrocyte Sedimentation Rate, DAS28-ESR = Disease Activity Score 28, VAS = Visual Analog Scale, US GS = Ultrasound Greyscale, US PD = Ultrasound Power Doppler, DAS28-ESR response = EULAR DAS28-ESR response at 12 months, response includes good and moderate response, NA = data not available.

Clinical group	Final classification	Age	Sex	ACPA	RF	SJC28	TJC28	CRP	ESR	DAS28-ESR	Physician global	Patient VAS	US GS	US PD	DAS28-ESR response 12 months
Res	UA	64	M	n	n	2	5	15	24	4.5	11	77	1	0	NA
Res	Parvovirus	40	F	n	n	7	7	0	5	3.9	NA	15	1	1	NA
sdRA	RA	49	F	n	n	8	9	8	12	4.7	36	40	2	0	NA
NonRA	PsA	41	M	n	n	4	2	25	50	4.8	NA	37	1	0	NA
sdRA	RA	45	F	n	n	3	3	12	24	3.8	20	0	2	0	NA
sdRA	RA	63	F	n	n	5	1	9	104	5.1	48	56	1	0	NA
sdRA	RA	48	F	n	n	6	8	102	4	3.5	NA	96	2	2	NA
ldRA	RA	46	M	p	p	16	13	7	34	6.7	NA	42	2	1	NA
ldRA	RA	61	F	n	n	6	15	9	8	4.9	53	16	1	1	NA
ldRA	RA	69	F	n	n	7	7	0	11	4.6	17	73	2	1	NA
sdRA	RA	59	M	n	n	20	4	22	14	5.0	65	78	3	2	NA
ldRA	RA	57	M	p	p	14	21	16	56	7.1	NA	60	1	2	NA
ldRA	RA	58	M	p	p	7	7	0	7	4.0	NA	2	3	2	NA
NonRA	PsA	55	F	n	p	2	0	16	19	3.9	6	23	1	1	NA
Res	ReA	32	M	n	n	1	1	10	10	2.9	24	100	2	2	NA
Res	UA	33	M	n	n	9	12	14	51	6.7	64	32	1	1	NA
Res	RA	74	M	n	n	23	0	13	45	4.8	35	0	1	0	NA
Res	UA	72	M	n	n	4	7	0	5	3.6	18	0	1	1	NA

sdRA	RA	43	M	n	n	4	19	0	58	6.9	17	61	1	0	NA
ldRA	RA	46	F	p	n	4	10	32	24	5.2	NA	67	2	2	NA
Res	Pseudogout	81	F	n	n	11	16	52	60	6.7	NA	0	2	2	NA
Res	Parvovirus	45	F	n	n	5	5	0	4	4.0	9	82	1	1	NA
sdRA	RA	55	M	p	n	4	0	45	58	3.5	80	79	2	0	NA
Res	ReA	35	M	n	n	1	1	7	51	4.1	21	64	1	1	NA
sdRA	RA	48	F	p	p	3	6	0	10	3.9	NA	0	2	0	NA
NonRA	PsA	43	F	p	n	10	11	70	97	7.1	59	60	2	1	NA
Res	UA	55	M	n	n	5	4	6	2	3.5	31	2	2	1	NA
NonRA	Sarcoid	39	F	n	n	0	5	15	27	4.7	9	28	1	0	NA
sdRA	RA	53	F	n	n	2	7	0	11	4.2	10	34	1	0	NA
NonRA	UA	69	M	n	n	3	5	38	44	5.6	62	74	1	1	NA
sdRA	RA	74	F	p	n	3	3	32	20	4.4	46	100	3	2	NA
Res	UA	35	M	n	n	3	1	9	2	1.6	22	12	2	1	NA
Res	UA	37	F	n	n	2	8	0	7	4.7	12	100	1	0	NA
ldRA	RA	67	M	p	p	1	1	48	29	4.2	23	66	3	2	NA
NonRA	AS	34	M	n	n	1	1	22	21	3.5	17	50	3	2	NA
Res	Parvovirus	41	F	n	n	2	0	9	5	1.7	13	0	2	0	NA
Res	UA	32	F	n	n	1	3	0	15	3.6	NA	7	1	0	NA
sdRA	RA	58	F	p	p	9	5	0	27	4.5	18	7	1	0	NA
NonRA	SLE	33	F	n	n	17	11	0	79	7.4	35	56	1	0	NA
ldRA	RA	22	F	n	n	6	6	79	81	6.4	46	90	3	2	NA
Res	ReA	28	M	n	n	1	2	8	18	4.5	NA	59	2	1	NA
ldRA	RA	72	F	n	n	16	21	43	53	7.4	85	50	2	3	NA
Res	ReA	27	M	n	n	2	2	28	37	3.8	NA	77	1	0	NA
ldRA	RA	65	M	p	p	12	3	81	72	5.3	61	78	3	3	NA
ldRA	RA	51	F	p	p	4	2	0	9	3.3	52	20	1	0	NA
NonRA	PMR	58	F	n	n	2	3	13	24	4.7	23	41	1	2	NA

ldRA	RA	65	F	n	n	6	4	10	50	5.0	37	33	2	1	NA
sdRA	RA	51	M	p	p	10	11	97	61	6.7	92	100	2	3	NA
sdRA	RA	63	M	n	n	1	1	0	2	2.3	NA	81	3	2	NA
ldRA	RA	63	F	p	p	3	6	0	35	5.2	78	30	1	2	NA
ldRA	RA	64	F	p	p	11	12	10	46	6.1	65	50	3	3	NA
sdRA	RA	61	M	n	n	16	14	25	20	5.3	86	75	3	3	NA
sdRA	RA	56	M	p	n	6	25	25	29	6.7	58	69	1	1	Yes
ldRA	RA	56	F	p	p	1	7	47	48	5.5	NA	79	2	1	NA
ldRA	RA	60	F	n	p	10	5	112	70	5.2	69	51	3	0	No
sdRA	RA	61	F	n	p	1	15	18	43	5.4	24	12	1	0	NA
NonRA	Peripheral SpA	24	M	n	p	1	1	122	44	4.3	63	66	2	2	NA
NonRA	MCTD	23	F	n	p	4	7	16	6	3.9	48	40	2	0	NA
NonRA	PsA	60	M	n	n	3	0	5	10	2.5	28	13	2	2	NA
ldRA	RA	77	F	n	n	4	11	5	4	4.2	62	59	2	3	No
sdRA	RA	59	M	p	p	11	22	8	26	6.5	79	40	3	2	NA
ldRA	RA	52	F	p	p	7	5	13	34	5.6	72	93	3	0	NA
ldRA	RA	33	F	p	p	6	8	15	48	5.9	56	69	3	1	Yes
ldRA	RA	51	M	p	p	6	15	113	96	7.3	28	59	2	0	Yes
ldRA	RA	29	F	p	p	6	24	49	60	6.8	NA	NA	NA	NA	NA
Res	UA	45	F	n	n	2	1	17	8	2.6	NA	NA	NA	NA	NA
NonRA	UA	57	F	n	n	6	8	76	90	6.6	NA	NA	NA	NA	NA
NonRA	PsA	39	F	n	n	1	4	3	27	4.0	39	NA	NA	NA	NA
sdRA	RA	55	M	p	p	14	13	6	8	5.5	NA	70	3	0	NA
ldRA	RA	54	F	n	n	6	22	0	6	5.3	NA	NA	NA	NA	NA
ldRA	RA	44	M	p	p	1	10	0	2	3.9	20	69	2	0	Yes
sdRA	RA	53	M	n	n	12	23	88	58	7.5	92	40	2	0	Yes
ldRA	RA	53	M	p	p	5	7	21	56	5.8	67	63	3	0	NA
ldRA	RA	63	M	n	n	10	28	28	27	7.5	61	53	2	0	Yes

ldRA	RA	62	F	p	p	7	8	8	39	6.2	45	50	2	0	Yes
ldRA	RA	76	M	p	p	5	2	11	53	4.2	NA	NA	NA	NA	NA
ldRA	RA	53	F	n	n	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ldRA	RA	55	F	p	n	3	15	3	12	5.1	NA	NA	NA	NA	NA
ldRA	RA	69	F	n	n	14	15	66	45	7.1	79	73	2	1	Yes
ldRA	RA	55	F	n	p	2	2	13	31	4.8	NA	NA	NA	NA	NA
ldRA	RA	60	M	p	p	1	3	5	24	3.9	33	50	2	3	No
Res	UA	37	M	n	n	0	0	18	37	3.2	7	56	1	0	NA
ldRA	RA	77	M	p	p	5	2	11	53	4.2	23	0	2	2	NA
ldRA	RA	39	M	p	n	15	26	58	9	6.9	70	87	1	0	NA
ldRA	RA	55	M	p	p	4	7	12	39	5.4	NA	NA	NA	NA	NA
ldRA	RA	24	F	p	p	3	20	1	15	NA	NA	NA	1	0	NA
sdRA	RA	76	F	n	n	20	19	39	39	7.0	51	35	1	0	NA
ldRA	RA	74	F	n	n	24	26	25	45	8.1	58	80	3	1	NA
sdRA	RA	54	F	p	p	3	7	0	21	4.5	NA	NA	NA	NA	NA
ldRA	RA	69	M	n	n	22	25	7	2	5.1	74	83	2	2	NA
ldRA	RA	57	M	p	p	4	9	10	21	5.4	45	68	3	1	NA
ldRA	RA	52	F	n	n	5	23	9	46	6.8	86	86	2	0	No
Res	UA	55	F	n	n	3	3	4	14	4.3	19	28	2	1	NA
sdRA	RA	70	M	p	p	16	19	29	56	7.3	NA	NA	NA	NA	NA
ldRA	RA	51	M	p	p	9	12	59	21	5.8	84	77	3	1	NA
ldRA	RA	46	F	n	n	4	14	10	13	5.6	NA	NA	NA	NA	NA
ldRA	RA	58	F	p	p	0	1	13	31	4.2	NA	NA	3	2	NA
ldRA	RA	62	F	p	p	1	8	6	63	5.3	62	48	3	3	NA
ldRA	RA	69	F	n	n	5	5	148	51	5.5	69	95	1	1	NA
ldRA	RA	62	M	n	p	17	18	4	12	5.7	74	48	2	2	Yes
ldRA	RA	49	M	p	n	7	7	13	14	5.4	49	47	3	1	Yes
sdRA	RA	71	M	n	n	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

ldRA	RA	60	F	n	n	15	14	11	4	5.0	4	7	2	2	NA
NonRA	PsA	64	M	n	n	2	1	11	2	1.6	11	25	2	3	No
ldRA	RA	18	F	p	p	11	12	29	35	6.1	64	71	1	0	NA
sdRA	RA	59	F	n	p	11	12	36	10	5.2	88	90	3	3	Yes
ldRA	RA	45	M	p	p	4	5	5	na	NA	50	50	1	0	NA
NonRA	PsA	41	M	n	n	16	0	25	18	4.2	75	0	2	0	NA
ldRA	RA	69	M	n	n	14	19	78	83	7.9	7	94	2	2	NA
ldRA	RA	61	M	p	n	3	18	0	38	6.0	28	60	1	0	NA
sdRA	RA	74	M	p	p	10	4	33	25	4.4	56	41	3	2	NA
ldRA	RA	30	F	p	p	21	28	57	48	8.1	84	93	2	0	Yes
sdRA	RA	68	F	n	p	7	18	5	34	7.0	NA	NA	NA	NA	NA
Res	Parvovirus	60	M	p	n	1	1	21	22	3.1	NA	NA	NA	NA	NA
ldRA	RA	74	M	p	p	3	10	44	47	5.9	49	71	2	2	Yes
ldRA	RA	68	F	p	n	5	4	6	38	4.8	NA	NA	NA	NA	NA
Res	PsA	56	F	n	n	6	4	3	9	4.7	21	22	2	1	NA
ldRA	RA	27	F	n	n	8	12	29	35	6.5	NA	NA	NA	NA	NA
ldRA	RA	50	F	p	p	2	26	2	2	4.6	NA	NA	NA	NA	NA
sdRA	RA		F	p	p	4	11	1	48	6.5	NA	NA	NA	NA	NA
ldRA	RA	46	M	p	p	2	2	22	22	4.8	NA	NA	NA	NA	NA
ldRA	RA	66	F	p	n	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ldRA	RA	73	F	p	p	3	1	34	28	4.6	54	NA	NA	NA	NA
ldRA	RA	50	F	n	n	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ldRA	RA	58	M	p	p	15	13	135	37	6.2	76	NA	NA	NA	NA
sdRA	RA	42	F	p	p	8	8	1	31	6.2	45	NA	NA	NA	NA
ldRA	RA	81	M	n	p	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 9.2: Patient characteristics for the AMP RA Phase II cohort. Samples from this cohort were used for external validation of BEACON scoring in Chapter 3. SJC28 = Swollen Joint Count 28, TJC28 = Tender Joint Count 28, CRP = C Reactive Protein, ESR = Erythrocyte Sedimentation Rate, DAS28-ESR = Disease Activity Score 28, VAS = Visual Analog Scale, US GS = Ultrasound Greyscale, US PD = Ultrasound Power Doppler, DAS28-ESR response = EULAR DAS28-ESR response, response includes good and moderate response, NA = data not available.

Age	Sex	SJC28	TJC28	CRP	ESR	DAS28-ESR	Physician global	Patient VAS	US GS	US PD	DAS28-ESR response 6 months
24	F	21	18	1.4	19	5.8	3	5.5	3	NA	Y
73	F	22	25	1.2	65	7.1	6	4.5	2	1	Y
60	F	22	10	1.0	NA	NA	5	3.5	3	1	NA
67	F	19	12	0.6	NA	NA	6	5.5	2	2	NA
39	F	10	14	0.4	24	5.3	7	9	2	2	Y
71	M	6	6	1.4	50	4.9	4	5	3	1	N
67	M	20	27	3.2	64	7.2	5	7	2	1	Y
28	F	2	9	0.8	25	4.4	5	3.5	NA	NA	Y
50	F	18	12	0.2	105	6.5	8	5.5	2	2	NA
56	F	15	13	5.1	25	5.4	8	5	2	1	N
61	F	2	6	0.0	27	4.2	5	6.5	3	0	N
72	F	12	0	0.2	8	2.5	8	3	NA	NA	Y
59	F	5	7	0.2	5	3.3	7	4	NA	NA	NA
66	F	2	2	2.1	32	3.7	3.5	3	NA	NA	NA
69	F	2	2	1.4	42	3.8	3	2	NA	NA	NA
67	M	13	18	2.1	22	5.7	7	7.5	3	3	NA
67	M	6	3	0.3	9	3.3	7	10	2	0	NA
51	F	4	1	0.7	38	3.8	5	8	3	0	NA
63	M	17	8	7.8	63	5.7	8	5	3	2	Y
50	F	6	8	0.1	43	5.0	5	7	2	1	Y
45	M	2	2	1.4	NA	NA	2	3	2	0	NA
55	M	5	8	0.6	24	4.5	5	4.5	2	2	N

70	M	10	9	3.7	39	5.3	5	9	2	1	Y
56	F	9	12	0.8	59	5.8	6	9	2	2	NA
41	M	2	2	0.7	13	3.0	7	4	2	1	N
35	F	5	4	4.5	31	4.2	8.5	6.5	2	3	NA
63	M	12	24	2.3	87	7.0	8.5	8	3	3	NA
70	F	7	12	4.3	39	5.3	6.5	5	2	1	NA
51	F	3	6	1.5	7	3.2	2	2	2	2	Y
36	M	24	24	1.7	80	7.3	8	6	2	2	NA
64	F	10	6	1.2	15	4.2	3	3	2	0	NA
77	F	1	1	1.3	25	3.2	6	7	3	3	NA
59	M	3	1	0.5	28	3.4	3.5	2	2	2	N
55	M	3	2	0.4	6	2.6	5	6	3	3	N
75	F	2	0	0.7	82	3.6	4	5	2	2	Y
68	F	13	10	0.8	4	3.8	7	5.5	2	1	Y
62	F	5	8	0.2	3	3.1	6	5.5	3	2	N
79	F	13	13	0.8	12	4.9	6	6.5	2	1	N
68	F	7	2	1.2	30	3.9	6	1	3	2	Y
61	F	7	3	1.3	28	4.1	2	4	1	1	NA
57	F	9	11	NA	NA	NA	5.2	3	3	3	NA
43	F	18	28	2.4	25	6.5	8	10	3	3	NA
52	F	6	7	NA	NA	NA	5.5	8	2	2	NA
27	F	7	12	NA	NA	NA	4.6	2	3	2	NA
65	F	14	16	0.3	36	5.9	7	6	2	1	N
60	F	11	5	NA	NA	NA	8.4	9	3	2	NA
64	F	16	6	NA	NA	NA	7.2	8	3	3	NA
54	F	10	2	0.2	43	4.4	5.1	5	3	2	Y
55	M	19	27	NA	33	6.6	6.8	5	2	1	Y
23	F	11	0	0.0	24	3.2	3	2	2	0	NA

66	F	6	21	NA	NA	NA	6	7	0	0	NA
68	F	2	2	0.0	14	3.0	4	0.5	2	1	Y
24	F	16	13	1.9	36	5.7	7	5	3	2	NA
69	M	14	0	0.7	8	2.6	3	5	3	0	NA
46	F	14	16	0.7	14	5.2	4	5	2	0	NA
42	F	4	4	0.0	8	3.2	2.5	6	1	0	NA
58	F	5	11	0.5	5	3.7	7.5	5.5	2	2	NA
34	F	NA	NA	0.0	8	NA	2	6	2	0	NA
51	F	7	14	0.8	3	3.7	7.8	7	3	2	N
77	M	16	21	17.4	34	6.2	9	1.5	2	2	Y
66	F	14	21	9.8	32	6.1	6.6	6	3	2	Y
69	F	6	2	0.6	19	3.6	2.2	4	2	2	Y
56	F	2	2	0.0	24	3.5	5.2	5	2	2	N
45	M	4	4	0.5	NA	NA	5	0	1	0	NA
39	F	9	16	0.0	23	5.3	6.9	4	3	2	NA
54	F	2	3	0.0	28	3.8	5	6.5	2	2	Y
61	M	3	18	0.0	38	5.5	2.8	7	1	0	Y
30	F	21	28	5.7	48	7.1	8.4	7.5	2	0	N
68	F	7	18	0.5	34	5.7	6.1	6	1	0	N
62	M	5	14	0.6	52	NA	9	NA	3	1	NA
44	F	16	24	1.0	12	NA	7	NA	2	0	NA
36	F	3	6	1.6	107	NA	4.9	NA	2	0	NA
75	F	5	9	0.8	31	NA	4.3	NA	3	0	NA
39	F	7	14	2.2	26	NA	7.7	NA	3	2	NA
74	M	3	10	4.4	47	5.0	4.9	5	2	2	NA
27	F	8	12	2.9	35	5.4	7.4	9.5	1	0	NA
30	F	21	28	0.6	19	NA	8.4	NA	2	1	NA
73	F	3	1	3.4	28	3.5	5.4	8.5	2	2	NA

38	M	3	4	1.4	84	4.8	8.8	9	2	2	Y
70	M	2	5	2.4	30	4.1	3	5	2	2	Y
55	F	11	22	0.2	43	6.3	3.1	5	3	2	Y
68	F	15	1	3.5	21	3.9	5.8	6.5	2	0	Y
30	F	3	10	1.4	36	NA	5.3	NA	2	1	NA
39	M	5	17	1.6	5	NA	7.4	NA	2	2	NA
76	F	13	7	0.3	40	5.1	6	5	3	3	NA
44	F	13	11	4.2	NA	NA	7	7	2	2	NA
51	F	13	17	NA	NA	NA	6	7	2	2	NA
82	M	8	5	8.9	NA	NA	1.5	6	2	1	NA
58	F	10	6	0.7	89	5.5	6	7	2	0	NA
75	F	15	10	0.6	29	5.3	6	6	2	1	NA
49	F	13	13	3.9	88	6.3	6	10	2	1	NA
75	F	22	19	5.7	109	7.1	7	7	2	0	Y
68	M	10	8	0.9	44	5.2	6	7	2	2	NA
32	F	21	21	0.0	4	4.9	5	6.5	2	0	NA
63	M	28	28	14.3	130	8.0	9	7	3	1	Y
47	F	16	16	0.1	12	5.2	6	5	2	0	NA
68	F	22	22	0.5	80	7.1	7	8.5	2	0	NA
44	F	23	3	0.8	42	5.0	6	8	2	1	NA
44	F	22	22	0.4	21	6.2	5	6	2	1	NA
81	F	24	10	0.8	86	6.4	7	8	2	1	NA
52	F	26	26	1.0	32	6.8	6	5	2	0	NA
42	F	3	7	1.1	45	4.7	8	8	2	0	NA
61	M	14	0	1.1	77	4.2	7	6.5	2	1	NA

Table 9.3: Patient characteristics for BEACON cohort samples used in bulk RNA sequencing. Samples from these patients were used in Chapters 4, 5 and 6. ACPA = Anti-citrullinated protein antibodies, RF = Rheumatoid Factor, TJC28 = Tender Joint Count 28, SJC28 = Swollen Joint Count 28, CRP = C Reactive Protein, ESR = Erythrocyte Sedimentation Rate, DAS28-ESR = Disease Activity Score 28, US GS = Ultrasound Greyscale, US PD = Ultrasound Power Doppler, DAS28-ESR response = EULAR DAS28 response, response includes good and moderate response, NA = data not available.

Clinical group	Final classification	Age	Sex	ACPA	RF	TJC28	SJC28	CRP	ESR	DAS28-ESR	US GS	US PD	DAS28-ESR response 12 months
Res	UA	64	M	n	n	5	2	15	24	4.5	1	0	NA
sdRA	RA	49	F	n	n	9	8	8	12	4.7	2	0	NA
sdRA	RA	45	F	n	n	3	3	12	24	3.8	2	0	NA
sdRA	RA	63	F	n	n	1	5	9	104	5.1	1	0	NA
ldRA	RA	61	F	n	n	15	6	9	8	4.9	1	1	NA
ldRA	RA	69	F	n	n	7	7	0	11	4.6	2	1	NA
sdRA	RA	59	M	n	n	4	20	22	14	5.0	3	2	NA
NonRA	PsA	55	F	n	p	0	2	16	19	3.1	1	1	NA
Res	ReA	32	M	n	n	1	1	10	10	2.9	2	2	NA
Res	UA	33	M	n	n	12	9	14	51	6.7	1	1	NA
Res	RA	74	M	n	n	0	23	13	45	4.8	1	0	NA
Res	UA	72	M	n	n	7	4	0	5	3.6	1	1	NA
sdRA	RA	43	M	n	n	19	4	0	58	6.9	1	0	NA
Res	Parvovirus	45	F	n	n	5	5	0	4	4.0	1	1	NA
sdRA	RA	55	M	p	n	0	4	45	58	3.5	2	0	NA
Res	ReA	35	M	n	n	1	1	7	51	4.1	1	1	NA
NonRA	PsA	43	F	p	n	11	10	70	97	7.1	2	1	NA
Res	UA	55	M	n	n	4	5	6	2	3.5	2	1	NA
NonRA	Sarcoid	39	F	n	n	5	0	15	27	4.7	1	0	NA
sdRA	RA	53	F	n	n	7	2	0	11	4.2	1	0	NA
NonRA	UA	69	M	n	n	5	3	38	44	5.6	1	1	NA
sdRA	RA	74	F	p	n	3	3	32	20	4.4	3	2	NA

Res	UA	35	M	n	n	1	3	9	2	1.6	2	1	NA
Res	UA	37	F	n	n	8	2	0	7	4.7	1	0	NA
ldRA	RA	67	M	p	p	1	1	48	29	4.2	3	2	NA
NonRA	AS	34	M	n	n	1	1	22	21	3.5	3	2	NA
Res	Parvovirus	41	F	n	n	0	2	9	5	1.7	2	0	NA
sdRA	RA	58	F	p	p	5	9	0	27	4.5	1	0	NA
NonRA	SLE	33	F	n	n	11	17	0	79	7.4	1	0	NA
ldRA	RA	22	F	n	n	6	6	79	81	6.4	3	2	NA
ldRA	RA	72	F	n	n	21	16	43	53	7.4	2	3	NA
ldRA	RA	65	M	p	p	3	12	81	72	5.3	3	3	NA
Norm	Norm	53	F	n	n	NA	NA	NA	NA	NA	NA	NA	NA
Norm	Norm	44	M	n	n	NA	NA	NA	NA	NA	NA	NA	NA
Norm	Norm	48	M	n	n	NA	NA	NA	NA	NA	NA	NA	NA
ldRA	RA	51	F	p	p	2	4	0	9	3.3	1	0	NA
Norm	Norm	22	F	n	n	NA	NA	NA	NA	NA	NA	NA	NA
Norm	Norm	24	F	n	n	NA	NA	NA	NA	NA	NA	NA	NA
NonRA	PMR	58	F	n	n	3	2	13	24	4.7	1	2	NA
ldRA	RA	65	F	n	n	4	6	10	50	5.0	2	1	NA
sdRA	RA	51	M	p	p	11	10	97	61	6.7	2	3	NA
ldRA	RA	63	F	p	p	6	3	0	35	5.2	1	2	NA
ldRA	RA	64	F	p	p	12	11	10	46	6.1	3	3	NA
sdRA	RA	61	M	n	n	14	16	25	20	NA	3	3	NA
sdRA	RA	56	M	p	n	25	6	25	29	6.7	1	1	Yes
ldRA	RA	60	F	n	p	5	10	112	70	5.2	3	0	No
sdRA	RA	61	F	n	p	15	1	18	43	5.4	1	0	NA
NonRA	Peripheral SpA	24	M	n	p	1	1	122	44	4.3	2	2	NA
NonRA	MCTD	23	F	n	p	7	4	16	6	3.9	2	0	NA
NonRA	PsA	60	M	n	n	0	3	5	10	2.5	2	2	NA

ldRA	RA	77	F	n	n	11	4	5	4	4.2	2	3	No
sdRA	RA	59	M	p	p	22	11	8	26	6.5	3	2	NA
ldRA	RA	52	F	p	p	5	7	13	34	5.6	3	0	NA
ldRA	RA	33	F	p	p	8	6	15	48	5.9	3	1	Yes
ldRA	RA	46	M	n	p	19	5	4	12	5.6	2	0	NA
ldRA	RA	51	M	p	p	15	6	113	96	7.3	2	0	Yes
NonRA	PsA	39	F	n	n	4	1	3	27	4.0	NA	NA	NA
ldRA	RA	44	M	p	p	10	1	0	2	3.9	2	0	Yes
sdRA	RA	53	M	n	n	23	12	88	58	7.5	2	0	Yes
ldRA	RA	53	M	p	p	7	5	21	56	5.8	3	0	NA
ldRA	RA	63	M	n	n	28	10	28	27	7.5	2	0	Yes
ldRA	RA	62	F	p	p	8	7	8	39	6.2	2	0	Yes
ldRA	RA	69	F	n	n	15	14	66	45	7.1	2	1	Yes
ldRA	RA	60	M	p	p	3	1	5	24	3.9	2	3	No
Res	UA	37	M	n	n	0	0	18	37	3.2	1	0	NA
ldRA	RA	77	M	p	p	2	5	11	53	4.2	2	2	NA
sdRA	RA	76	F	n	n	19	20	39	39	7.0	1	0	NA
ldRA	RA	74	F	n	n	26	24	25	45	8.1	3	1	NA
ldRA	RA	52	F	n	n	23	5	9	46	6.8	2	0	No
Res	UA	55	F	n	n	3	3	4	14	4.3	2	1	NA
ldRA	RA	62	F	p	p	8	1	6	63	5.3	3	3	NA
ldRA	RA	69	F	n	n	5	5	148	51	5.5	1	1	NA
ldRA	RA	62	M	n	p	18	17	4	12	5.7	2	2	Yes
ldRA	RA	49	M	p	n	7	7	13	14	5.4	3	1	Yes
ldRA	RA	60	F	n	n	14	15	11	4	5.0	2	2	NA
NonRA	PsA	64	M	n	n	1	2	11	2	1.6	2	3	No
sdRA	RA	59	F	n	p	12	11	36	10	5.2	3	3	Yes
NonRA	PsA	41	M	n	n	0	16	25	18	4.2	2	0	NA

ldRA	RA	52	M	P	P	12	0	1	8	4.1	2	0	No
ldRA	RA	69	M	n	n	19	14	78	83	7.9	2	2	NA
ldRA	RA	61	M	p	n	18	3	0	38	6.0	1	0	NA
sdRA	RA	74	M	p	p	4	10	33	25	4.4	3	2	NA
ldRA	RA	30	F	p	p	28	21	57	48	8.1	2	0	Yes
ldRA	RA	74	M	p	p	10	3	44	47	5.9	2	2	Yes
Res	PsA	56	F	n	n	2	6	3	9	4.3	2	1	NA

9.2 BEACON SCORING SYSTEM ATLAS

Birmingham Density and Aggregate Grading Atlas

UNIVERSITY OF BIRMINGHAM

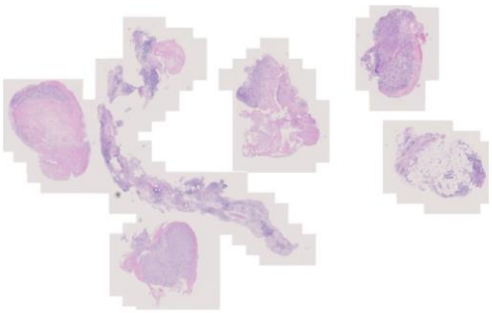
v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A

BEACON
BIRMINGHAM EARLY ARTHRITIS COHORT

2

Reporting Tissue samples

- Synovial biopsy tissue samples are made up of multiple fragments to account for tissue heterogeneity. 6-8 fragments in large joints, at least 4 fragments for small joints.
- Only fragments agreed to contain SYNOVIAL tissue should be graded.
- If a large single piece of tissue (for instance from arthroplasty or synovectomy) is being graded then the scorer is expected to examine multiple discrete areas, at least 6 for a large joint and at least 4 for a small joint.



UNIVERSITY OF BIRMINGHAM

v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A

BEACON

3

Assignment of pathotype

Pathotypes are assigned as a hierarchy with Lymphoid highest:

Lymphoid: presence of significant aggregates:

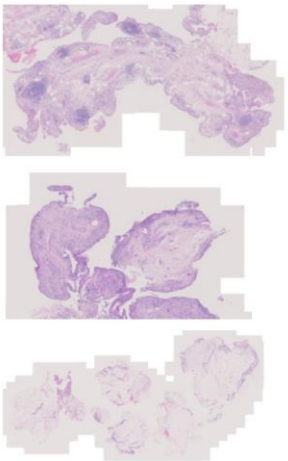
- Presence of ≥1 grade 1 aggregate *in at least two fragments*, or any grade 2 aggregate, or any grade 3 aggregate

Diffuse: presence of immune infiltrate but absence of significant aggregates:

- Does not meet lymphoid criteria, mean fragment density grade ≥1

Pauci-immune: Absence of immune infiltrate

- Does not meet lymphoid criteria, mean density grade <1



UNIVERSITY OF BIRMINGHAM

v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A

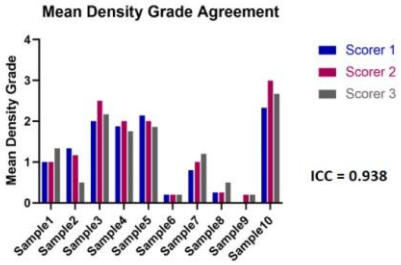
BEACON

Birmingham Density Grading

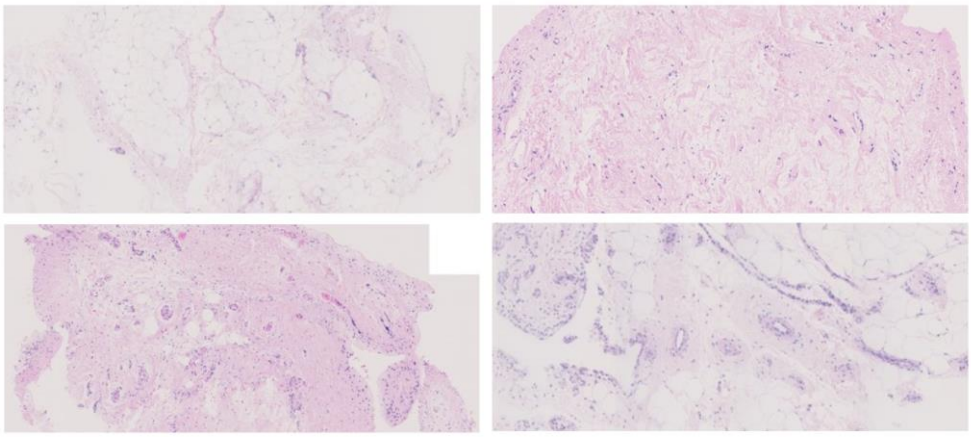
- Grade fragments individually

- Grade 0 no infiltrate
- Grade 1 low density (scarce) infiltrate
- Grade 2 medium density infiltrate
- Grade 3 high density or band-like infiltrate

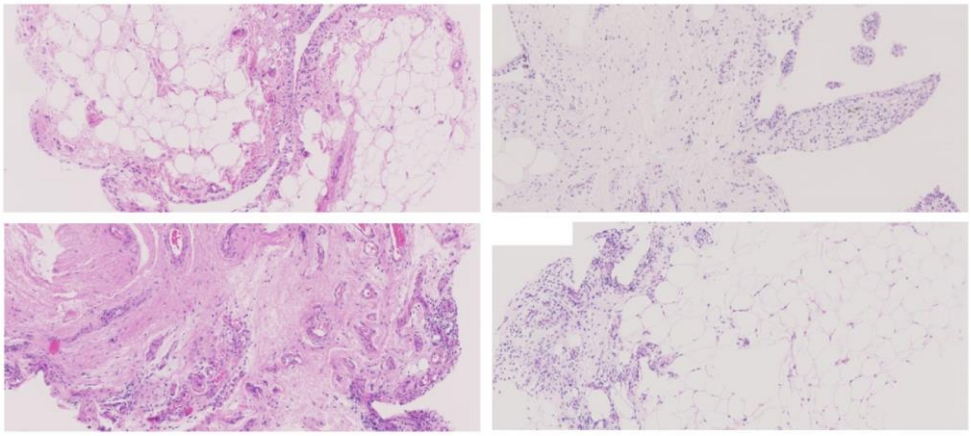
- Whole tissue sample reported as mean of fragment grades
- Note only fragments containing synovial tissue should be graded



Grade 0 Density

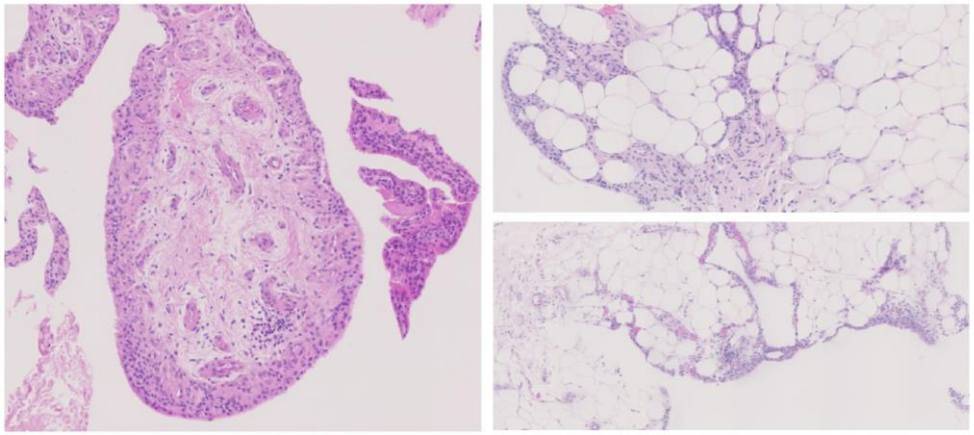


Grade 1 Density



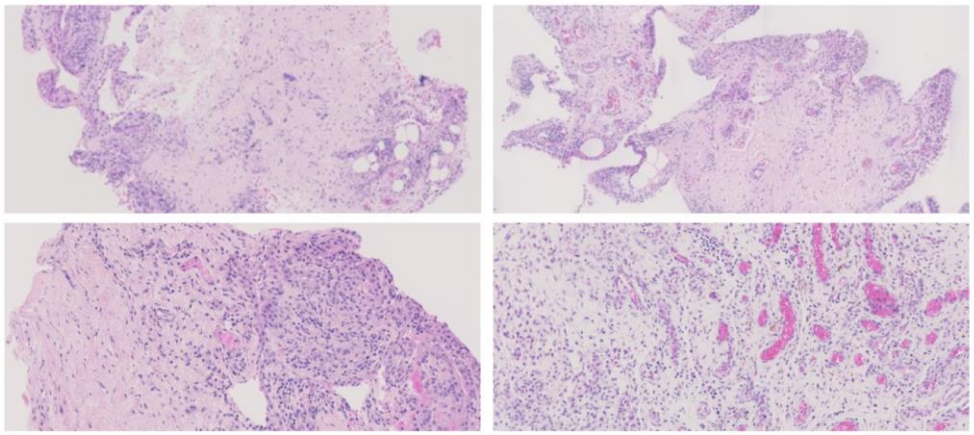
Grade 1 Density

7



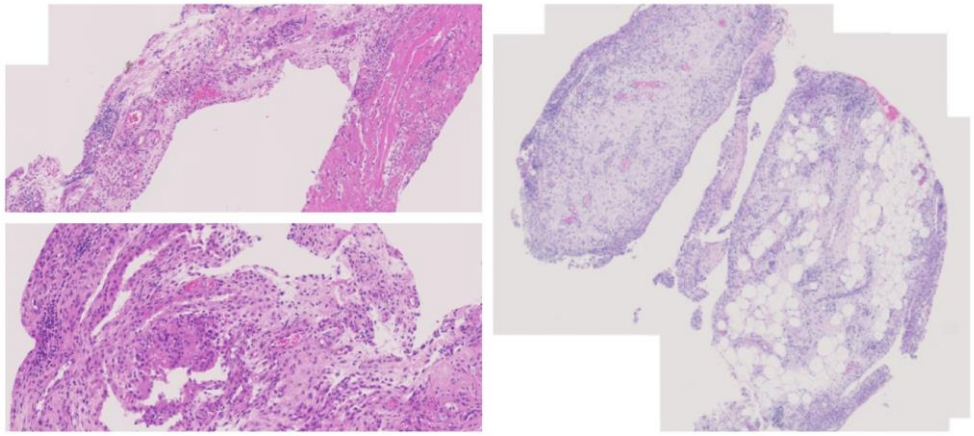
Grade 2 Density

8



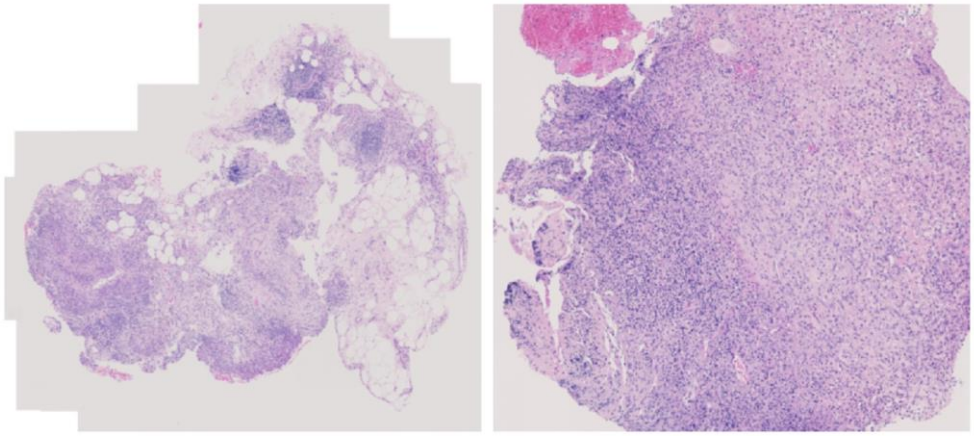
Grade 2 Density

9



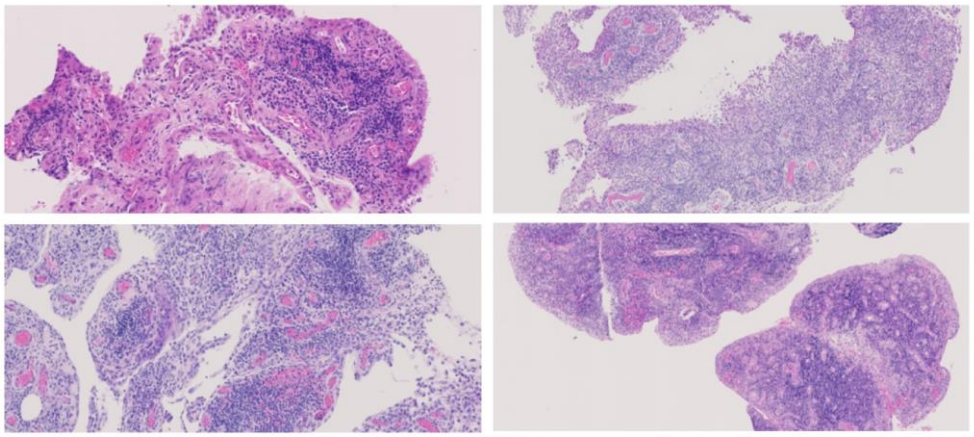
10

Grade 3 Density



11

Grade 3 Density

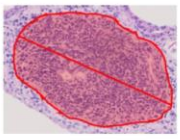


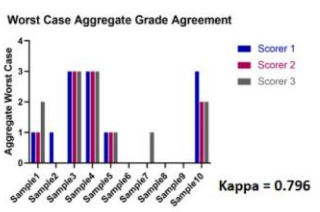
12

Birmingham Aggregate Grading

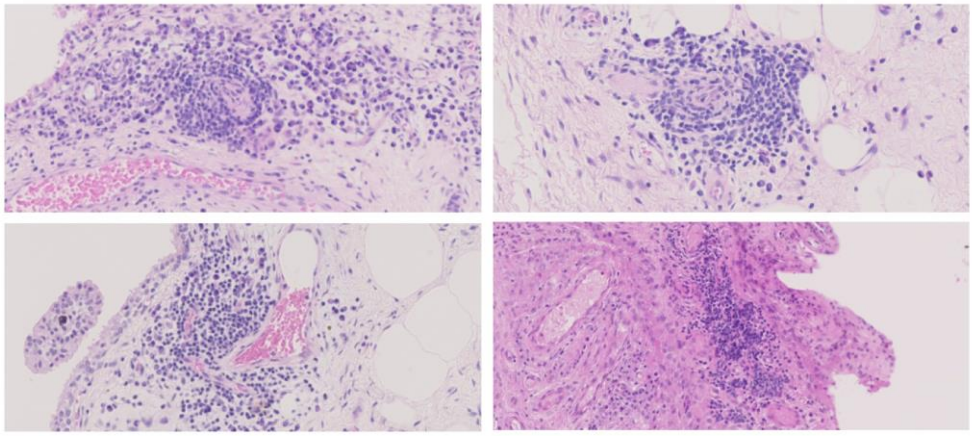
- Grade fragments individually
- Both lymphocyte and plasma cell aggregates allowed
- Grade maximum longitudinal diameter of structure

- Grade 0 no aggregates
- Grade 1 low 6-9 radial cell count
- Grade 2 medium 10-19 radial cell count
- Grade 3 high ≥ 20 radial cell count
- Whole tissue sample reported as worst case of any fragment, but to achieve grade 1, aggregates must be present in at least two fragments
- Note only fragments containing synovial tissue should be graded



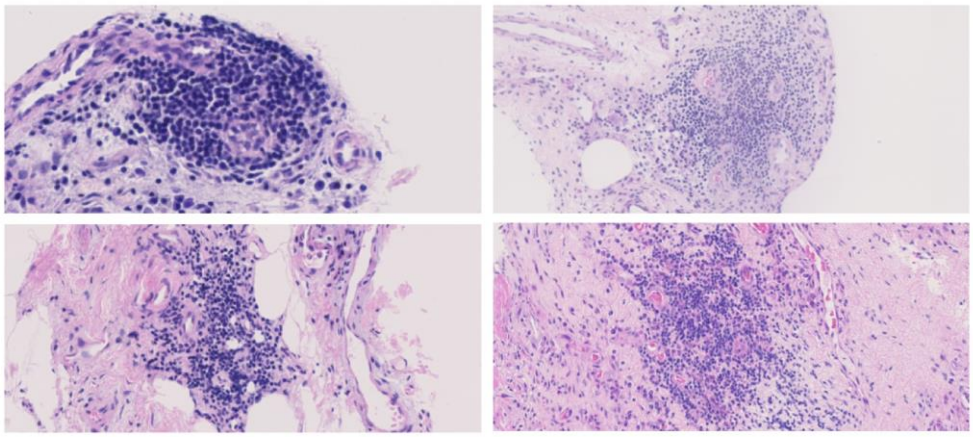


Grade 1 Aggregate



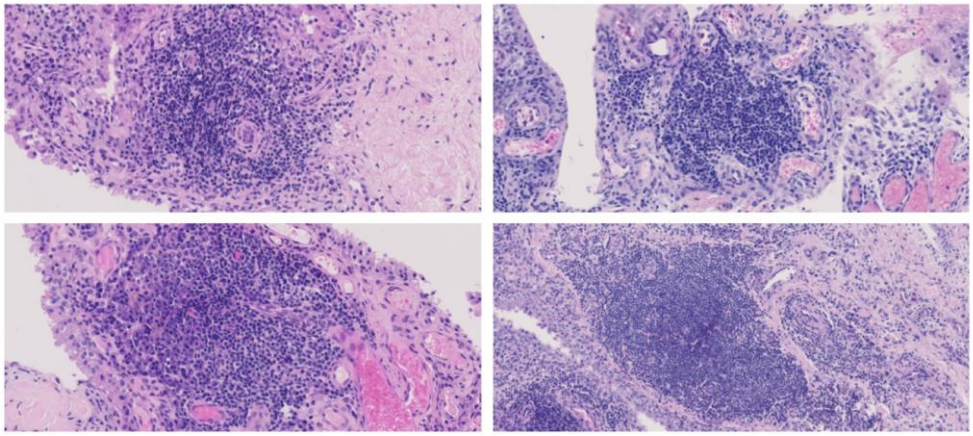
UNIVERSITY OF BIRMINGHAM v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A BEACON

Grade 2 Aggregate



UNIVERSITY OF BIRMINGHAM v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A BEACON

Grade 3 Aggregate



UNIVERSITY OF BIRMINGHAM v1.1 18.02.2019 Carr, H; Scheel-Toellner, D; Filer, A BEACON

9.3 SCRIPT FOR DATA PREPARATION, ALIGNMENT, AND READ SUMMARISATION

```
#!/bin/sh

# Description
flowcell=GB120918-AF
sampledirs=( BX006 BX008 BX010 BX011 BX012 BX013 BX014 BX017 BX018 BX020 BX023 BX024 BX026
BX028 BX030 BX031 BX038 BX042 BX045 BX048 \
    BX051 BX054 BX056 BX057 BX062 BX063 BX064 BX065 BX066 BX067 BX071 BX073 BX074 BX075 BX077
BX093 BX101 BX102 BX103 BX104 \
    BX106 BX107 BX110 BX112 BX113 BX115 BX117 BX118 BX119 BX121 BX123 BX127 BX130 BX131 BX135
BX138 BX139 BX141 BX150 BX151 \
    BX152 BX153 BX160 BX164 BX165 BX166 BX167 BX168 BX173 BX175 BX178 BX188 BX194 BX195 BX201
BX202 BX217 BX221 BX230 BX240 \
    BX242 BX245 BX248 BX272 BX275 BX278 BX281 BX285 BX287-p1 BX287-p2 BX287 p3 BX287 p4
BX287_p6 BX296 BX303 )

samples=( BX006 BX008 BX010 BX011 BX012 BX013 BX014 BX017 BX018 BX020 BX023 BX024 BX026 BX028
BX030 BX031 BX038 BX042 BX045 BX048 \
    BX051 BX054 BX056 BX057 BX062 BX063 BX064 BX065 BX066 BX067 BX071 BX073 BX074 BX075 BX077
BX093 BX101 BX102 BX103 BX104 \
    BX106 BX107 BX110 BX112 BX113 BX115 BX117 BX118 BX119 BX121 BX123 BX127 BX130 BX131 BX135
BX138 BX139 BX141 BX150 BX151 \
    BX152 BX153 BX160 BX164 BX165 BX166 BX167 BX168 BX173 BX175 BX178 BX188 BX194 BX195 BX201
BX202 BX217 BX221 BX230 BX240 \
    BX242 BX245 BX248 BX272 BX275 BX278 BX281 BX285 BX287-p1 BX287-p2 BX287-p3 BX287-p4 BX287-
p6 BX296 BX303 )
nsamples_minus_1=94

STARindexdir="/castles/nr/projects/f/filera-rheumatoid-arthritis-rna-
seq/sequencing/Genomes/hg38/STAR indices/sjdb overhang 100"
assembly=hg38

for i in `seq 0 ${nsamples_minus_1}` ; do

sampledir=${sampledirs[${i}]}
sample=${samples[${i}]}

echo $sampledir $sample

if [ ! -d $sampledir ] ; then mkdir $sampledir ; fi
cd $sampledir
currentdir=$PWD

# link raw data files
ln -s ../../../Raw_Data/${flowcell}/${sampledir}/${sample}*.fastq.gz .

# Merge lanes
if [ ! -s ${sample}_R1.fastq.gz ] ; then
    echo "...merging files for R1..."
    zcat ${sample}*L001_R1_*fastq.gz ${sample}*L002_R1_*fastq.gz ${sample}*L003_R1_*fastq.gz
${sample}*L004_R1_*fastq.gz | gzip -c > ${sample}_R1.fastq.gz
fi
if [ ! -s ${sample}_R2.fastq.gz ] ; then
    echo "...merging files for R2..."
    zcat ${sample}*L001_R2_*fastq.gz ${sample}*L002_R2_*fastq.gz ${sample}*L003_R2_*fastq.gz
${sample}*L004_R2_*fastq.gz | gzip -c > ${sample}_R2.fastq.gz
fi

# Quality control
module purge
module load bluebear
module load FastQC/v0.11.5
if [ ! -s ${sample}_R1_fastqc.html ] ; then fastqc ${sample}_R1.fastq.gz  ; fi
if [ ! -s ${sample}_R2_fastqc.html ] ; then fastqc ${sample}_R2.fastq.gz  ; fi

cd ..

done
# Trim adaptor sequences
```

```
for i in `seq 0 ${nsamples_minus_1}` ; do

sample=${samples[${i}]}

cd $sample
currentdir=$PWD

echo $sample

# Trim adaptors from merged R1 and R2 files
cutadapt -a GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATGTCGTATGC -o ${sample}_Trim_R1.fastq.gz -
p ${sample}_Trim_R2.fastq.gz ${sample}_R1.fastq.gz ${sample}_R2.fastq.gz -m 5 --pair-
filter=any

cd ..

done

#!/bin/bash
#SBATCH -J StarMapping                  # A single job name for the array
#SBATCH -n 8                           # Number of cores
#SBATCH -N 1                           # All cores on one machine
#SBATCH --mem 40000                     # in MB
#SBATCH -t 0-08:00                      # Maximum execution time (D-HH:MM)
#SBATCH -o job %A %a.log        # Standard output
#SBATCH -e job %A %a.log        # Standard error
####SBATCH --qos bbshort
####SBATCH --qos bbdefault
#SBATCH --mail-type ALL         # Send e-mail upon starting/completion/error
##SBATCH --account=filera-rheumatoid-arthritis-rna-seq

# This is a slurm job submission script for STAR alignment

nThreads=8

STARindexdir="/castles/nr/projects/f/filera-rheumatoid-arthritis-rna-
seq/sequencing/Genomes/Human/GRCh38/STAR indices/sjdb overhang 100"

workingdir=$(pwd)

module purge; module load bluebear
module load apps/star-aligner/v2.5.2a
module load apps/samtools/1.4

sampledirs=( BX006 BX008 BX010 BX011 BX012 BX013 BX014 BX017 BX018 BX020 BX023 BX024 BX026
BX028 BX030 BX031 BX038 BX042 BX045 BX048 \
    BX051 BX054 BX056 BX057 BX062 BX063 BX064 BX065 BX066 BX067 BX071 BX073 BX074 BX075 BX077
BX093 BX101 BX102 BX103 BX104 \
    BX106 BX107 BX110 BX112 BX113 BX115 BX117 BX118 BX119 BX121 BX123 BX127 BX130 BX131 BX135
BX138 BX139 BX141 BX150 BX151 \
    BX152 BX153 BX160 BX164 BX165 BX166 BX167 BX168 BX173 BX175 BX178 BX188 BX194 BX195 BX201
BX202 BX217 BX221 BX230 BX240 \
    BX242 BX245 BX248 BX272 BX275 BX278 BX281 BX285 BX287-p1 BX287-p2 BX287-p3 BX287-p4 BX287-
p6 BX296 BX303 )

samples=( BX006 BX008 BX010 BX011 BX012 BX013 BX014 BX017 BX018 BX020 BX023 BX024 BX026 BX028
BX030 BX031 BX038 BX042 BX045 BX048 \
    BX051 BX054 BX056 BX057 BX062 BX063 BX064 BX065 BX066 BX067 BX071 BX073 BX074 BX075 BX077
BX093 BX101 BX102 BX103 BX104 \
    BX106 BX107 BX110 BX112 BX113 BX115 BX117 BX118 BX119 BX121 BX123 BX127 BX130 BX131 BX135
BX138 BX139 BX141 BX150 BX151 \
    BX152 BX153 BX160 BX164 BX165 BX166 BX167 BX168 BX173 BX175 BX178 BX188 BX194 BX195 BX201
BX202 BX217 BX221 BX230 BX240 \
    BX242 BX245 BX248 BX272 BX275 BX278 BX281 BX285 BX287-p1 BX287-p2 BX287-p3 BX287-p4 BX287-
p6 BX296 BX303 )
nsamples_minus_1=94
#nsamples_minus_1=9

for i in `seq 0 ${nsamples_minus_1}` ; do

    sampledir=${sampledirs[${i}]}
    sample=${samples[${i}]}

    currentdir=${workingdir}/${sampledir}
```

```
    cd $currentdir

    echo "================================================="

    echo "Sample: $sample"

    if [ ! -s ${currentdir}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam.idxstats ] ;
then
        rm -rf ${currentdir}/${sample}_Trim_hg38__STARtmp
    # align data using STAR
    echo "$(date): Started STAR alignment. Sample ${sample}"
    STAR --runThreadN ${nThreads} --runMode alignReads --genomeDir ${STARindexdir} --
readFilesIn ${currentdir}/${sample}_Trim_R1.fastq.gz ${currentdir}/${sample}_Trim_R2.fastq.gz
--readFilesCommand zcat --outSAMtype BAM Unsorted --outFileNamePrefix
${currentdir}/${sample}_Trim_hg38_ --chimSegmentMin 20 --quantMode GeneCounts --
outReadsUnmapped Fastx > ${currentdir}/${sample}_Trim_hg38_star.out 2>&1
    echo "$(date): Finished STAR alignment. Sample ${sample}"

    # sort and index BAM files using samtools
    echo "$(date): Started sorting the bam file. Sample ${sample}"
    samtools sort -@${nThreads} ${currentdir}/${sample}_Trim_hg38_Aligned.out.bam >
${currentdir}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam
    echo "$(date): Finished sorting the bam file. Sample ${sample}"

    echo "$(date): Started indexing the bam file. Sample ${sample}"
    samtools index ${currentdir}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam
    echo "$(date): Finished indexing the bam file. Sample ${sample}"

    samtools idxstats ${currentdir}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam >
${currentdir}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam.idxstats

    else
    echo "     output files present."
    fi
    echo "================================================="
    cd -

done

echo "$(date) FINISHED"

#!/bin/sh

# Description
samples=( BX006 BX008 BX010 BX011 BX012 BX013 BX014 BX017 BX018 BX020 BX023 BX024 BX026 BX028
BX030 BX031 BX038 BX042 BX045 BX048 \
    BX051 BX054 BX056 BX057 BX062 BX063 BX064 BX065 BX066 BX067 BX071 BX073 BX074 BX075 BX077
BX093 BX101 BX102 BX103 BX104 \
    BX106 BX107 BX110 BX112 BX113 BX115 BX117 BX118 BX119 BX121 BX123 BX127 BX130 BX131 BX135
BX138 BX139 BX141 BX150 BX151 \
    BX152 BX153 BX160 BX164 BX165 BX166 BX167 BX168 BX173 BX175 BX178 BX188 BX194 BX195 BX201
BX202 BX217 BX221 BX230 BX240 \
    BX242 BX245 BX248 BX272 BX275 BX278 BX281 BX285 BX287-p1 BX287-p2 BX287-p3 BX287-p4 BX287-
p6 BX296 BX303 )
nsamples_minus_1=94

module purge; module load bluebear
module load apps/samtools/1.4

for i in `seq 3 ${nsamples_minus_1}` ; do
#for i in `seq 0 2` ; do

sample=${samples[${i}]}

echo $sample

# Merge Runs and sort
if [ ! -s ${sample}_Trim.bam ] ; then
    echo "merging files"
    samtools merge -n ${sample}_Trim.bam ../GB120918-
AF/${sample}/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam ../GB120918-
AF_01/${sample}*/${sample}_Trim_hg38_Aligned.sortedByCoord.out.bam
fi
if [ ! -s ${sample}_Trim_sorted.bam ] ; then
```

```
    echo "sorting file"
    samtools sort ${sample}_Trim.bam > ${sample}_Trim_sorted.bam
fi

done

# Create Read Counts File
module purge; module load bluebear
module load Subread/1.5.3
featureCounts -a /rds/homes/h/hlc711/filera-rheumatoid-arthritis-rna-
seq/sequencing/Genomes/Human/#GRCh38/gencode.v31.annotation.gtf -o readcounts_Trim.txt
${bam_files} -p -s 2 > readcounts_Trim.out 2>&1 &
```

9.4 R SCRIPT FOR ANALYSIS OF THE RNA SEQUENCING DATA

9.4.1 Setting up of datasets

```
library(DESeq2)
library(gplots)
library(RColorBrewer)
library(ggplot2)
library(pheatmap)
library(factoextra)
library(vioplot)
library(ggpubr)
library(data.table)
library(olsrr)
library(ComplexHeatmap)
library(circlize)
library(FSA)
library(biomaRt)

####Set up data in R####
# Set-up and read in counts
outdir = "U:/R/DESeq2/Output"
nsamples = 95
readcounts <-
read.csv(file="Z:/sequencing/Processed_Data/BAM_files_merged/readcounts_Trim_noY.txt",
                        header = TRUE, row.names=1, sep = "\t", skip = 1)

# Remove chromosome Y
ChY <- grepl("chrY", readcounts[,1])
readcounts_noY <- readcounts[!ChY,]

# Remove unnecessary columns
colnames(readcounts_noY)[6:ncol(readcounts_noY)] =
  gsub("_Trim_sorted.bam","",gsub("...BAM_files.","",
                          colnames(readcounts_noY[6:ncol(readcounts_noY)])))
counts_noY <- readcounts_noY[,6:(nsamples+5)]
rownames(counts_noY) <- gsub("[[:punct:]].*","",rownames(counts_noY))
rownames(readcounts_noY) <- gsub("[[:punct:]].*","",rownames(readcounts_noY))

# Read in coldata and create factors
coldata <- read.csv(file="U:/R/DESeq2/coldata.csv", header = TRUE)
coldata$Diagnosis <- factor(coldata$Diagnosis)
coldata$sex <- factor(coldata$sex)
coldata$ccp <- factor(coldata$ccp)
coldata$rhf <- factor(coldata$rhf)
coldata$ExMethod <- factor(coldata$ExMethod)
coldata$Plate <- factor(coldata$Plate)
coldata$Pathotype <- factor(coldata$Pathotype)
coldata$joint <- factor(coldata$joint)
coldata$biopsy_side <- factor(coldata$biopsy_side)
coldata$therapy.Nsaid <- factor(coldata$therapy.Nsaid)
coldata$therapy.pred <- factor(coldata$therapy.pred)
coldata$Muscle <- factor(coldata$Muscle)
coldata$DAS28ESR_resp_12 <- factor(coldata$DAS28ESR_resp_12)

####Use BiomaRt to get gene names####
ensembl = useMart(biomart = "ENSEMBL_MART_ENSEMBL", dataset="hsapiens_gene_ensembl",
                  host="www.ensembl.org")
```

```
Attributes <- listAttributes(ensembl)

searchAttributes(mart = ensembl, pattern = "entrezgene")

Filters <- listFilters(ensembl)

searchFilters(mart = ensembl, pattern = "ensembl.*id")

keep <- rowSums(counts_noY) >= 50
counts_noY_filter <- counts_noY[keep,]

BMimport_counts =getBM(attributes= c("ensembl_gene_id", "hgnc_symbol", "entrezgene_id"),
               values= rownames(counts_noY_filter),filters="ensembl_gene_id", mart= ensembl)

anyDuplicated(rownames(counts_noY_filter))
anyDuplicated(BMimport_counts$ensembl_gene_id)

BMimport_nodup_counts <- BMimport_counts[!duplicated(BMimport_counts$hgnc_symbol),]
BMimport_match_counts <- BMimport_nodup_counts[match(rownames(counts_noY_filter),
                                BMimport_nodup_counts$ensembl_gene_id),]

counts_gene_names <- counts_noY_filter
rownames(counts_gene_names)[BMimport_match_counts$hgnc_symbol!="" &
       !is.na(BMimport_match_counts$hgnc_symbol)] <-
               BMimport_match_counts$hgnc_symbol[BMimport_match_counts$hgnc_symbol!="" &
       !is.na(BMimport_match_counts$hgnc_symbol)]

####Remove XIST, TSIX & IGV genes####
# Remove XIST & TSIX
X <- grepl("XIST", rownames(counts_gene_names)) |
  grepl(c("TSIX"), rownames(counts_gene_names))

counts_gene_names <- counts_gene_names[!X,]

# Remove IGV genes
IGVDJ <- grepl("IGHV", rownames(counts_gene_names)) |
  grepl(c("IGLV"), rownames(counts_gene_names)) |
  grepl(c("IGKV"), rownames(counts_gene_names)) |
  grepl(c("IGHD"), rownames(counts_gene_names)) |
  grepl(c("IGHJ"), rownames(counts_gene_names)) |
  grepl(c("IGLJ"), rownames(counts_gene_names)) |
  grepl(c("IGKJ"), rownames(counts_gene_names))

counts noY noIGV <- counts gene names[!IGVDJ,]

####Remove samples that failed QC####
# Remove poor quality samples from counts and coldata
counts_noIGV <- counts_noY_noIGV[,-c(1, 3, 5, 19, 37, 50)]
coldata <- coldata[-c(1, 3, 5, 19, 37, 50),]

# Remove repeats
counts_noIGV <- counts_noIGV[,-c(84:87)]
coldata <- coldata[-c(84:87),]

# Remove samples with muscle contamination - used for analyses
#where muscle not corrected for
counts_noIGV_nomusc <- counts_noIGV[,-c(1:2, 5, 11, 22, 24, 37, 42, 49, 69,
                                        75)]
coldata_nomusc <- coldata[-c(1:2, 5, 11, 22, 24, 37, 42, 49, 69, 75),]

####Set up DESeq2####
#This was also done without muscle to produce dds_nomusc
dds <- DESeqDataSetFromMatrix(countData=counts_noIGV, colData=coldata, design=~Diagnosis)
#Filter out genes with low expression
keep <- rowSums(counts(dds)) >= 50
dds <- dds[keep,]

dds$Diagnosis <- relevel(dds$Diagnosis, ref = "sdRA")

# Run DESeq2
dds <- DESeq(dds)

# vst transform data
vsd <- vst(dds)
```

```
####Create summary muscle score for use in correction####
musc_genes <- c("IDI2", "DUPD1", "MYH1", "LRRC30", "MYH4", "SMTNL1",
                "ACTN3", "PPP1R27","MYADML2","ANKRD23","UCP3","CHRNA10")

z_score <- t(scale(t(assay(vsd)), center=T, scale=T))

select <- rownames(z_score)%in%musc_genes

z_score <- z_score[select,]

#Calculate muscle score for each sample
musc_score <- colSums(z_score)/sqrt(nrow(z_score))

####Create new DESeq2 dataset that includes muscle score correction####
coldata$musc_score <- as.numeric(musc_score)

dds_musc <- DESeqDataSetFromMatrix(countData=counts_noIGV, colData=coldata,
                                   design=~ musc_score + Diagnosis)
keep <- rowSums(counts(dds_musc)) >= 50
dds_musc <- dds_musc[keep,]
dds_musc$Diagnosis <- relevel(dds_musc$Diagnosis, ref = "sdRA")
dds_musc <- DESeq(dds_musc)
vsd_musc <- vst(dds_musc)
```

9.4.2 QC and data exploration

```
####Plot heatmap of top 100 most variable genes####
#NOTE: these scripts produce plots after removal of samples containing
#muscle, this was also run including muscle for original QC plots.
library(genefilter)

topVarGenes <- head( order( rowVars( assay(vsd_nomusc) ), decreasing=TRUE ), 100 )

#Create csv file containing list of most variable genes#
write.csv(assay(vsd_nomusc)[topVarGenes,], file = "Variable_genes_musc.csv")

#Set the colour function for complexheatmap
col_fun <- colorRamp2(breaks = seq(-3, 3, length.out=101),
                      colorRampPalette(rev(brewer.pal(n=11, "RdYlBu")))(101))

#Create pdf containing heatmap
pdf(file=paste0(outdir,"/Heatmap_Var100.pdf"),width = 15,
    height = 15, paper = "special", onefile = T, title = "")
par(mar=c(6,4,2.5,7),mgp=c(4,1,0))
Heatmap(t(scale(t(assay(vsd_nomusc)[ topVarGenes,]))), name = "expression",
        col = col_fun, #border = TRUE, column_title = "Top 100 most variable genes",
        column_title_gp = gpar(fontsize = 16, fontface = "bold"),
        row_names_gp = gpar(fontsize = 3), column_names_gp = gpar(fontsize = 10),
        top_annotation = HeatmapAnnotation(Diagnosis = vsd_nomusc$Diagnosis,
                        Pathotype = vsd_nomusc$Pathotype,
                        col = list(Diagnosis = c("Norm"="black", "Res"="gold",
                        "NonRA"="green3","sdRA"="blue","ldRA"="red"),
                        Pathotype = c("Pauci_immune" = "palegreen3",
                        "Diffuse" = "cornflowerblue","Lymphoid" = "tomato2")),
                        na_col = "darkgrey"),
        cluster_columns = TRUE, row_dend_reorder = TRUE,
        column_dend_reorder = TRUE,
        clustering distance rows = function(x) as.dist(1-cor(t(x))),
        clustering_method_rows = "average",
        clustering_distance_columns = function(x) as.dist(1-cor(t(x))),
        clustering_method_columns = "average"
)
dev.off()

####PCA plotting####
#This script was adapted from source script for plotPCA to create a function
#for plotting of PC3&4
library(genefilter)
library(ggplot2)
library(ggrepel)

plotPCA.san34 <- function (object, intgroup = "condition", ntop = 500,
```

```
                          returnData = FALSE)
{
  rv <- rowVars(assay(object))
  select <- order(rv, decreasing = TRUE)[seq_len(min(ntop,
                                                     length(rv)))]
  pca <- prcomp(t(assay(object)[select, ]))
  percentVar <- pca$sdev^2/sum(pca$sdev^2)
  if (!all(intgroup %in% names(colData(object)))) {
    stop("the argument 'intgroup' should specify columns of colData(dds)")
  }
  intgroup.df <- as.data.frame(colData(object)[, intgroup, drop = FALSE])
  group <- if (length(intgroup) > 1) {
    factor(apply(intgroup.df, 1, paste, collapse = " : "))
  }
  else {
    colData(object)[[intgroup]]
  }

  ## Select the PCAs and percentVar that you like instead of 1 and 2
  d <- data.frame(PC3 = pca$x[, 3], PC4 = pca$x[, 4], group = group,
                  intgroup.df, name = colData(object)[,1])
  if (returnData) {
    attr(d, "percentVar") <- percentVar[3:4]
    return(d)
  }
  ggplot(data = d, aes_string(x = "PC3", y = "PC4", color = "group",
  label = "name")) + geom_point(size = 3) +
  xlab(paste0("PC3: ", round(percentVar[3] * 100), "% variance")) +
  ylab(paste0("PC4: ", round(percentVar[4] * 100), "% variance")) +
  coord_fixed()

}

#Produce PCA plots into a pdf
rv <- rowVars(assay(vsd_nomusc))
select <- order(rv, decreasing = TRUE)[seq_len(min(500, length(rv)))]

pca <- prcomp(t(assay(vsd_nomusc)[select,]))

pdf(file=paste0(outdir,"/PCA_plots_nomusc.pdf"),width = 15, height = 10,
    paper = "special",onefile = T, title = "")
par(mar=c(6,7,2.5,3.5),mgp=c(4,1.5,0))
print( plotPCA( vsd_nomusc, "Diagnosis" ) +
    geom_label(aes(label = vsd_nomusc$sample), label.size = 0.25, size = 4))
print( plotPCA( vsd_nomusc, intgroup = c( "Diagnosis") )  +
        ggtitle("Diagnosis"))
print( plotPCA( vsd_nomusc, intgroup = c( "sex") ) + ggtitle("Sex") )
print( plotPCA( vsd_nomusc, intgroup = c( "age") ) + ggtitle("Age") +
         scale_color_gradient(low="blue", high="red") )
print( plotPCA( vsd_nomusc, intgroup = c( "ExMethod") ) +
        ggtitle("Extraction Method") +
         scale_color_discrete(name="ExMethod",breaks=c("TS", "WT"),
                         labels=c("Thick Section", "Whole Tissue")))
print( plotPCA( vsd_nomusc, intgroup = c( "Plate") ) + ggtitle("Plate Number") )
print( plotPCA( vsd_nomusc, intgroup = c("Pathotype")) + ggtitle("Pathotype"))
print( plotPCA( vsd_nomusc, intgroup = c( "ccp") ) + ggtitle("CCP")  +
         scale_color_discrete(name="ccp", breaks=c("n", "p"), labels=c("CCP-", "CCP+")))
print( plotPCA( vsd_nomusc, intgroup = c( "rhf") ) + ggtitle("RHF") +
         scale_color_discrete(name="rhf",breaks=c("n", "p"),labels=c("RhF-", "RhF+")) )
print( plotPCA( vsd_nomusc, intgroup = c( "joint") ) + ggtitle("Joint Biopsied") )
print( plotPCA( vsd_nomusc, intgroup = c( "biopsy side") ) + ggtitle("Biopsy Side") )
print( plotPCA( vsd_nomusc, intgroup = c( "RIN") ) + ggtitle("RIN") +
         scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "LL.median") ) + ggtitle("Lining layer median") +
         scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "Density.mean") ) + ggtitle("Density Median") +
         scale_color_gradient(low="blue", high="red") )
print( plotPCA( vsd_nomusc, intgroup = c( "Agg.worst") ) + ggtitle("Aggregate worst") +
         scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "US.GS.worst") ) + ggtitle("US GS worst") +
         scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "US.PD.worst") ) + ggtitle("US PD worst") +
         scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "DAS28_ESR") ) + ggtitle("DAS28_ESR") +
```

```
            scale_color_gradient(low="blue", high="red"))
print( plotPCA( vsd_nomusc, intgroup = c( "DAS28_CRP") ) + ggtitle("DAS28_CRP") +
            scale_color_gradient(low="blue", high="red"))
print( plotPCA(vsd_nomusc, intgroup = c("therapy.Nsaid")) + ggtitle("Nsaid therapy"))
print( plotPCA(vsd_nomusc, intgroup = c("therapy.pred")) + ggtitle("Pred therapy"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("Diagnosis")) +
    geom_label(aes(label = vsd_nomusc$sample), label.size = 0.25, size = 4))
print( plotPCA.san34(vsd_nomusc, intgroup = c("sex")) + ggtitle("Sex"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("Diagnosis")) + ggtitle("Diagnosis"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "age") ) + ggtitle("Age") +
        scale_color_gradient(low="blue", high="red") )
print( plotPCA.san34(vsd_nomusc, intgroup = c("ExMethod")) + ggtitle("Extraction Method") +
        scale_color_discrete(name="ExMethod",breaks=c("TS", "WT"),
                             labels=c("Thick Section", "Whole Tissue")))
print( plotPCA.san34(vsd_nomusc, intgroup = c("Plate")) + ggtitle("Plate Number"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("Pathotype")) + ggtitle("Pathotype"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("ccp")) + ggtitle("CCP")  +
        scale_color_discrete(name="ccp",breaks=c("n", "p"),labels=c("CCP-", "CCP+")))
print( plotPCA.san34(vsd_nomusc, intgroup = c("rhf")) + ggtitle("RHF") +
        scale_color_discrete(name="rhf",breaks=c("n", "p"),labels=c("RhF-", "RhF+")))
print( plotPCA.san34(vsd_nomusc, intgroup = c("joint")) + ggtitle("Joint Biopsied"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("biopsy_side")) + ggtitle("Biopsy Side"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("RIN")) + ggtitle("RIN") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "LL.median") ) +
       ggtitle("Lining Layer median") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "Density.mean") ) +
       ggtitle("Density Median") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "Agg.worst") ) + ggtitle("Aggregate worst") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "US.GS.worst") ) + ggtitle("US GS worst") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "US.PD.worst") ) + ggtitle("US PD worst") +
        scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "DAS28_ESR") ) + ggtitle("DAS28_ESR") +
            scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34( vsd_nomusc, intgroup = c( "DAS28_CRP") ) + ggtitle("DAS28_CRP") +
            scale_color_gradient(low="blue", high="red"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("therapy.Nsaid")) + ggtitle("Nsaid Therapy"))
print( plotPCA.san34(vsd_nomusc, intgroup = c("therapy.pred")) + ggtitle("Pred Therapy"))
print( plotPCA.san56(vsd_nomusc, intgroup = c("Diagnosis")) +
    geom label(aes(label = vsd nomusc$sample), label.size = 0.25, size = 4))
print( plotPCA.san56(vsd_nomusc, intgroup = c("Diagnosis")) + ggtitle("Diagnosis") )
fviz_pca_var(pca, col.var = "contrib", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE, select.var = list(contrib = 20),
             labelsize = 6) + theme(text = element_text(size=18))
fviz_pca_var(pca, axes = c(3,4), col.var = "contrib",
        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE, select.var = list(contrib = 20),
             labelsize = 6) + theme(text = element_text(size=18))
fviz_pca_var(pca, axes = c(5,6),col.var = "contrib",
        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE, select.var = list(contrib = 20),
             labelsize = 6) + theme(text = element text(size=18))
fviz_eig(pca, ncp = 10) + theme_minimal(base_size = 20)
dev.off()

####Run stepwise regression analysis for identification of confounders####
#Run linear model on top variable genes
# Extract most variable genes and transpose
n = 100 #number of genes - analysis repeated for top 1000
topVarGenes <- head( order( rowVars( assay(vsd_nomusc) ), decreasing=TRUE ), n )
topVarGenes <- assay(vsd_nomusc)[ topVarGenes, ]
topVarGenes <- t(topVarGenes)

# create df containing log(counts)/sd(log(counts))
Var.log <- log(topVarGenes)
Var.sdlog <- Var.log
colstd <- rep(0, n)
for (i in 1:n) {
  colstd[i] <- sd(Var.log[,i])
  Var.sdlog[,i] <- Var.log[,i]/colstd[i]
```

```
}

# Calculate coefficients and create df
df <- as.data.frame(cbind(coldata_nomusc[,c(1:24,32)],as.numeric(Var.sdlog[,1])))
colnames(df)[26] = 'counts'
for (i in 2:n) {
  if (i %% 100 == 0) { message(i,'/',n) }
  dftoadd <- as.data.frame(cbind(coldata_nomusc[,c(1:24)],
    as.numeric(Var.sdlog[,i])))
  colnames(dftoadd)[25] = 'counts'
  df <- rbindlist(list(df,dftoadd))
}

fit <- lm(counts ~ Diagnosis + sex + age + ccp + rhf + ExMethod + Plate +
            Pathotype + joint + biopsy_side + therapy.Nsaid + therapy.pred +
            RIN + LL.median + DAS28_ESR + DAS28_CRP + Density.mean +
            Agg.worst + PC.worst + US.GS.worst + US.PD.worst, df)

ols_step_both_p(fit, pent=0.05)

#Run linear model on sinlge principal component
ppp <- plotPCA( vsd_nomusc, intgroup = c( "Diagnosis"))
ppa <- plotPCA.san34( vsd_nomusc, intgroup = c( "Diagnosis") )

#Example for PC4, the same script was used for PC1-3
df <- as.data.frame(cbind(coldata_nomusc[,c(1:24)], ppa$data$PC4))
colnames(df)[25] = 'PC4'

fit <- lm(PC4 ~ Diagnosis + sex + age + ccp + rhf + ExMethod + Plate +
            Pathotype + joint + biopsy_side + therapy.Nsaid + therapy.pred +
            RIN + LL.median + DAS28_ESR + DAS28_CRP + Density.mean +
            Agg.worst + PC.worst + US.GS.worst + US.PD.worst, df)

ols_step_both_p(fit, pent=0.05)

#Run linear model on all principal components
df <- as.data.frame(cbind(coldata_nomusc[,c(1:24)], ppp$data$PC1,rep("PC1",85)))
colnames(df)[25] = 'PC_coef'
colnames(df)[26] = 'PC'
dftoadd <- as.data.frame(cbind(coldata_nomusc[,c(1:24)],ppp$data$PC2,rep("PC2",85)))
colnames(dftoadd)[25] = 'PC_coef'
colnames(dftoadd)[26] = 'PC'
df <- as.data.frame(rbind(df, dftoadd))
dftoadd <- as.data.frame(cbind(coldata nomusc[,c(1:24)],ppa$data$PC3,rep("PC3",85)))
colnames(dftoadd)[25] = 'PC_coef'
colnames(dftoadd)[26] = 'PC'
df <- as.data.frame(rbind(df, dftoadd))
dftoadd <- as.data.frame(cbind(coldata_nomusc[,c(1:24)],ppa$data$PC4,rep("PC4",85)))
colnames(dftoadd)[25] = 'PC_coef'
colnames(dftoadd)[26] = 'PC'
df <- as.data.frame(rbind(df, dftoadd))

fit <- lm(PC_coef ~ Diagnosis + sex + age + ccp + rhf + ExMethod + Plate +
            Pathotype + joint + biopsy_side + therapy.Nsaid + therapy.pred +
            RIN + LL.median + DAS28_ESR + DAS28_CRP + Density.mean +
            Agg.worst + PC.worst + US.GS.worst + US.PD.worst + PC, df)

ols_step_both_p(fit, pent=0.05)

#UMAP plotting & k-means clustering
library(umap)

#Normalise data
orig_count_table = counts(dds_nomusc)
q3_columns = apply(orig_count_table,2,quantile,probs=0.75)
column_factor = median(q3_columns) / q3_columns
count_table <- t(t(orig_count_table) * column_factor)
sd_rows = apply(count_table,1,sd)
mean_rows = apply(count_table,1,mean)
count_table <- (count_table - mean_rows) / sd_rows
count_table <- count_table[order(sd_rows, decreasing = TRUE),]

#Run UMAP & plot
set.seed(12)
```

```
count.table.umap_1000 <- umap(t(count_table[1:1000,]))
toplot <- cbind(as.data.frame(count.table.umap_1000$layout),
           coldata_nomusc$Diagnosis, coldata_nomusc$sample)
colnames(toplot) <- c("x","y","Diagnosis","sample")

#Plot UMAP & create pdf
toplot <- cbind(as.data.frame(count.table.umap_1000$layout),
               coldata_nomusc$Diagnosis, coldata_nomusc$sample)
colnames(toplot) <- c("x","y","Diagnosis","sample")

ggplot(toplot, aes(x,y, color = Diagnosis)) + geom_point( size = 3) + ggtitle("Diffuse")

pdf(file=paste0(outdir,"/UMAP_plots_dds1000_nomusc.pdf"),width = 15, height = 10, paper =
"special",
    onefile = T, title = "")
par(mar=c(6,7,2.5,3.5),mgp=c(4,1.5,0))
ggplot(toplot, aes(x,y, color = Diagnosis)) + geom_point( size = 3)
ggplot(toplot, aes(x,y, color = coldata_nomusc$sex)) +
    geom_point( size = 3) + scale_color_discrete(name="Sex",
                        breaks=c("F", "M"), labels=c("Female", "Male"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$age)) + geom_point( size = 3) +
  scale_color_continuous(name="Age", type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$ExMethod)) +
    geom_point( size = 3) + scale_color_discrete(name="Extraction Method",
           breaks=c("TS", "WT"), labels=c("Thick Section", "Whole Tissue"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$RIN)) + geom_point( size = 3) +
  scale_color_continuous(name="RIN", type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$Plate)) +
    geom_point( size = 3) + scale_color_discrete(name="Plate No.")
ggplot(toplot, aes(x,y, color = coldata_nomusc$ccp)) +
    geom_point( size = 3) + scale_color_discrete(name="CCP",
        breaks=c("p", "n"), labels=c("Positive", "Negative"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$rhf)) +
    geom_point( size = 3) + scale_color_discrete(name="RHF",
        breaks=c("p", "n"), labels=c("Positive", "Negative"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$joint)) +
    geom_point( size = 3) + scale_color_discrete(name="Biopsy Joint")
ggplot(toplot, aes(x,y, color = coldata_nomusc$biopsy_side)) +
    geom_point( size = 3) + scale_color_discrete(name="Biopsy Side",
        breaks=c("L", "R"), labels=c("Left", "Right"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$US.GS.worst)) +
    geom_point( size = 3) + scale_color_continuous(name="US GS Worst",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata nomusc$US.PD.worst)) +
    geom_point( size = 3) + scale_color_continuous(name="US PD Worst",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$DAS28_ESR)) +
    geom_point( size = 3) + scale_color_continuous(name="DAS28-ESR",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata nomusc$DAS28 CRP)) +
    geom_point( size = 3) + scale_color_continuous(name="DAS28-CRP",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$therapy.Nsaid)) +
    geom_point( size = 3) + scale_color_discrete(name="Therapy Nsaid",
        breaks=c("y", "n"), labels=c("Yes", "No"))
ggplot(toplot, aes(x,y, color = coldata nomusc$therapy.pred)) +
    geom_point( size = 3) + scale_color_discrete(name="Therapy Pred",
        breaks=c("y", "n"), labels=c("Yes", "No"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$Pathotype)) +
    geom_point( size = 3) + scale_color_discrete(name="Pathotype",
        breaks=c("Lymphoid", "Diffuse","Pauci immune"))
ggplot(toplot, aes(x,y, color = coldata_nomusc$Density.mean)) +
    geom_point( size = 3) + scale_color_continuous(name="Mean Density",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$Agg.worst)) +
    geom_point( size = 3) + scale_color_continuous(name="Aggregate Worst",
        type = "viridis")
ggplot(toplot, aes(x,y, color = coldata_nomusc$LL.median)) +
    geom_point( size = 3) + scale_color_continuous(name="Lining Layer",
        type = "viridis")
dev.off()

#Cluster Identification
library(tidyverse)
```

```
library(cluster)
library(factoextra)

set.seed(12)
fviz_nbclust(count.table.umap_1000$layout, kmeans, method = "wss")
clust_test <- fviz_nbclust(count.table.umap_1000$layout, kmeans,
                           method = "wss")
ggplot(clust_test$data, aes(as.numeric(clusters), y)) +
    geom_point(size = 5) + geom_line(size = 2) + geom_vline(xintercept=3,
        linetype="dashed", size=2) + scale_x_continuous(breaks = c(1:10)) +
  labs(x="Number of clusters k", y="Total Within Sum of Squares") +
  theme_classic(base_size = 30)

fviz_nbclust(count.table.umap_1000$layout, kmeans, method = "silhouette")
clust_test2 <- fviz_nbclust(count.table.umap_1000$layout, kmeans,
                           method = "silhouette")
ggplot(clust_test2$data, aes(as.numeric(clusters), y)) +
    geom_point(size = 5) + geom_line(size = 2) +
  geom_vline(xintercept=3, linetype="dashed", size=2) +
  scale_x_continuous(breaks = c(1:10)) + labs(x="Number of clusters k",
           y="Average silhouette width") + theme_classic(base_size = 30)

set.seed(12)
gap_stat <- clusGap(count.table.umap_1000$layout, FUN = kmeans, nstart = 25,
                    K.max = 10, B = 50)
fviz_gap_stat(gap_stat)
clust_test3 <- fviz_gap_stat(gap_stat)
ggplot(clust_test3$data, aes(as.numeric(clusters), gap)) +
    geom_point(size = 5) + geom_line(size = 2) +
    geom_errorbar(aes(ymin=gap-SE.sim, ymax=gap+SE.sim),
                        width = .4, size = 1.5) +
  geom_vline(xintercept=3, linetype="dashed", size=2) +
  scale_x_continuous(breaks = c(1:10)) + labs(x="Number of clusters k",
           y="Gap statistic (k)") + theme_classic(base_size = 30)

#Run k-means clustering & plot
set.seed(12)
k3 <- kmeans(count.table.umap_1000$layout, centers = 3, nstart = 25)
toplot_umap <- cbind(as.data.frame(count.table.umap_1000$layout),
                coldata_nomusc$Pathotype, coldata_nomusc$sample,
                coldata_nomusc$Diagnosis)
colnames(toplot_umap) <- c("x","y","Pathotype","sample", "Diagnosis")
k3cluster <- factor(k3$cluster)
toplot umap <- cbind(toplot umap, k3cluster)

pdf(file=paste0(outdir,"/UMAP_plots_kmeansclust_nomusc.pdf"),width = 15,
    height = 10, paper = "special", onefile = T, title = "")
par(mar=c(6,7,2.5,3.5),mgp=c(4,1.5,0))
ggplot(toplot_umap, aes(x,y, color = k3cluster)) + geom_point( size = 3)+
  scale color discrete(name="k3cluster")
fviz_cluster(list(data = as.data.frame(count.table.umap_1000$layout),
                cluster = k3cluster))
dev.off()

#Add clustering to coldata & produce csv
coldata nomusc clust <- coldata nomusc
coldata_nomusc_clust$kmeanclust <- k3cluster
write.csv(as.data.frame(coldata_nomusc_clust),
    file="coldata_nomusc_kmeansclust.csv", row.names = FALSE)

#Find differences between clusters using differential expression
#NOTE adaptation of this script was also used for differential expression analysis of the
#clinical comparisons
dds_nomusc_clust <- DESeqDataSetFromMatrix(countData=counts_noIGV_nomusc,
                        colData=coldata_nomusc_clust, design=~0 + kmeanclust)
keep <- rowSums(counts(dds_nomusc_clust)) >= 50
dds_nomusc_clust <- dds_nomusc_clust[keep,]
dds_nomusc_clust <- DESeq(dds_nomusc_clust)
vsd_clust <- vst(dds_nomusc_clust)

#Example for finding DEGs associated with single cluster, same
#script used for the other clusters
resultsNames(dds_nomusc_clust) # lists the coefficients
res_clust1 <- results(dds_nomusc_clust,
```

```
                    contrast=list(c("kmeanclust1"),
                                  c("kmeanclust2","kmeanclust3")),
                    listValues=c(1, -1/2))

# Find significant results
sum( res_clust1$padj < 0.05, na.rm=TRUE ) #Number of DEGs
resSig1 <- res_clust1[ which(res_clust1$padj < 0.05 ), ]
resLog1 <- resSig1[which(abs(resSig1$log2FoldChange) >2),]
nrow(resLog1) #Number of DEGs with LFC>2
resOrdered1 <- resLog1[order(resLog1$log2FoldChange),]
write.csv(as.data.frame(resOrdered1),file="UMAP_kmeansclust1_diffex.csv")

#Same script used to produce resLog2 & resLog3
resLog1 <- resSig1[which(resSig1$log2FoldChange >2),]
resLog2 <- resSig2[which(resSig2$log2FoldChange >2),]
resLog3 <- resSig3[which(resSig3$log2FoldChange >2),]
resAll <- rbind(resLog1, resLog2, resLog3)
resOrdered <- resAll[order(resAll$padj),]

vsd_clust$Pathotype <- relevel(vsd_clust$Pathotype, ref = "Diffuse")
vsd_clust$Pathotype <- relevel(vsd_clust$Pathotype, ref = "Pauci_immune")
vsd_clust$Diagnosis <- relevel(vsd_clust$Diagnosis, ref = "NonRA")
vsd_clust$Diagnosis <- relevel(vsd_clust$Diagnosis, ref = "Res")
vsd_clust$Diagnosis <- relevel(vsd_clust$Diagnosis, ref = "Norm")
vsd_clust_sort <- vsd_clust[,order(vsd_clust$Pathotype)]
vsd_clust_sort <- vsd_clust_sort[,order(vsd_clust_sort$Diagnosis)]
vsd_clust_sort <- vsd_clust_sort[,order(vsd_clust_sort$kmeanclust)]

toplot <- assay(vsd_clust_sort,normalized=TRUE)
select<-
  rownames(toplot)%in%rownames(resAll)
toplot <- toplot[select,]

pdf(file=paste0(outdir,"/kmeansclust_diffex.pdf"),width = 12,
    height = 14, paper = "special", onefile = T, title = "")
par(mar=c(3,3,10,6))
draw(Heatmap(t(scale(t(toplot))), name = "expression", col = col_fun,
    column_title = "Genes enriched in UMAP clusters padj<0.05 LFC>2",
    column_title_gp = gpar(fontsize = 16, fontface = "bold"),
    row_names_gp = gpar(fontsize = 8), column_names_gp = gpar(fontsize = 8),
    top_annotation = HeatmapAnnotation(Cluster = vsd_clust_sort$kmeanclust,
        Diagnosis = vsd_clust_sort$Diagnosis,
        Pathotype = vsd_clust_sort$Pathotype,
        col = list(Cluster = c("1" = "lightseagreen", "2" = "orchid3",
            "3" = "royalblue"), Diagnosis = c("Norm"="darkgrey",
            "Res"="gold", "NonRA"="green3","sdRA"="darkorchid3",
            "ldRA"="darkorange"),Pathotype = c("Pauci_immune" = "palegreen3",
            "Diffuse" = "cornflowerblue","Lymphoid" = "tomato2")),
            na_col = "darkgrey"), cluster_columns = FALSE,
    row dend reorder = TRUE,
    clustering_distance_rows = function(x) as.dist(1-cor(t(x))),
    clustering_method_rows = "average"))
dev.off()

#GO biological processes analysis on resulting DEGs
#NOTE this script was also used for GO biological process analysis following differential
#expression analysis of the clinical comparisons
library("clusterProfiler")
library("org.Hs.eg.db")
library(DOSE)
library(enrichplot)
library(ggplot2)

#Shows script for cluster 1 as example
d <- resLog1
#Get Entrez IDs for DEGs
d$entrez <- mapIds(org.Hs.eg.db,
                   keys=rownames(d),
                   column="ENTREZID",
                   keytype="SYMBOL",
                   multiVals="first")
geneList <- as.vector(d$log2FoldChange)
names(geneList) <- as.character(d$entrez)
geneList <- geneList[na.omit(names(geneList))]
```

```
geneList <- sort(geneList, decreasing = TRUE)

#Run clusterProfiler
OrgDb <- org.Hs.eg.db
edo <- clusterProfiler::enrichGO(gene = names(geneList),
            OrgDb = OrgDb,ont = "MF",pAdjustMethod = "BH",readable= F)

#simplify removes redundant terms
edos <- simplify(edo, cutoff=0.5, by="p.adjust", select_fun=min)

##Convert Entrez ID to Symbol so genes are readable & produce
#table of pathways as output
edox <- setReadable(edos, 'org.Hs.eg.db', 'ENTREZID')
pathway_table <- edox[,c("Description", "p.adjust", "geneID", "Count")]
write.csv(as.data.frame(pathway_table),
    file="UMAP_cluster_p005LFC2_pathway_table_simplify05_clust1_all.csv")

heatplot.orderCat(edox, foldChange=geneList, showCategory = 10,
    label_format = 22) + theme(axis.text.y = element_text(size=14),
    axis.text.x = element_text(size=8))
```

9.4.3 Clinical comparisons & mechanisms of resolution

Script for differential expression analysis and GO biological process analysis is included above for identification of genes associated with each UMAP cluster. The same script was used with only small adaptations for the same analyses in both the clinical comparisons and mechanisms of resolution chapters so to avoid repeats this will not be included again.

```
####Testing for average gene expression across groups####
#NOTE: Script shown is for average expression of genes associated with
#response across pathotypes as an example but slightly adapted script is
#used on multiple occasions, including for the exploration of metabolic
#gene expression across sdRA and resolving disease.

#resSig is the output from differential expression analysis of responders
#versus non-responders
resp_genes <- rownames(resSig[which(resSig$log2FoldChange >1),])

#Calculate z score & select response genes
z_score_nomusc <- t(scale(t(assay(vsd_nomusc)), center=T, scale=T))
select<- rownames(z_score_nomusc)%in%resp_genes
z_score_resp <- z_score_nomusc[select,]

#Calculate summary expression for each sample
resp_summary <- colSums(z_score_resp)/sqrt(nrow(z_score_resp))

#Set up data
to_plot <- as.data.frame(cbind(resp_summary, as.data.frame(coldata_nomusc$Pathotype),
    as.data.frame(coldata_nomusc$Diagnosis)))
colnames(to_plot)[2] <- "Pathotype"
colnames(to_plot)[3] <- "Diagnosis"
to_plot <- to_plot[to_plot$Diagnosis %in% c("sdRA","ldRA"),]
to_plot$Pathotype = factor(to_plot$Pathotype, levels=c("Pauci_immune", "Diffuse", "Lymphoid"))

#Plot against pathotype
vioplot(resp_summary ~ Pathotype, data = to_plot,
        col = c("palegreen3","cornflowerblue","tomato2"),
        xlab = 'Pathotype', ylab = 'Summary score: response',
        main = 'Response genes relative expression')

#Test for signficance
library(FSA)
Summarize(resp_summary ~ Pathotype, data = to_plot)
kruskal.test(resp_summary ~ Pathotype, data = to_plot)
PT <- dunnTest(resp_summary ~ Pathotype, data = to_plot, method="bh")
PT <- PT$res
PT

####Exploration of individual gene expression against groups####
```

```
#NOTE: this is again an example script and was used to test multiple genes
#against pathotype as well as clinical group
EGR2_nc <- counts(dds_musc, normalized = TRUE)["EGR2",]
EGR2_nc[EGR2_nc==0] <- 1
EGR2_log <- log(EGR2_nc)

#Combine gene expression & Diagnosis
to_plot <- as.data.frame(cbind(EGR2_log, as.data.frame(dds_musc$Diagnosis)))
colnames(to_plot) <- c("EGR2","Diagnosis")
to_plot$Diagnosis = factor(to_plot$Diagnosis,
                           levels=c("Norm", "Res", "NonRA", "sdRA", "ldRA"))

vioplot(EGR2 ~ Diagnosis, data = to_plot,
        col = c("darkgrey","gold","green3","darkorchid3","darkorange"),
        xlab = 'Clinical group', ylab = 'log EGR2 expression',
        main = 'Diagnosis EGR2 expression')

Summarize(EGR2 ~ Diagnosis, data = to_plot)
kruskal.test(EGR2 ~ Diagnosis, data = to_plot)
PT <- dunnTest(EGR2 ~ Diagnosis, data = to_plot, method="bh")
PT <- PT$res
PT

####Creating heatmaps of genes involved in metabolic processes###
#NOTE: shows example script. Gene lists for other metabolic processes can
#be found in Appendix 9.7.
FA_synthesis_genes <- c("ACACA", "ACACB", "ACSF3", "MCAT", "FASN", "OXSM",
                    "CBR4", "HSD17B8", "HTD2", "MECR", "OLAH", "ACSL6",
                    "ACSL4", "ACSL1", "ACSL5", "ACSL3", "ACSBG1", "ACSBG2")

#Set up data
vsd_nomusc$Pathotype <- relevel(vsd_nomusc$Pathotype, ref = "Diffuse")
vsd_nomusc$Pathotype <- relevel(vsd_nomusc$Pathotype, ref = "Pauci_immune")
vsd_nomusc$Diagnosis <- relevel(vsd_nomusc$Diagnosis, ref = "Res")
vsd_test_counts_sort <- vsd_nomusc[,order(vsd_nomusc$Pathotype)]
vsd_test_counts_sort <- vsd_test_counts_sort[,order(vsd_test_counts_sort$Diagnosis)]
vsd_test_counts_sort <-
    vsd_test_counts_sort[,vsd_test_counts_sort$Diagnosis %in% c("Res","sdRA")]
vsd_test_counts_sort$Diagnosis <- factor(vsd_test_counts_sort$Diagnosis)

col_fun = colorRamp2(breaks = seq(-3, 3, length.out=101),
colorRampPalette(rev(brewer.pal(n=11, "RdYlBu")))(101))

toplot <- assay(vsd test counts sort,normalized=TRUE)
select <- rownames(toplot)%in%FA_synthesis_genes
toplot <- toplot[select,]

pdf(file=paste0(outdir,"/metabolism_compheatmap_FA_synthesis.pdf"),
    width = 18, height = 10, paper = "special", onefile = T, title = "")
par(mar=c(3,3,10,6))
draw(Heatmap(t(scale(t(toplot))), name = "expression", col = col_fun,
    column_title = "FA synthesis genes", column_title_gp = gpar(fontsize = 16,
        fontface = "bold"),row_names_gp = gpar(fontsize = 8),
    column_names_gp = gpar(fontsize = 10),
    top_annotation = HeatmapAnnotation(Diagnosis = vsd_test_counts_sort$Diagnosis,
        Pathotype = vsd test counts sort$Pathotype,
        col = list(Diagnosis = c("Norm" = "darkgrey", "Res" = "gold",
        "NonRA" = "green3","sdRA" = "darkorchid3", "ldRA" = "darkorange"),
        Pathotype = c("Lymphoid" = "tomato2", "Diffuse" = "cornflowerblue",
        "Pauci_immune" = "palegreen3")), na_col = "darkgrey"),
        cluster columns = TRUE, row dend reorder = TRUE,
        column_dend_reorder = TRUE,
        clustering_distance_rows = function(x) as.dist(1-cor(t(x))),
        clustering_method_rows = "average",
        clustering_distance_columns = function(x) as.dist(1-cor(t(x))),
        clustering_method_columns = "average",
))
dev.off()
```

9.5 GO BIOLOGICAL PROCESS LISTS

9.5.1 Processes associated with UMAP clusters

Table 9.4: GO biological processes associated with DEGs upregulated in UMAP cluster C1. DEGs from padj<0.05, LFC>2.

GO ID	Description	p.adjust	geneID	Count
GO:0048705	skeletal system morphogenesis	0.000121754	COL1A1/PAX1/COL11A1/ACAN/CHAD/SOX11/RFLNA	7
GO:0030198	extracellular matrix organization	0.000121754	WT1/IBSP/ADAM12/COL1A1/COL11A1/COL10A1/ACAN/POSTN	8
GO:0043062	extracellular structure organization	0.000121754	WT1/IBSP/ADAM12/COL1A1/COL11A1/COL10A1/ACAN/POSTN	8
GO:0001503	ossification	0.002106495	IBSP/CHRDL2/COL1A1/CTHRC1/COL11A1/SOX11/RFLNA	7
GO:0010721	negative regulation of cell development	0.018377648	PAEP/CCL11/POSTN/SOX11/RFLNA	5
GO:0060231	mesenchymal to epithelial transition	0.018426575	WT1/GDNF	2
GO:0009612	response to mechanical stimulus	0.018426575	ANKRD1/COL1A1/COL11A1/POSTN	4
GO:0098743	cell aggregation	0.018426575	COL11A1/ACAN	2
GO:1903532	positive regulation of secretion by cell	0.019131234	PAEP/ANKRD1/GDNF/POSTN/SOX11	5
GO:1903510	mucopolysaccharide metabolic process	0.021834447	STAB2/CHST6/ACAN	3
GO:0071560	cellular response to transforming growth factor beta stimulus	0.021834447	ANKRD1/COL1A1/POSTN/SOX11	4
GO:0018146	keratan sulfate biosynthetic process	0.021834447	CHST6/ACAN	2
GO:0071295	cellular response to vitamin	0.025451606	COL1A1/POSTN	2
GO:0071356	cellular response to tumor necrosis factor	0.030541294	CCL11/ANKRD1/COL1A1/POSTN	4
GO:0051146	striated muscle cell differentiation	0.030541294	WT1/ALPK2/ANKRD1/ADAM12	4
GO:0110148	biomineralization	0.040770934	IBSP/COL1A1/RFLNA	3

Table 9.5: GO biological processes associated with DEGs downregulated in UMAP cluster C1. DEGs from padj<0.05, LFC<2.

GO ID	Description	p.adjust	geneID	Count

GO:0016042	lipid catabolic process	0.012990623	LIPE/PLIN1/CYP3A4/CIDEA/LEP	5
GO:0008343	adult feeding behavior	0.012990623	LEP/BRS3	2
GO:0032275	luteinizing hormone secretion	0.012990623	LEP/CRH	2
GO:0070091	glucagon secretion	0.012990623	LEP/CRH	2
GO:0032276	regulation of gonadotropin secretion	0.012990623	LEP/CRH	2
GO:0046173	polyol biosynthetic process	0.014206175	CYP3A4/LEP/MAS1	3
GO:0034389	lipid droplet organization	0.030768815	CIDEC/CIDEA	2
GO:0042738	exogenous drug catabolic process	0.034215169	CYP3A4/CYP4B1	2
GO:0051953	negative regulation of amine transport	0.034215169	LEP/CRH	2
GO:0007586	digestion	0.038498682	CHIT1/LEP/CRH	3
GO:1905952	regulation of lipid localization	0.044031295	CIDEA/LEP/CRH	3
GO:0044058	regulation of digestive system process	0.044031295	LEP/CRH	2
GO:0006631	fatty acid metabolic process	0.044031295	LIPE/ACOD1/CYP3A4/LEP	4
GO:0140353	lipid export from cell	0.044031295	LEP/CRH	2
GO:0044060	regulation of endocrine process	0.044031295	LEP/CRH	2
GO:0071466	cellular response to xenobiotic stimulus	0.044031295	CYP3A4/GLYAT/CRH	3
GO:0008217	regulation of blood pressure	0.044031295	LEP/CRH/BRS3	3
GO:0043949	regulation of cAMP-mediated signaling	0.044483198	PDE11A/CRH	2
GO:0007565	female pregnancy	0.044483198	ACOD1/LEP/CRH	3

Table 9.6: GO biological processes associated with DEGs upregulated in UMAP cluster C2. DEGs from padj<0.05, LFC>2.

GO ID	Description	p.adjust	geneID	Count
GO:0016999	antibiotic metabolic process	3.05E-06	CYP4B1/DUOXA2/DUOX2/ADH4/ADH1A/ADH1B/HBG2/HBA1/PCK1/ HBB/AKR1B10/ADH1C/HBQ1/HBZ/HBA2	15
GO:0015671	oxygen transport	6.99E-05	HBG2/HBA1/HBB/HBQ1/HBZ/HBA2	6
GO:0042572	retinol metabolic process	0.000173397	ADH4/ADH1A/ADH1B/CYP3A4/AKR1B15/RBP4/AKR1B10/ADH1C	8

GO:0006936	muscle contraction	0.000548104	MYH14/KCNIP2/SPX/ATP1A2/MYOT/SCN7A/ADRA1A/SORBS1/NKX2-5/CRYAB/KCNJ3/PGAM2/MYH2/DES/MYOCD/CTNNA3/SCN4A/CAV3/CASQ2/MYH11	20
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	0.000589151	ADIPOQ/FAT2/CBLN1/PCDH15/CLDN10/MYOT/CLDN9/CDH22/CDH20/CDH19/PCDH10/CDH7/ANXA3/VSTM2L/MYPN/NLGN1/CLDN2	17
GO:0015701	bicarbonate transport	0.000881306	HBA1/HBB/CA3/CA4/HBA2/SLC4A9/SLC4A1	7
GO:1990748	cellular detoxification	0.002096938	DUOX2/ADH4/HBG2/HBA1/HBB/AKR1B10/HBQ1/MGST1/HBZ/HBA2	10
GO:0098754	detoxification	0.002096938	DUOX2/ADH4/HBG2/HBA1/MTARC1/HBB/AKR1B10/HBQ1/MGST1/HBZ/HBA2	11
GO:0042391	regulation of membrane potential	0.003480157	BEST2/CHRNA4/MYH14/KCNIP2/CBLN1/MYOC/ATP1A2/SCN7A/GRIN2B/ADRA1A/GABRG1/KCNJ3/GABRE/CTNNA3/SCN4A/NLGN1/CAV3/GRIK5/MAPT/CASQ2	20
GO:0042593	glucose homeostasis	0.01223055	LEP/MLXIPL/ADIPOQ/PCK1/CRH/RBP4/SSTR5/KLF15/ADCY5/PDK4/CAV3/GRIK5/OPRK1	13
GO:0008217	regulation of blood pressure	0.013705943	BRS3/LEP/ADIPOQ/NPY1R/SPX/ATP1A2/HBB/ADRA1A/CRH/NPR3/TAC4	11
GO:0050804	modulation of chemical synaptic transmission	0.016167233	SHISA9/CHRNA4/SCGN/ADIPOQ/NPY5R/CBLN1/ATP1A2/GRIN2B/ADRA1A/LRRTM1/CALB2/CRH/SORCS3/LGI1/NLGN1/GRIK5/MAPT/LRFN2	18
GO:0099177	regulation of trans-synaptic signaling	0.016167233	SHISA9/CHRNA4/SCGN/ADIPOQ/NPY5R/CBLN1/ATP1A2/GRIN2B/ADRA1A/LRRTM1/CALB2/CRH/SORCS3/LGI1/NLGN1/GRIK5/MAPT/LRFN2	18
GO:0034389	lipid droplet organization	0.019201472	CIDEA/CIDEC/PLIN5/HSD17B13	4
GO:0046890	regulation of lipid biosynthetic process	0.020607641	LEP/MLXIPL/ACADL/ADIPOQ/PLIN5/HSD17B13/THRSP/SORBS1/BMP5/GPAM/PDK4	11
GO:0030539	male genitalia development	0.023348426	LHCGR/BMP5/FGF8/GREB1L	4
GO:0032868	response to insulin	0.023348426	LEP/OTC/TRARG1/ADIPOQ/GRB14/PFKFB1/PCK1/SORBS1/LPL/KLF15/ACVR1C/PDK4/OPRK1	13
GO:0032275	luteinizing hormone secretion	0.023348426	LEP/CRH/OPRK1	3
GO:0035637	multicellular organismal signaling	0.023348426	MYH14/KCNIP2/ATP1A2/SCN7A/RYR3/NKX2-5/KCNJ3/FXYD3/CTNNA3/SCN4A/CASQ2	11
GO:0034284	response to monosaccharide	0.025942419	LEP/MLXIPL/ADIPOQ/PCK1/CRH/LPL/ADCY5/ACVR1C/SLC30A8/GRIK5/OPRK1	11
GO:0099054	presynapse assembly	0.028777744	CBLN1/CBLN2/LRRTM1/LRRTM3/NLGN1	5

GO:1901379	regulation of potassium ion transmembrane transport	0.028777744	DPP10/KCNIP2/DPP6/CAV3/KCNAB1/CASQ2/OPRK1	7
GO:0060037	pharyngeal system development	0.03178012	NKX2-5/BMP5/FGF8/EYA1	4
GO:0010889	regulation of sequestering of triglyceride	0.032820458	CIDEA/PLIN5/LPL	3
GO:0050953	sensory perception of light stimulus	0.035471452	EYS/RORB/PCDH15/ZIC2/MYO3A/RBP4/SFRP5/USH1C/POU6F2/VSX1/ SLITRK6	11
GO:0016042	lipid catabolic process	0.036337284	LEP/PLIN1/CIDEA/ACADL/LIPE/ADIPOQ/PLIN5/CYP3A4/FABP4/PCK1/ LIPH/AKR1B10/LPL/LGALS12	14
GO:0050746	regulation of lipoprotein metabolic process	0.036984623	LEP/APOD/ANGPTL8	3
GO:0060347	heart trabecula formation	0.036984623	NKX2-5/RBP4/CAV3	3
GO:0042487	regulation of odontogenesis of dentin-containing tooth	0.040704623	FGF8/DMP1/ENAM	3
GO:0007626	locomotory behavior	0.041407841	CHRNA4/PAK5/NPY1R/ATP1A2/PPP1R1B/LRRTM1/CRH/ADCY5/OPRK1/ SLITRK6	10
GO:0014706	striated muscle tissue development	0.045974863	MYH14/MYF6/ADRA1A/TCF21/SGCG/NKX2-5/RBP4/BMP5/FGF8/MYOCD/ EYA1/CAV3/HLF/KCNAB1/MYH11	15
GO:0061384	heart trabecula morphogenesis	0.045974863	NKX2-5/RBP4/BMP5/CAV3	4

Table 9.7: GO biological processes associated with DEGs downregulated in UMAP cluster C2. DEGs from padj<0.05, LFC<2.

GO ID	Description	p.adjust	geneID	Count
GO:0042113	B cell activation	6.63E-14	IL6/LAX1/FOXP3/TNFRSF13C/FCRL3/IGLL5/CTLA4/BATF/IGKC/ CD19/CXCR5/CD79A/IGHG1/IGHG3/IGHA1/IFNE/IGHM/IGLC2/ IGHG4/IGHG2/SLAMF8/MZB1/CR2/IGLC3/IGLC6/IGLC7/IL21/IGHE	28
GO:0006959	humoral immune response	6.01E-13	IL6/CXCL8/LTA/IGLL5/CD5L/POU2AF1/IGKC/CD19/CXCL1/CXCL9/ IGHG1/IGHG3/IGHA1/IFNE/IGHM/PDCD1/IGLC2/IGHG4/IGHG2/ IFNG/CR2/IGLC3/IGLC6/CR1/CXCL13/IGLC7/CXCL6/IGHE	28
GO:0042742	defense response to bacterium	8.52E-10	IL6/LTA/IGLL5/HAMP/IGKC/CLEC4E/NLRP10/IGHG1/IGHG3/IGHA1/ IFNE/IGHM/IGLC2/IGHG4/IGHG2/SLAMF8/IGLC3/IGLC6/CXCL13/ IGLC7/CXCL6/CLEC4D/IGHE	23
GO:0006910	phagocytosis, recognition	1.83E-09	IGLL5/IGKC/IGHG1/IGHG3/IGHA1/IGHM/IGLC2/IGHG4/IGHG2/	13

			IGLC3/IGLC6/IGLC7/IGHE	
GO:0050900	leukocyte migration	2.71E-09	CCR5/GPR18/IL6/PLA2G7/SLC7A5/ADAM8/CXCL8/ADGRE2/IGKC/ CXCR5/CXCL1/CXCL9/IGHA1/IGHM/IGLC2/CCL15/SLAMF8/IGLC3/ IGLC6/FCAMR/CXCL13/MMP1/IGLC7/CXCL6/CCL18/CCL11/CCL7	27
GO:0070098	chemokine-mediated signaling pathway	3.51E-08	CCR5/CXCL8/CXCR5/CXCR6/CXCL1/CXCL9/CCL15/CXCL13/CXCL6/ CCL18/CCL11/CCL7	12
GO:0001819	positive regulation of cytokine production	1.51E-06	IL6/FOXP3/TNFRSF13C/KIR2DL4/IRF4/CD80/ADAM8/LTA/CSF2/ NLRP2/TIGIT/CLEC4E/NLRP10/CCBE1/GBP5/IDO1/NLRP7/IFNG/ AIM2/CLEC6A/IL21/PAEP	22
GO:0050727	regulation of inflammatory response	8.26E-05	IL6/PLA2G7/FOXP3/ADAM8/LTA/IL22RA2/SIGLEC10/NLRP10/GBP5/ SMPDL3B/IDO1/NLRP7/SLAMF8/IL2RA/MMP3/IL21/CCL18	17
GO:0002544	chronic inflammatory response	9.73E-05	FOXP3/LTA/IDO1/CXCL13/CCL11	5
GO:0032461	positive regulation of protein oligomerization	0.000370541	TCL1A/BIK/AIM2/MMP3/MMP1	5
GO:0001776	leukocyte homeostasis	0.001817666	TNFRSF17/IL6/FOXP3/TNFRSF13C/GPR174/IL2RA/CXCL6	7
GO:0002377	immunoglobulin production	0.002221911	IL6/LAX1/FOXP3/TRAV9-2/FCRL3/BATF/IGKC/MZB1/TRDV1/TRDV2	10
GO:0050715	positive regulation of cytokine secretion	0.004621851	NLRP2/CLEC4E/NLRP10/NLRP7/IFNG/AIM2/CLEC6A/PAEP	8
GO:0070234	positive regulation of T cell apoptotic process	0.009581431	ADAM8/IDO1/PDCD1	3
GO:0072503	cellular divalent inorganic cation homeostasis	0.010694635	CCR5/GPR18/GPR174/P2RX2/CD19/CXCR5/CXCR6/CXCL9/MCHR1/ NTSR1/MT1H/CCL15/CXCL13/CCL11/CCL7	15
GO:0071887	leukocyte apoptotic process	0.019058632	CCR5/IL6/ADAM8/IDO1/PDCD1/IL2RA	6
GO:0010524	positive regulation of calcium ion transport into cytosol	0.042728414	P2RX2/CD19/CXCL9/NTSR1	4

Table 9.8: GO biological processes associated with DEGs upregulated in UMAP cluster C3. DEGs from padj<0.05, LFC>2.

GO ID	Description	p.adjust	geneID	Count
GO:0042113	B cell activation	2.07E-15	IGHE/IL21/IGHM/CR2/IGHA1/CD19/SLAMF8/IGLL5/CTLA4/IGLC7/FCRL1/ CXCR5/IGHG2/IGHG3/MZB1/MS4A1/FCRL3/FOXP3/IGKC/CD79A/IGHG1/ IGLC2/CD27/IGHG4/CD38/BATF/LAX1	27
GO:0051249	regulation of lymphocyte activation	2.07E-15	PLA2G2D/IFNG/IDO1/IGHE/IL21/IGHM/CD80/IL27/IGHA1/CD19/PDCD1/ SLAMF8/IGLL5/CTLA4/IGLC7/CD274/TIGIT/IGHG2/IGHG3/MZB1/FCRL3/ FOXP3/IGKC/LILRB4/IGHG1/IGLC2/CD27/IGHG4/ADAM8/CD38/LAX1/SIRPG	32

GO:0002429	immune response-activating cell surface receptor signaling pathway	4.54E-15	CLEC6A/CLEC4D/IGHE/IGHM/CLEC4E/CR2/IGHA1/BTN1A1/CD19/CLEC4C/ IGLL5/CTLA4/FCGR1A/IGLC7/IGHG2/IGHG3/MS4A1/FCRL3/FOXP3/TRAV8-4/ MUC16/IGKC/LILRB4/CD79A/IGHG1/IGLC2/HLA-DQB2/IGHG4/CD38/ LAX1/PAX5	31
GO:0002757	immune response-activating signal transduction	4.54E-15	CLEC6A/CLEC4D/IGHE/IGHM/CLEC4E/CR2/IGHA1/BTN1A1/CD19/CLEC4C/ IGLL5/CTLA4/FCGR1A/IGLC7/IGHG2/IGHG3/MS4A1/FCRL3/FOXP3/TRAV8-4/ MUC16/IGKC/LILRB4/CD79A/IGHG1/IGLC2/HLA-DQB2/IGHG4/CD38/LAX1/PAX5	31
GO:0070098	chemokine-mediated signaling pathway	4.14E-09	CCL7/CXCL9/CXCL10/CXCL13/CXCL11/CCL18/CXCR5/CCL8/CCR8/CCR5/ CXCR3/CXCL6	12
GO:0006910	phagocytosis, recognition	3.52E-08	IGHE/IGHM/IGHA1/IGLL5/IGLC7/IGHG2/IGHG3/IGKC/IGHG1/IGLC2/IGHG4	11
GO:0050900	leukocyte migration	1.24E-07	CCL7/CXCL9/CXCL10/CXCL13/CXCL11/FCAMR/IGHM/IGHA1/CCL18/SLAMF8/ MMP1/IGLC7/CXCR5/PLA2G7/CCL8/IGKC/IGLC2/CCR5/ADAM8/CXCR3/ SIRPG/CXCL6	22
GO:0042742	defense response to bacterium	1.49E-07	CXCL13/CLEC4D/IGHE/IGHM/CLEC4E/IGHA1/SLAMF8/IGLL5/IGLC7/IGHG2/ IGHG3/LTA/IGKC/IGHG1/IGLC2/IGHG4/LYZ/CXCL6	18
GO:0002507	tolerance induction	0.000114	ACOD1/IDO1/PDCD1/CD274/FOXP3	5
GO:0002831	regulation of response to biotic stimulus	0.000178	CLEC6A/ACOD1/IFNG/GBP5/CLEC4D/IL21/CLEC4E/IL27/AIM2/CLEC4C/ SLAMF8/CD274/MUC16/ADAM8/CXCL6	15
GO:0070234	positive regulation of T cell apoptotic process	0.000206	IDO1/PDCD1/CD274/ADAM8	4
GO:0050727	regulation of inflammatory response	0.000321	ACOD1/GBP5/IDO1/IL21/SIGLEC10/CCL18/SLAMF8/IL22RA2/PLA2G7/LTA/ SMPDL3B/FOXP3/TLR10/ADAM8	14
GO:0002544	chronic inflammatory response	0.0006	CXCL13/IDO1/LTA/FOXP3	4
GO:0001819	positive regulation of cytokine production	0.000697	CLEC6A/IFNG/GBP5/IDO1/IL21/CD80/CLEC4E/IL27/AIM2/CD274/TIGIT/LTA/ FOXP3/NLRP2/ADAM8	15
GO:0051281	positive regulation of release of sequestered calcium ion into cytosol	0.000742	CXCL9/CXCL10/CXCL11/CD19/NTSR1	5
GO:0055074	calcium ion homeostasis	0.000797	CCL7/CXCL9/CXCL10/CXCL13/CXCL11/CD19/NTSR1/CXCR5/MS4A1/CCL8/ SNX10/CCR8/CCR5/CD38/CXCR3	15
GO:0046717	acid secretion	0.002657	PLA2G2D/NTSR1/TRH/KMO/AQP9/SLC22A16/SNX10	7
GO:0014048	regulation of glutamate secretion	0.005838	NTSR1/TRH/KMO	3
GO:1903532	positive regulation of secretion by cell	0.005838	CLEC6A/IFNG/SPP1/CLEC4E/AIM2/NTSR1/TRH/CD274/KMO/NLRP2/ ADAM8/CD38	12

GO:0050663	cytokine secretion	0.005986	CLEC6A/IFNG/GBP5/CLEC4E/AIM2/CD274/FOXP3/NLRP2/TLR10	9
GO:0002377	immunoglobulin production	0.006374	TRDV1/TRAV14DV4/MZB1/FCRL3/FOXP3/IGKC/BATF/LAX1	8
GO:0001894	tissue homeostasis	0.016736	SPP1/IGHA1/IGHG3/IGKC/SNX10/ADAM8/CD38/LYZ	8
GO:0032461	positive regulation of protein oligomerization	0.018102	TCL1A/AIM2/MMP1	3
GO:0001773	myeloid dendritic cell activation	0.023657	UBD/CLEC4D/BATF	3
GO:0009074	aromatic amino acid family catabolic process	0.023657	IL4I1/IDO1/KMO	3

Table 9.9: GO biological processes associated with DEGs downregulated in UMAP cluster C3. DEGs from padj<0.05, LFC<2.

GO ID	Description	p.adjust	geneID	Count
GO:0006936	muscle contraction	0.001663634	DES/MYH11/KCNJ12/CTNNA3/MYOCD/CACNA1S/KCNB2/ KCNJ3/NKX2-5/MYOT/SCN7A/MYBPC1/MYL1/MYH1/MYH2	15
GO:0042445	hormone metabolic process	0.004906162	CRABP1/LRAT/BMP5/AKR1B10/AKR1B15/DIO3/CHST9/ PCSK2/CYP2C9/DUOX2/DUOXA2	11
GO:0071772	response to BMP	0.007932764	SOSTDC1/BMP5/SFRP2/TNMD/TBXT/SFRP5/BMPR1B/ NKX2-5/VSTM2A	9
GO:0071773	cellular response to BMP stimulus	0.007932764	SOSTDC1/BMP5/SFRP2/TNMD/TBXT/SFRP5/BMPR1B/ NKX2-5/VSTM2A	9
GO:0060562	epithelial tube morphogenesis	0.007932764	SALL1/SOSTDC1/SHROOM3/BMP5/TCF21/SFRP2/ALOX12/ ALX1/TBXT/EYA1/NKX2-5/HHIP	12
GO:0051965	positive regulation of synapse assembly	0.007932764	LRRTM3/ADGRL3/SLITRK6/CBLN1/GRID2/LRRTM1	6
GO:1901890	positive regulation of cell junction assembly	0.007932764	LRRTM3/ADGRL3/SLITRK6/CBLN1/GRID2/MYOC/LRRTM1	7
GO:0007389	pattern specification process	0.009675471	SOSTDC1/SHROOM3/BMP5/SFRP2/PCDH8/ALX1/TBXT/EYA1/ TDRD5/MYF6/BMPR1B/NKX2-5/HHIP/PAX1	14
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	0.024914492	TENM2/ADGRL3/PCDH8/CBLN1/CDH7/GRID2/UNC5D/CDH19/ MYOT/PCDH10	10
GO:1900452	regulation of long-term synaptic depression	0.028484386	CBLN1/GRID2/SORCS3	3

GO:0010002	cardioblast differentiation	0.041313358	MYOCD/TBXT/NKX2-5	3
GO:0014706	striated muscle tissue development	0.041313358	RBM24/MYH11/BMP5/JPH2/TCF21/MYOCD/EYA1/MYF6/ NKX2-5/ANKRD1/MYBPC1	11
GO:0050953	sensory perception of light stimulus	0.044197612	CRYBB2/LRAT/MYO3A/SLITRK6/POU6F2/SFRP5/USH1C/ZIC2	8
GO:0045445	myoblast differentiation	0.04684225	RBM24/SOSTDC1/MYOCD/TBXT/MYF6	5
GO:0048562	embryonic organ morphogenesis	0.049929453	SALL1/TCF21/MYO3A/ALX1/SLITRK6/TBXT/EYA1/NKX2-5/USH1C	9

9.5.2 Processes associated with response to treatment

Table 9.10: GO biological processes associated with DEGs upregulated in responders. DEGs from unadjusted p<0.05, absolute LFC>1.

GO ID	Description	padj	geneID	Count
GO:0042110	T cell activation	2.90E-38	IFNG/IDO1/IL21/PLA2G2D/WNT1/CCL19/CD80/CLEC4E/CD274/PDCD1/ EOMES/TNFRSF13C/SIRPG/CTLA4/CCL5/THEMIS/IL2RA/TIGIT/ICOS/ITK/ LILRB4/BATF/TBX21/CEACAM1/GPR18/CD3D/CD2/XCL1/CD40LG/CD6/ CD8A/CD3E/RASGRP1/LCK/LAG3/SIT1/SIRPB1/IL10/CARD11/IRF1/ SLA2/EBI3/CD83/SOCS1/CD3G/GATA3/TNFSF14/CD5/ HLA-DOA/PDCD1LG2/NLRC3/RHOH/BCL11B/TESPA1/ZAP70/SEMA4A/ ZNF683/IL12RB1/PTPRC/IL18/CD160/CD8B/ITGAL/ICAM1/FUT7/ RUNX3/FCER1G/TCF7/PRKCQ/RIPOR2/GRAP2/CORO1A/P2RX7/SASH3/ HLA-DPA1/PTPN6/CD74/FZD5/PIK3CD	79
GO:0007159	leukocyte cell-cell adhesion	6.16E-27	IFNG/IDO1/IL21/PLA2G2D/CCL19/CD80/CD274/PDCD1/TNFRSF13C/ SIRPG/S100A8/CTLA4/CCL5/IL2RA/TIGIT/ICOS/LILRB4/TBX21/CEACAM1/ XCL1/TNF/CD40LG/CD6/CD3E/RASGRP1/LCK/LAG3/SIRPB1/IL10/ CARD11/IRF1/EBI3/CD83/SOCS1/SKAP1/GATA3/TNFSF14/CD5/PDCD1LG2/ TESPA1/ZAP70/IL12RB1/PTPRC/IL18/CD160/ITGAL/SEMA4D/ICAM1/ RUNX3/PRKCQ/RIPOR2/GRAP2/CORO1A/SASH3/HLA-DPA1/PTPN6/CD74	57
GO:0002449	lymphocyte mediated immunity	1.19E-19	IL21/AICDA/CR2/KIR2DL4/CD19/LTA/NCR3/SH2D1A/BATF/TBX21/ CEACAM1/GZMB/XCL1/TNF/CD226/CD40LG/TRBC2/CD8A/RASGRP1/ LAG3/IL10/SLA2/TRBC1/GATA3/PRF1/CD96/NCR1/GZMM/CLEC12B/ IL12RB1/PTPRC/IL18/CD160/C1QB/ICAM1/FCER1G/C1QC/CORO1A/P2RX7/ C1QA/SERPINB9/SASH3/CTSH/C3/PTPN6/CD74/FZD5/HLA-H/CD40	49
GO:0001819	positive regulation of cytokine production	3.44E-18	IFNG/IDO1/IL21/CLEC6A/CCL19/GBP5/CD80/CLEC4E/CD274/KIR2DL4/ TNFRSF13C/SIGLEC16/LTA/TIGIT/CD2/PYHIN1/XCL1/MCOLN2/TNF/CD226/	54

			CD40LG/CD6/CD3E/RASGRP1/IL10/CARD11/IRF1/EBI3/CD83/BIRC3/ GATA3/LTB/STAT1/HTR2B/IL12RB1/PTPRC/IL18/CD160/TLR7/CD244/ TLR1/FCER1G/PRKCQ/P2RX7/SASH3/FFAR2/HLA-DPA1/C3/HPSE/CD74/ FZD5/CD40/LRRK2/MNDA	
GO:0050851	antigen receptor-mediated signaling pathway	1.27E-17	FCRL3/PAX5/MS4A1/CD22/CD19/GBP1/BLK/CTLA4/TRBV7-9/THEMIS/ITK/ LILRB4/TRAT1/CEACAM1/STAP1/CD3D/UBASH3A/CD247/CD226/TRBC2/ CD3E/TRAC/LCK/CARD11/SLA2/TRBC1/CD3G/SKAP1/GATA3/HLA-DQA1/ TESPA1/ZAP70/LCP2/PTPRC/CD160/PRKCQ/GRAP2/HLA-DRA/PSMB9/ HLA-DPA1/PTPN6/LPXN/PIK3CD/MNDA	44
GO:0050900	leukocyte migration	1.46E-12	CXCL11/CXCL9/CXCL10/CCL19/CXCL13/CCL17/CXCR5/VPREB3/GPR15/ CNR2/SIRPG/S100A8/CCL4/CCL5/TBX21/CEACAM1/GPR18/CXCR3/STAP1/ CD2/CCL8/P2RY12/XCL1/MCOLN2/TNF/LCK/IL10/CD48/GATA3/TNFSF14/ XCL2/ZAP70/SDC3/CD244/ITGAL/ICAM1/FUT7/CD84/FCER1G/RIPOR2/ TRPM2/CORO1A/FFAR2/PTPN6/CD74/PIK3CD/CSF3R	47
GO:0050663	cytokine secretion	1.59E-09	IFNG/CLEC6A/CCL19/GBP5/CLEC4E/NOS2/CD274/GBP1/CD2/MCOLN2/ TNF/RASGRP1/IL10/CARD11/SOCS1/GATA3/HTR2B/LCP2/CD160/ CD244/TLR1/TNFAIP3/P2RX7/FFAR2/BANK1/ABCA1/FZD5/LRRK2	28
GO:0002544	chronic inflammatory response	3.71E-07	IDO1/CXCL13/LTA/S100A8/CCL5/TNF/IL10/TNFAIP3	8
GO:0007204	positive regulation of cytosolic calcium ion concentration	6.56E-07	CXCL11/CXCL9/CXCL10/CCL19/CXCL13/CXCR5/DRD1/MS4A1/CCR8/ CD19/GPR18/CXCR3/FASLG/XCL1/GPR174/MCOLN2/LCK/GLP1R/CD52/ GPR65/HTR2B/PLCH2/PTPRC/TRPM2/CORO1A/P2RX7/PTPN6/S1PR4	28
GO:0071887	leukocyte apoptotic process	1.09E-05	IDO1/CCL19/CD274/PDCD1/CCL5/IL2RA/FASLG/IL10/CD3G/FCER1G/ PRKCQ/P2RX7/CD74/PIK3CD	14
GO:0001776	leukocyte homeostasis	4.16E-05	TNFRSF13B/TNFRSF13C/IL2RA/GPR174/SIT1/TNFSF14/TNFAIP3/ FCER1G/CORO1A/P2RX7/CD74/PIK3CD	12
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	5.25E-05	IFNG/IL21/IL31RA/CD80/DOK7/CCL5/FCGR1A/STAP1/TNF/CD3E/NCF1/ PTPRC/IL18/SEMA4D/ICAM1/BANK1/CD74/CD40	18
GO:0016444	somatic cell DNA recombination	0.000426	AICDA/BATF/TBX21/CD40LG/IL10/BCL11B/PTPRC/TCF7/CD40	9
GO:0048525	negative regulation of viral process	0.000845	AICDA/CCL4/CCL5/CCL8/LTF/OASL/TNF/APOBEC3H/TRIM14/STAT1/ APOBEC3G	11
GO:0098883	synapse pruning	0.001557	C1QB/C1QC/C1QA/C3	4
GO:0033623	regulation of integrin activation	0.0029	CXCL13/P2RY12/SKAP1/PLEK	4
GO:0045730	respiratory burst	0.003716	NCF1/CD52/NCF1B/NCF1C/MPO/PIK3CD	6

GO:0051341	regulation of oxidoreductase activity	0.004827	IFNG/HP/CNR2/GCH1/GZMA/TNF/HTR2B/APOE/GFI1/LRRK2	10
GO:0006801	superoxide metabolic process	0.00581	NOS2/GCH1/TNF/NCF1/NCF1B/NCF1C/MPO/ACP5	8
GO:0010543	regulation of platelet activation	0.010062	CEACAM1/PLEK/APOE/FCER1G/PRKCQ	5
GO:0045348	positive regulation of MHC class II biosynthetic process	0.015561	IFNG/IL10/CIITA	3
GO:1901739	regulation of myoblast fusion	0.015986	CXCL9/CXCL10/TNFSF14/RIPOR2	4
GO:2000810	regulation of bicellular tight junction assembly	0.015986	OCLN/TNF/GPBAR1/FZD5	4
GO:0031529	ruffle organization	0.018215	SNX10/STAP1/P2RY12/PLEK/CARMIL2/ICAM1	6
GO:0043547	positive regulation of GTPase activity	0.018215	CCL19/CXCL13/CCL17/CCL4/CCL5/CCL8/XCL1/RASGRP1/XCL2/ALDH1A1/ ADRB1/GPR65/HTR2B/SEMA4D/ICAM1/ARAP2/TBC1D10C/ARHGAP25/ ACAP1/CD40/LRRK2	21
GO:0051043	regulation of membrane protein ectodomain proteolysis	0.021254	IFNG/TNF/IL10/APOE	4
GO:0006216	cytidine catabolic process	0.024465	AICDA/APOBEC3H/APOBEC3G	3
GO:0009972	cytidine deamination	0.024465	AICDA/APOBEC3H/APOBEC3G	3
GO:0033700	phospholipid efflux	0.024465	APOC1/APOE/ABCA1	3
GO:0042368	vitamin D biosynthetic process	0.024465	IFNG/TNF/GFI1	3
GO:0046087	cytidine metabolic process	0.024465	AICDA/APOBEC3H/APOBEC3G	3
GO:0006809	nitric oxide biosynthetic process	0.028571	IFNG/NOS2/GCH1/TNF/IL10/ICAM1/ACP5	7
GO:2000116	regulation of cysteine-type endopeptidase activity	0.029533	S100A8/FASLG/CASP5/LTF/CST7/TNF/LCK/NAIP/BIRC3/TNFSF14/ TP63/PSMB9/SERPINB9/CTSH	14

GO:0001845	phagolysosome assembly	0.029533	RAB39A/CORO1A/P2RX7	3
GO:0030656	regulation of vitamin metabolic process	0.029533	IFNG/TNF/GFI1	3
GO:0051851	modulation by host of symbiont process	0.032278	CCL4/CCL5/CCL8/LTF/APOE/IFI27	6
GO:0045834	positive regulation of lipid metabolic process	0.034091	IFNG/CCL19/CD19/P2RY12/TNF/APOC1/HTR2B/APOE/NR1H3/CD74	10
GO:0014002	astrocyte development	0.036728	IFNG/DRD1/S100A8/TNF/C1QA	5
GO:0032352	positive regulation of hormone metabolic process	0.041036	IFNG/TNF/GATA3	3
GO:2001057	reactive nitrogen species metabolic process	0.042565	IFNG/NOS2/GCH1/TNF/IL10/ICAM1/ACP5	7
GO:0010875	positive regulation of cholesterol efflux	0.048039	APOE/NR1H3/ABCA1	3

Table 9.11: GO biological processes associated with DEGs upregulated in non-responders. DEGs from unadjusted p<0.05, absolute LFC>1.

GO ID	Description	padj	geneID	Count
GO:0001503	ossification	6.37E-08	CCN2/ROR2/ENPP1/CLEC11A/COL1A2/WNT3/COL13A1/GJA1/SOX8/ SMOC1/BMP8B/COL1A1/ASPN/CREB3L1/SCX/BMP8A/IGSF10/COMP/ HAND2/FGF9/SOX11/TAC1/DMP1/PENK/COL2A1/CLEC3A	26
GO:0030199	collagen fibril organization	2.88E-06	P3H4/COL1A2/COL12A1/FMOD/COL1A1/SCX/COMP/MMP11/ACAN/COL2A1	10
GO:0048705	skeletal system morphogenesis	0.000385	CCN2/ROR2/IRX5/COL7A1/COL12A1/COL13A1/COL1A1/SCX/COMP/ HOXD10/HHIP/ACAN/CHAD/SOX11/COL2A1	15
GO:0051216	cartilage development	0.000385	ADAMTS12/CCN2/ROR2/COL7A1/COL12A1/BMP8B/COL1A1/SCX/BMP8A/ COMP/HAND2/ACAN/FGF9/COL2A1	14
GO:0018146	keratan sulfate biosynthetic process	0.000545	FMOD/B3GNT4/OGN/KERA/CHST6/ACAN	6
GO:0050769	positive regulation of neurogenesis	0.001063	SLIT2/ITPKA/WNT3/PTK7/RET/PTPRD/SOX8/PROX1/CNTN1/CHODL/ZNF365/ IRX3/VLDLR/NEFL/SERPINE2/DRD2/NKX2-5/PAX6/SOX11/SYT4	20

GO:0055001	muscle cell development	0.001239	FHOD3/UCHL1/RGS4/PROX1/MYOM3/ALPK2/COMP/MEGF10/TNNT1/MYF6/NKX2-5/DNER	12
GO:0060562	epithelial tube morphogenesis	0.003598	SLIT2/ADAMTS12/LAMA1/PTK7/GJA1/RET/IRX2/SOX8/PROX1/IRX3/KIF26B/HAND2/HHIP/NKX2-5/SOX11	15
GO:0045926	negative regulation of growth	0.003689	SLIT2/ENPP1/WNT3/GJA1/MT1E/MT1A/SEMA3D/RGS4/KCNK2/MT1M/TBX5/SERPINE2/MT1X	13
GO:0010273	detoxification of copper ion	0.00413	MT1E/MT1A/MT1M/MT1X	4
GO:1990169	stress response to copper ion	0.00413	MT1E/MT1A/MT1M/MT1X	4
GO:0044272	sulfur compound biosynthetic process	0.00524	CHPF/CDO1/FMOD/ELOVL4/UST/B3GNT4/OGN/GSTO2/KERA/CHST6/ACAN	11
GO:0007409	axonogenesis	0.005336	SLIT2/DOK5/LAMA1/EPHB3/WNT3/UCHL1/RET/SEMA3D/ANK3/CHODL/UST/EFNB3/VLDLR/NEFL/NRTN/DRD2/VSTM2L/PAX6	18
GO:0015872	dopamine transport	0.005348	MAPK15/SYT5/DRD2/SYT4/SYT13/SLC6A3	6
GO:0003333	amino acid transmembrane transport	0.008277	LRRC8E/RGS4/SLC7A9/SLC7A4/SLC3A1/SLC38A3/SLC6A20	7
GO:0032963	collagen metabolic process	0.009535	CCN2/P3H4/COL1A2/COL13A1/COL1A1/CREB3L1/SCX/MMP11	8
GO:0045109	intermediate filament organization	0.012222	KRT17/NEFM/NEFL/DES	4
GO:0098743	cell aggregation	0.012222	CCN2/ROR2/ACAN/COL2A1	4
GO:0051965	positive regulation of synapse assembly	0.015207	THBS2/EPHB3/PTPRD/LRRN1/SYNDIG1/LINGO2	6
GO:0042271	susceptibility to natural killer cell mediated cytotoxicity	0.016662	RAET1E/ULBP2/ULBP1	3
GO:0071772	response to BMP	0.017887	ADAMTS12/ROR2/BMP8B/SCX/BMP8A/COMP/NKX2-5/SOX11/COL2A1	9
GO:0071773	cellular response to BMP stimulus	0.017887	ADAMTS12/ROR2/BMP8B/SCX/BMP8A/COMP/NKX2-5/SOX11/COL2A1	9
GO:0050905	neuromuscular process	0.019651	UCHL1/NEFL/DRD2/COMP/HOXD10/PENK/SLC6A3	7
GO:0019896	axonal transport of mitochondrion	0.019651	UCHL1/NEFL/MGARP	3
GO:0070208	protein heterotrimerization	0.019651	COL1A2/C1QTNF6/COL1A1	3

GO:0014855	striated muscle cell proliferation	0.022137	GJA1/KCNK2/TBX5/MEGF10/NKX2-5/FGF9	6
GO:0043200	response to amino acid	0.022269	CCN2/COL16A1/COL1A2/CDO1/CPEB1/COL1A1/F7	7
GO:0044273	sulfur compound catabolic process	0.022269	CDO1/FMOD/OGN/KERA/ACAN	5
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	0.022269	NREP/ROR2/COL1A2/FMOD/BMP8B/ASPN/SCX/BMP8A/CILP/COMP/ NKX2-5/FGF9/SOX11	13
GO:0002028	regulation of sodium ion transport	0.022876	ANK3/CNTN1/SERPINE2/DRD2/NKX2-5/NKAIN4	6
GO:0006882	cellular zinc ion homeostasis	0.027655	MT1E/MT1A/MT1M/MT1X	4
GO:0060047	heart contraction	0.029287	CCN2/IRX5/GJA1/RGS4/HBEGF/APLN/DRD2/DES/TNNT1/NKX2-5/TAC1	11
GO:0007589	body fluid secretion	0.03015	P2RY2/GJA1/CDO1/APLN/TAC1/SLC6A3	6
GO:0110148	biomineralization	0.030231	ROR2/ENPP1/COL1A2/GJA1/COL1A1/ASPN/COMP/DMP1	8
GO:1902742	apoptotic process involved in development	0.030853	SLIT2/MEGF10/HAND2/NKX2-5	4
GO:0072210	metanephric nephron development	0.03417	RET/IRX2/SOX8/KIF26B	4
GO:0030010	establishment of cell polarity	0.036568	GPSM2/LAMA1/PTK7/UST/ALPK2/KIF26B/PAX6	7
GO:0009713	catechol-containing compound biosynthetic process	0.040345	HAND2/PNMT/SLC6A3	3
GO:0042423	catecholamine biosynthetic process	0.040345	HAND2/PNMT/SLC6A3	3
GO:0071467	cellular response to pH	0.048523	GJA1/INSRR/SLC38A3	3
GO:0050803	regulation of synapse structure or activity	0.048523	ITPKA/THBS2/EPHB3/PTPRD/LRRN1/DRD2/NGEF/SYNDIG1/LINGO2	9

9.5.3 Processes associated with sdRA & resolving signatures

Table 9.12: GO biological processes associated with DEGs upregulated in sdRA compared to normal. DEGs from padj<0.05, absolute LFC>1.

GO ID	Description	padj	geneID	Count
GO:0042110	T cell activation	5.06E-31	CTPS1/MICB/RELB/BCL3/ICAM1/RUNX1/FANCA/RASAL3/CORO1A/GATA3/PTPRC/ TNFRSF4/TCF7/MAD1L1/LEF1/TNFSF14/F2RL1/BATF/TFRC/ITGAL/CD6/CD2/ SIT1/RHOH/SLA2/ZAP70/CCL2/CAMK4/LCK/CD7/EBI3/CD8B/NLRC3/JAK3/ TREML2/SLAMF6/EGR1/CD3D/CD3E/CD5/CD3G/EOMES/CCL5/TNFSF11/ITK/ EGR3/CCR7/THEMIS/CD8A/GPR18/RASGRP1/ZNF683/ADAM8/BCL11B/BTLA/ IL7R/CLEC4E/TIGIT/FOXP3/CD80/ICOS/PDCD1/IL1B/IL2RA/CR1/IDO1/IFNG/ IL6/PLA2G2D/IL21	70
GO:0030595	leukocyte chemotaxis	1.48E-22	CORO1A/THBS1/TNFSF14/CXCR4/F2RL1/ADGRE2/CCL2/FFAR2/S100A9/CCL5/ TNFSF11/CXCR3/CCR7/CCL8/GPR18/SERPINE1/ADAM8/CCL4/CXCR2/PF4/ CH25H/CXCL2/CCL3/FPR2/CXCR1/PPBP/CXCL3/SLAMF8/CXCL10/CXCL1/IL1B/ CCL3L1/CXCL11/S100A12/CXCL6/CXCL5/PF4V1/CCL18/IL6/CXCL8/CXCL9/ CCL7/CXCL13	43
GO:0060326	cell chemotaxis	1.48E-22	CORO1A/SMOC2/THBS1/LEF1/TNFSF14/CXCR4/F2RL1/ADGRE2/CCR4/CCL2/ FFAR2/S100A9/CCL5/TNFSF11/EGR3/CXCR3/CCR7/CCL8/GPR18/SERPINE1/ ADAM8/CCL4/CXCR2/CXCR6/PF4/CH25H/NR4A1/CXCL2/CCL3/FPR2/CXCR1/ PPBP/CXCL3/SLAMF8/CXCL10/CXCL1/IL1B/CCL3L1/CXCL11/S100A12/CXCL6/ CXCL5/PF4V1/CCL18/IL6/CXCL8/CXCL9/CCL7/CXCL13	49
GO:0070098	chemokine-mediated signaling pathway	6.04E-22	HIF1A/CXCR4/CCR4/CCL2/CCL5/CXCR3/CCR7/CCL8/CCL4/CXCR2/CXCR6/PF4/ CXCL2/CCL3/CXCR1/PPBP/CXCL3/CXCL10/CXCL1/CCL3L1/CXCL11/CXCL6/ CXCL5/PF4V1/CCL18/CXCL8/CXCL9/CCL7/CXCL13	29
GO:0007159	leukocyte cell-cell adhesion	2.94E-16	ICAM1/RUNX1/RASAL3/CORO1A/GATA3/SEMA4D/PTPRC/ITGA4/MAD1L1/ ITGB7/TNFSF14/TFRC/ITGAL/CD6/SELL/ZAP70/CCL2/LCK/EBI3/JAK3/CD3E/ S100A9/CD5/CCL5/TNFSF11/EGR3/CCR7/RASGRP1/ADAM8/BTLA/IL7R/ TIGIT/FOXP3/CD80/ICOS/PDCD1/IL1B/IL2RA/IDO1/IFNG/IL6/PLA2G2D/IL21	43
GO:0030198	extracellular matrix organization	1.27E-15	BCL3/ICAM1/COL4A1/TIMP1/COL6A3/ADAMTS9/PXDN/ADAMTS2/TNC/ COL5A2/SMOC2/MMP16/LOXL1/MMP17/LOXL2/ITGA4/THBS1/ITGB7/ITGAX/ COL3A1/SPOCK2/COL24A1/ITGAL/ADAM19/COL5A1/CRISPLD2/CARMIL2/ POSTN/SERPINE1/ADAM8/COL1A1/MMP25/ADAMTS4/CCN1/ADAM12/ COL11A1/EGFL6/ACAN/ADAMTS14/MMP3/IL6/MMP1/IBSP/MMP13	44
GO:0032496	response to lipopolysaccharide	3.16E-12	SBNO2/TRIB1/ICAM1/ZFP36/JUNB/CD14/LOXL1/NOCT/CD96/CD6/ALPL/CCL2/ CCL5/CCR7/PTGS2/SERPINE1/GJB2/PF4/CXCL2/CD80/CCL3/FOS/PPBP/CXCL3/ CXCL10/CXCL1/IL1B/CXCL11/GJB6/CXCL6/IDO1/CXCL5/PF4V1/IL6/CXCL8/	37

			CXCL9/CXCL13	
GO:0050727	regulation of inflammatory response	2.84E-11	SBNO2/SEMA7A/FANCA/GATA3/LDLR/MEFV/CCN4/PTPRC/LILRA5/CST7/CD6/ NLRC3/FFAR2/SOCS3/S100A9/SIGLEC10/CCL5/TNFSF11/CCR7/PTGS2/ SERPINE1/ADAM8/TNFAIP6/GBP5/FOXP3/CCL3/FPR2/OSM/SLAMF8/IL1B/ CCL3L1/IL2RA/S100A12/IDO1/CCL18/MMP3/IL6/IL21	38
GO:0042119	neutrophil activation	1.81E-10	TMC6/PLAU/ARHGAP9/CD14/HSPA6/LILRB3/SLC2A3/PTPRC/SLC2A5/ITGAX/ TBC1D10C/QPCT/F2RL1/ITGAL/SELL/CRISPLD2/SERPINA1/S100A9/CCL5/ ADAM8/LRG1/MMP25/TNFAIP6/CEACAM3/FPR1/CXCR2/GPR84/FCN1/FPR2/ CXCR1/ADGRE3/PPBP/ADGRG3/FCAR/MGAM/CXCL1/S100A12/FCGR3B/ CXCL6/CR1/HBB/CXCL8/S100P	43
GO:0002446	neutrophil mediated immunity	7.46E-10	TMC6/PLAU/ARHGAP9/CD14/HSPA6/LILRB3/SLC2A3/PTPRC/SLC2A5/ITGAX/ TBC1D10C/QPCT/F2RL1/ITGAL/SELL/CRISPLD2/SERPINA1/S100A9/ADAM8/ LRG1/MMP25/TNFAIP6/CEACAM3/FPR1/CXCR2/GPR84/FCN1/FPR2/CXCR1/ ADGRE3/PPBP/ADGRG3/FCAR/MGAM/CXCL1/S100A12/FCGR3B/CXCL6/CR1/ HBB/IL6/S100P	42
GO:0001819	positive regulation of cytokine production	4.02E-09	IL27RA/BCL3/SEMA7A/RUNX1/HIF1A/CD14/GATA3/PTPRC/THBS1/LILRA5/ F2RL1/CD6/CD2/EBI3/FFAR2/SLAMF6/EGR1/CD3E/POSTN/CCR7/PTGS2/ SERPINE1/RASGRP1/ADAM8/CLEC4E/TIGIT/GBP5/FOXP3/PF4/CD80/FCN1/ CCL3/OSM/AIM2/IL1B/IDO1/IFNG/IL6/IL21	39
GO:0002429	immune response-activating cell surface receptor signaling pathway	1.82E-06	MICB/RELB/RUNX1/CYFIP2/GATA3/PTPRC/MUC3A/CD247/HLA-DQA1/TRAC/ SLA2/ZAP70/LCK/UBASH3A/TRAT1/FFAR2/CD3D/CD3E/CD3G/TRBC2/ITK/ CCR7/THEMIS/TRBC1/CLEC4E/FPR1/FCGR1A/FOXP3/FCN1/FPR2/FCRL3/CD19/ CR1/MUC16	34
GO:0002757	immune response-activating signal transduction	1.82E-06	MICB/RELB/RUNX1/CYFIP2/GATA3/PTPRC/MUC3A/CD247/HLA-DQA1/TRAC/ SLA2/ZAP70/LCK/UBASH3A/TRAT1/FFAR2/CD3D/CD3E/CD3G/TRBC2/ITK/CCR7/ THEMIS/TRBC1/CLEC4E/FPR1/FCGR1A/FOXP3/FCN1/FPR2/FCRL3/CD19/CR1/ MUC16	34
GO:0050663	cytokine secretion	1.91E-06	IL27RA/CD14/SRGN/GATA3/CCN4/TNFRSF4/LILRA5/F2RL1/CD2/FFAR2/POSTN/ CCR7/RASGRP1/CLEC4E/GBP5/FOXP3/FCN1/CCL3/OSM/AIM2/IL1B/S100A12/IFNG	23
GO:0007059	chromosome segregation	1.98E-06	NDC80/CENPN/ECT2/KIF23/CDC6/KIF4A/MAD1L1/HASPIN/KIFC1/HJURP/KNL1/ SKA1/CENPF/BUB1B/AURKB/SKA3/BIRC5/TOP2A/NCAPG/BUB1/SPC25/KIF18B/ SGO1/KIF14/MKI67/FAM83D/DLGAP5	27
GO:0007204	positive regulation of cytosolic calcium ion concentration	1.83E-05	CORO1A/S1PR4/PTPRC/CXCR4/F2RL1/CCR4/LCK/CXCR3/CCR7/GPR18/FPR1/ CXCR2/CXCR6/P2RX5/CCL3/FPR2/CXCR1/CD19/MCHR1/CXCL10/CXCL11/ TRPA1/PROK2/CXCL9/CXCL13	25
GO:0015671	oxygen transport	3.30E-05	IPCEF1/HBA2/HBB/HBA1/HBD/HBM	6

GO:0048872	homeostasis of number of cells	3.53E-05	ZFP36/HIF1A/PMAIP1/CORO1A/GATA3/EZH2/TNFSF14/INHBA/CCNB2/SIT1/ CCR4/JAK3/CCR7/IL7R/CXCR2/FOXP3/IL2RA/CXCL6/KLF1/ALAS2/IL6	21
GO:0009612	response to mechanical stimulus	4.68E-05	ANGPT2/JUNB/TNC/PIEZO2/THBS1/COL3A1/TNFSF14/CXCR4/FOSL1/POSTN/ PTGS2/COL1A1/CASP5/COL11A1/FOS/CXCL10/IL1B/FOSB/TRPA1	19
GO:0071887	leukocyte apoptotic process	5.94E-05	HIF1A/JAK3/AURKB/CD3G/CCL5/CCR7/ADAM8/IL7R/CXCR2/PDCD1/IL2RA/ IDO1/IL6	13
GO:0001503	ossification	8.55E-05	ECM1/SBNO2/SEMA7A/HIF1A/SNAI1/JUNB/SRGN/TNC/COL5A2/SEMA4D/ MMP16/NOCT/CCN4/LEF1/MSX2/ALPL/TNFSF11/PTGS2/COL1A1/PHOSPHO1/ CCN1/COL11A1/CCL3/EGR2/IL6/IBSP/MMP13	27
GO:0043547	positive regulation of GTPase activity	0.000111	ICAM1/ARAP2/TAGAP/RGS2/ECT2/ARHGAP9/RASAL3/RGS16/RGS18/ SEMA4D/EZH2/DEPDC1B/TBC1D10C/F2RL1/CCL2/ACAP1/CCL5/CCR7/CCL8/ RASGRP1/CCL4/EPHA1/CCL3/CCL3L1/CCL18/CCL7/CXCL13	27
GO:0110148	biomineralization	0.000329	ECM1/SBNO2/SLC20A1/HIF1A/SRGN/MSX2/ALPL/PTGS2/COL1A1/PHOSPHO1/ CCN1/CCL3/IBSP/MMP13/AMTN	15
GO:0030574	collagen catabolic process	0.000401	ADAMTS2/MMP16/MMP17/MMP25/ADAMTS14/MMP3/MMP1/MMP13	8
GO:0032352	positive regulation of hormone metabolic process	0.000462	HIF1A/GATA3/EGR1/IL1B/IFNG	5
GO:0019058	viral life cycle	0.000561	ICAM1/KPNA2/LDLR/TNFRSF4/ITGB7/CXCR4/TFRC/APOBEC3H/NUP210/SLAMF1/ CCL2/TOP2A/CCL5/CCL8/LAMP3/CXCR6/CD80/FCN1/LRRC15/APOBEC3A/CR1/ CXCL8	22
GO:0046677	response to antibiotic	0.001364	ICAM1/RGS2/ECT2/CD14/TNC/GATA3/EZH2/ALPL/CCR4/FOSL1/ EGR1/TYMS/COL1A1/GJB2/HBA2/NEFL/HBB/HBA1/TRPA1/IL6/CCL7	21
GO:0072593	reactive oxygen species metabolic process	0.001671	ICAM1/HIF1A/PMAIP1/PXDN/THBS1/F2RL1/FOXM1/PTGS2/CCN1/ FPR2/IL1B/CCN6/HBA2/HBB/MMP3/HBA1/HBD/IFNG/HBM	19
GO:0032570	response to progesterone	0.001836	THBS1/FOSL1/TYMS/GJB2/FOS/FOSB/CSN1S1	7
GO:0061430	bone trabecula morphogenesis	0.003107	SBNO2/SEMA4D/MSX2/COL1A1	4
GO:0032461	positive regulation of protein oligomerization	0.004229	PMAIP1/AIM2/TCL1A/MMP3/MMP1	5
GO:0034501	protein localization to kinetochore	0.007616	NDC80/HASPIN/KNL1/AURKB	4
GO:0035987	endodermal cell differentiation	0.00875	COL5A2/ITGA4/INHBA/COL5A1/EOMES/COL11A1	6

GO:2000116	regulation of cysteine-type endopeptidase activity	0.010469	CYFIP2/PMAIP1/MEFV/THBS1/LEF1/TNFSF14/CST7/LCK/S100A9/BIRC5/ PTGS2/LAMP3/CCN1/CASP5/AIM2	15
GO:0007565	female pregnancy	0.010908	ANGPT2/RGS2/TIMP1/JUNB/NAMPT/FOSL1/PTGS2/GJB2/FOS/IL1B/FOSB/ KLF1/IDO1	13
GO:0033280	response to vitamin D	0.01862	TNC/PIM1/ALPL/PTGS2/CXCL10	5
GO:0042368	vitamin D biosynthetic process	0.023083	SNAI1/IL1B/IFNG	3
GO:0030656	regulation of vitamin metabolic process	0.02791	SNAI1/IL1B/IFNG	3
GO:2000191	regulation of fatty acid transport	0.030996	THBS1/TNFSF11/ERFE/IL1B	4
GO:0008360	regulation of cell shape	0.032317	ATP10A/ICAM1/CORO1A/SEMA4D/BAMBI/PLXNC1/RHOH/CCL2/CCL3/CCL7	10
GO:0018146	keratan sulfate biosynthetic process	0.033715	B3GNT3/ACAN/KERA/CHST6	4
GO:0007596	blood coagulation	0.033731	PAPSS2/PLAU/GATA3/THBS1/COL3A1/F2RL1/LCK/SERPINA1/ SERPINE1/COL1A1/H3C10/PF4/P2RX5/HBB/PF4V1/HBD/IL6	17
GO:0050974	detection of mechanical stimulus involved in sensory perception	0.037376	PIEZO2/CXCR4/COL11A1/TRPA1	4
GO:0051770	positive regulation of nitric-oxide synthase biosynthetic process	0.037376	NAMPT/CCL2/IFNG	3
GO:0000302	response to reactive oxygen species	0.038021	PLK3/ECT2/EZH2/FOSL1/CCR7/COL1A1/FOS/HBA2/HBB/MMP3/HBA1/TRPA1/IL6	13
GO:0007623	circadian rhythm	0.040803	RELB/BHLHE40/EZH2/NOCT/NAMPT/EGR1/TYMS/TOP2A/EGR3/SERPINE1/ RPE65/PROK2	12
GO:0110150	negative regulation of biomineralization	0.040845	ECM1/HIF1A/SRGN/CCL3	4
GO:0044319	wound healing, spreading of cells	0.044878	MSX2/COL5A1/CARMIL2/CCN1	4
GO:0090505	epiboly involved in wound healing	0.044878	MSX2/COL5A1/CARMIL2/CCN1	4

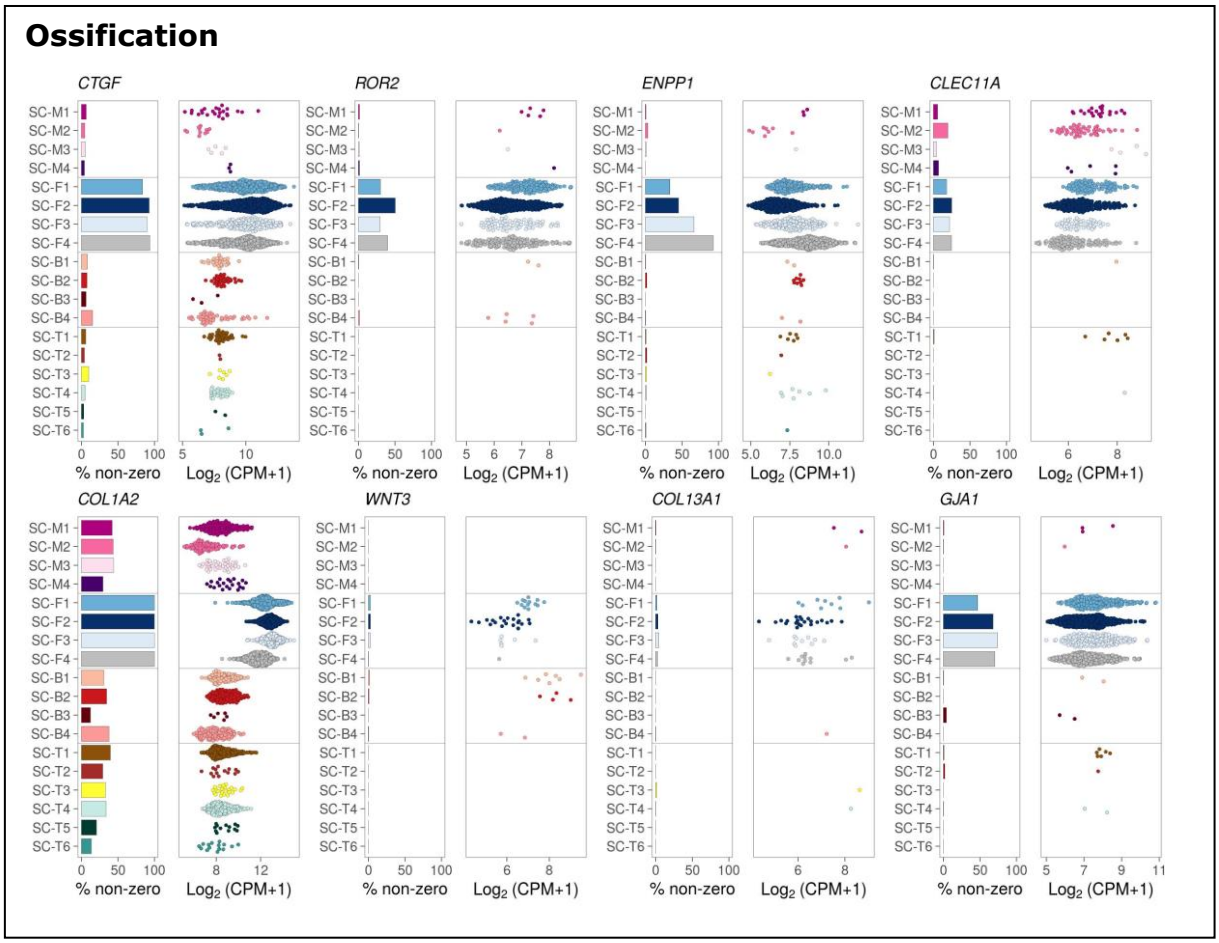
Table 9.13: GO biological processes associated with DEGs upregulated in Res compared to normal. DEGs from padj<0.05, absolute LFC>1.

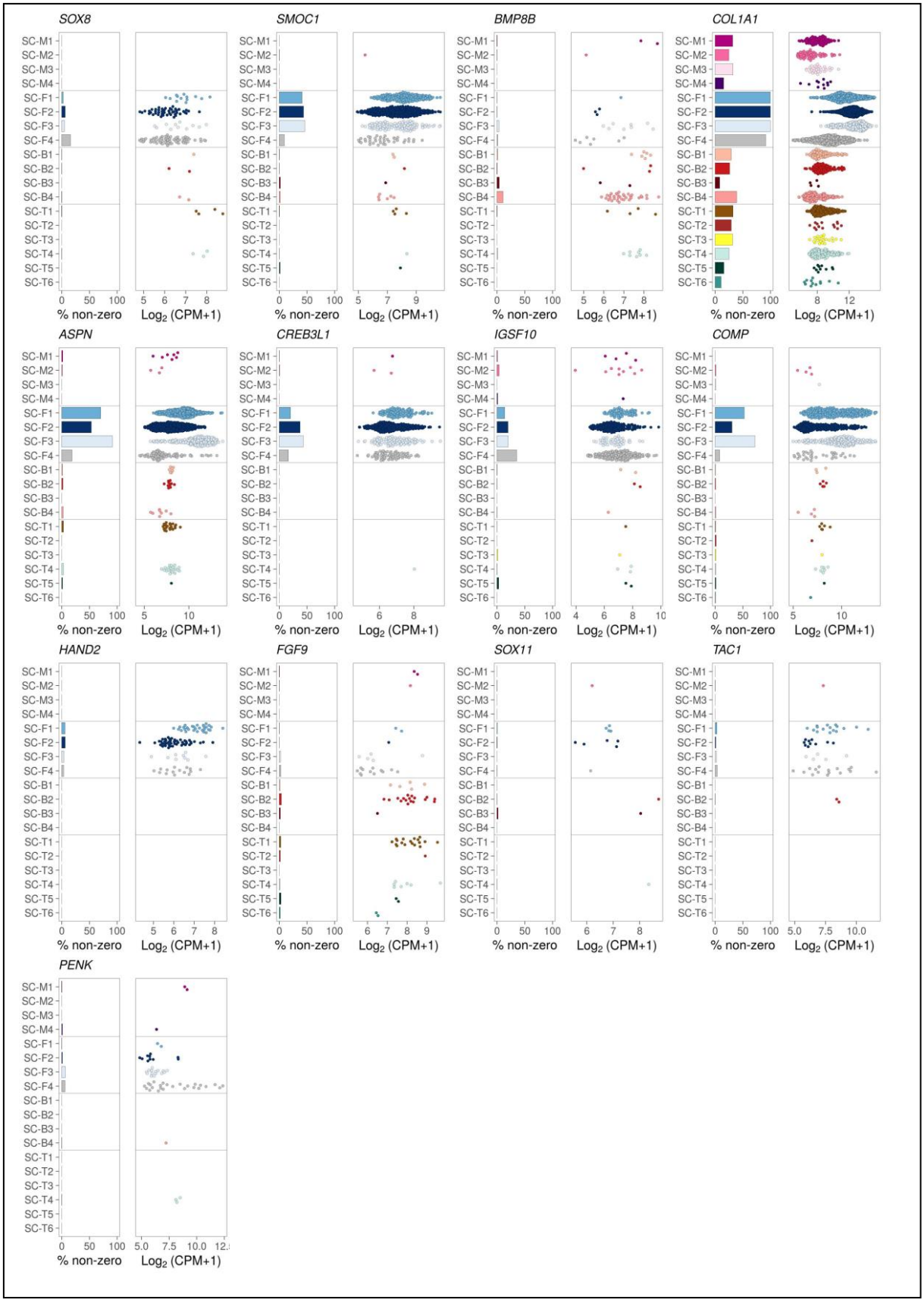
GO ID	Description	p.adjust	geneID	Count
GO:0030595	leukocyte chemotaxis	3.80E-13	DYSF/CCL2/FFAR2/SERPINE1/CXCL2/CXCL3/ADAM8/IL1B/FPR2/ S100A12/CXCR2/SLAMF8/CXCR1/CXCL8/CXCL10/IL6/CXCL9	17
GO:0042119	neutrophil activation	2.00E-11	NBEAL2/SLC2A3/QPCT/SLC2A5/FPR1/TNFAIP6/SERPINA1/ADAM8/ CEACAM3/MMP25/FPR2/S100A12/CXCR2/FCN1/CXCR1/MGAM/ ADGRG3/FCGR3B/CXCL8/S100P	20
GO:0002446	neutrophil mediated immunity	2.00E-11	NBEAL2/SLC2A3/QPCT/SLC2A5/FPR1/TNFAIP6/SERPINA1/ADAM8/ CEACAM3/MMP25/FPR2/S100A12/CXCR2/FCN1/CXCR1/MGAM/ ADGRG3/FCGR3B/IL6/S100P	20
GO:0070098	chemokine-mediated signaling pathway	9.18E-07	CCL2/CXCL2/CXCL3/CXCR2/CXCR1/CXCL8/CXCL10/CXCL9	8
GO:0031349	positive regulation of defense response	2.11E-06	LDLR/MUC3A/FFAR2/SERPINE1/ADAM8/IL1B/CLEC4E/FPR2/GBP5/ S100A12/FCN1/IL6/CLEC6A	13
GO:0071222	cellular response to lipopolysaccharide	3.31E-05	CCL2/SERPINE1/CXCL2/CXCL3/IL1B/CXCL8/CXCL10/IL6/CXCL9	9
GO:0061844	antimicrobial humoral immune response mediated by antimicrobial peptide	6.98E-05	CXCL2/CXCL3/S100A12/CXCL8/CXCL10/CXCL9	6
GO:0048143	astrocyte activation	0.000339	LDLR/IL1B/FPR2/IL6	4
GO:0050663	cytokine secretion	0.000645	DYSF/FFAR2/IL1B/CLEC4E/GBP5/S100A12/FCN1/CLEC6A	8
GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	0.00096	CCL2/FPR1/ADGRE1/FPR2/ADGRG3/MCHR1/CXCL10/CXCL9	8
GO:0010718	positive regulation of epithelial to mesenchymal transition	0.002963	BAMBI/LOXL2/IL1B/IL6	4
GO:0006898	receptor-mediated endocytosis	0.003394	LDLR/TFRC/SERPINE1/FPR2/CXCR2/CXCR1/HBA1/CXCL8	8
GO:0007204	positive regulation of cytosolic calcium ion concentration	0.003719	FPR1/FPR2/CXCR2/CXCR1/MCHR1/CXCL10/PROK2/CXCL9	8
GO:0051928	positive regulation of calcium ion transport	0.007378	HOMER1/CCL2/MCHR1/CXCL10/CXCL9	5
GO:0019915	lipid storage	0.007595	DYSF/FFAR2/IL1B/IL6	4
GO:0045765	regulation of angiogenesis	0.009886	TNFRSF12A/SERPINE1/IL1B/CXCR2/CXCL8/CXCL10/PROK2/IL6	8
GO:0042542	response to hydrogen peroxide	0.013799	PPIF/ECT2/FOSL1/HBA1/IL6	5

GO:0051090	regulation of DNA-binding transcription factor activity	0.019195	PIM1/BHLHE40/FOSL1/ADAM8/IL1B/S100A12/ADGRG3/IL6	8
GO:0045124	regulation of bone resorption	0.019321	TFRC/ADAM8/IL6	3
GO:0032147	activation of protein kinase activity	0.019392	ECT2/TPX2/CLSPN/FPR1/KIF14/IL1B/PROK2	7
GO:0006509	membrane protein ectodomain proteolysis	0.020808	ADAM19/ADAM8/IL1B	3
GO:0050764	regulation of phagocytosis	0.022843	DYSF/CCL2/IL1B/FPR2	4
GO:0022409	positive regulation of cell-cell adhesion	0.023078	ADAM19/CCL2/TFRC/ADAM8/IL1B/IL6	6
GO:0051983	regulation of chromosome segregation	0.025582	ECT2/MAD1L1/CENPF/MKI67	4
GO:0140014	mitotic nuclear division	0.025945	MAD1L1/TPX2/CENPF/MKI67/KIF14/IL1B	6
GO:0030198	extracellular matrix organization	0.027466	ADAM19/MMP17/LOXL2/SERPINE1/ADAM8/MMP25/IL6	7
GO:0045073	regulation of chemokine biosynthetic process	0.035728	IL1B/IL6	2
GO:0051044	positive regulation of membrane protein ectodomain proteolysis	0.035728	ADAM8/IL1B	2
GO:0051310	metaphase plate congression	0.038082	MAD1L1/CENPF/KIF14	3
GO:0001780	neutrophil homeostasis	0.038082	CXCR2/IL6	2
GO:0051382	kinetochore assembly	0.038082	CENPN/CENPF	2
GO:0007623	circadian rhythm	0.040548	NOCT/BHLHE40/TYMS/SERPINE1/PROK2	5
GO:0044409	entry into host	0.046158	LDLR/TFRC/FCN1/CXCL8	4
GO:0045444	fat cell differentiation	0.047731	DYSF/NOCT/FFAR2/IL6/GDF6	5
GO:0046677	response to antibiotic	0.049397	PPIF/ECT2/FOSL1/TYMS/HBA1/IL6	6
GO:0019058	viral life cycle	0.049545	LDLR/CCL2/TFRC/FCN1/APOBEC3A/CXCL8	6

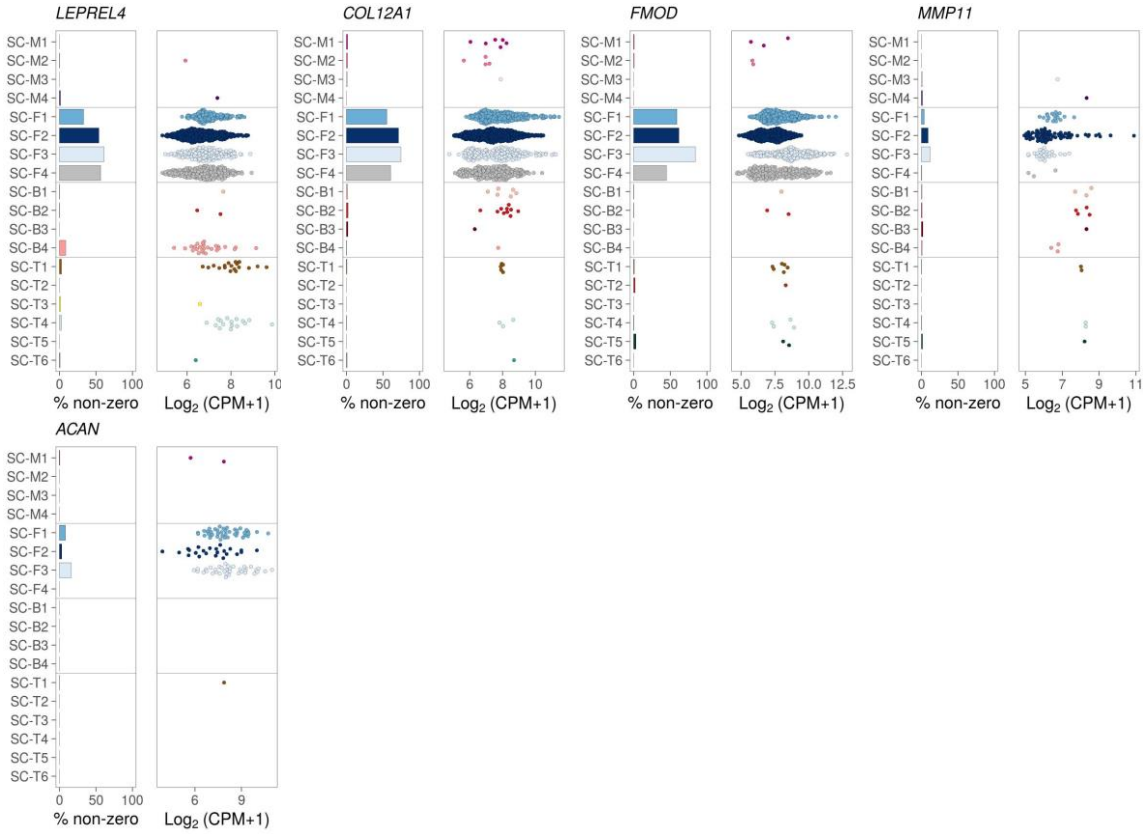
9.6 Exploration of genes involved in GO biological processes associated with non-response

Figure 9.1: Exploration of genes involved in the top three GO BP that are downregulated in EULAR DAS28-ESR responders compared to non-responders at 12 months in the AMP RA Phase I dataset (Zhang et al., 2019). Genes identified as downregulated in responders based on unadj p<0.05, absolute LFC>1 that are involved in ossification, collagen fibril organisation, skeletal system morphogenesis, or cartilage development. Single cell RNA sequencing data of selected cellular populations (monocytes, fibroblasts, B cells & T cells). M1 = IL1B+ pro-inflammatory monocytes, M2 = NUPR1+ monocytes, M3 = C1QA+ monocytes, M4 = IFN-activated monocytes, F1 = CD34+ sublining fibroblasts, F2 = HLA+ sublining fibroblasts, F3 = DKK3+ sublining fibroblasts, F4 = CD55+ lining fibroblasts, B1 = IGHD+ CD270 naive B cells, B2 = IGHG3+ CD27- memory B cells, B3 = autoimmune-associated cells (ABC), B4 = Plasmablasts, T1 = CCR7+ CD4+ T cells, T2 = FOXP3+ Tregs, T3 = PD-1+ Tph/Tfh, T4 = GZMK+ CD8+ T cells, T5 = GNLY+ GZMB+ CTLs, T6 = GZMK+/GZMB+ T cells. Graphs taken directly from https://immunogenomics.io/ampra/ (Zhang et al., 2019).

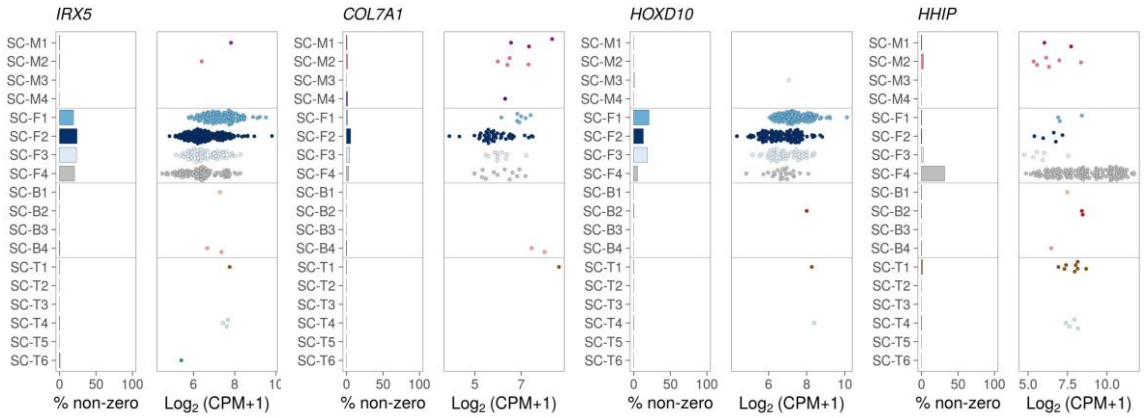



SOX8
SMOC1
BMP8B
COL1A1
ASPN
CREB3L1
IGSF10
COMP
HAND2
FGF9
SOX11
TAC1
PENK
SC-M1
SC-M2
SC-M3
SC-M4
SC-F1
SC-F2
SC-F3
SC-F4
SC-B1
SC-B2
SC-B3
SC-B4
SC-T1
SC-T2
SC-T3
SC-T4
SC-T5
SC-T6
% non-zero
Log2 (CPM+1)

Collagen fibril organization



Skeletal system morphogenesis



9.7 METABOLIC GENE LISTS

Glucose Metabolism

Glycolysis	ALDOA	ALDOB	ALDOC	BPGM	ENO1	ENO2	ENO3	ENO4	GALM	GCK
	GPI	HK2	HK3	LDHA	LDHB	LDHC	LDHD	PFKL	PFKM	PGAM2
	PGK1	PGK2	PGM1	PGM2	PGM3	PKLR	PKM	TPI1	FBP1	G6PC
	G6PC3	PDK1	PDK2	PDK3	PDK4	PDP2	PDPR			
Glucose transporters	SLC2A1	SLC2A10	SLC2A11	SLC2A12	SLC2A13	SLC2A14	SLC2A2	SLC2A3	SLC2A4	SLC2A5
	SLC2A6	SLC2A7	SLC2A8	SLC2A9	SLC5A1	SLC5A10	SLC5A11	SLC5A12	SLC5A2	SLC5A3
	SLC5A4	SLC5A5	SLC5A6	SLC5A7	SLC5A8	SLC5A9				
Lactate-pyruvate-citrate	SLC16A1	SLC16A2	SLC16A3	SLC16A4	SLC16A5	SLC16A6	SLC16A7	SLC16A8	SLC16A9	SLC16A10
	SLC16A11	SLC16A12	SLC16A13	SLC16A14	SLC13A5					
Glycogenolysis	GYG1	PYGM	PGM1	PGM2	PGM3	PGM5	GBE1			
PPP (pentose phosphate pathway)	G6PD	H6PD	PGLS	PRPS1	PRPS2	RBKS	RPE	RPIA	TALDO1	TKT
Glycogen metabolism	PGM1	GYS1	UGP2	PYGM						
Glycerol-phophate	SLC37A1	SLC37A2	SLC37A3	SLC37A4						

Lipid Metabolism

FA oxidation	ACAT2	ACAT1	ACAA1	ACAA2	HADHB	HADH	HADHA	EHHADH	ECHS1	ACOX3
	ACOX1	ACADS	ACADM	ACADL	ACADSB	ACADVL	GCDH	ACSL6	ACSL4	ACSL1
	ACSL5	ACSL3	ACSBG1	ACSBG2	CPT1	CPT1B	CPT1C	CPT2	ECI1	ECI2
	CYP4A11	CYP4A22	ADH1A	ADH1B	ADH1C	ADH7	ADH4	ADH5	ADH6	ALDH2
	ALDH3A2	ALDH1B1	ALDH7A1	ALDH9A1						
FA synthesis	ACACA	ACACB	ACSF3	MCAT	FASN	OXSM	CBR4	HSD17B8	HTD2	MECR
	OLAH	ACSL6	ACSL4	ACSL1	ACSL5	ACSL3	ACSBG1	ACSBG2		
FA elongation	ACAA2	HADHB	HADH	HADHA	ECHS1	MECR	PPT1	PPT2	ELOVL1	ELOVL2
	ELOVL3	ELOVL4	ELOVL5	ELOVL6	ELOVL7	HSD17B12	HACD2	HACD1	HACD4	HACD3
	TECR	ACOT4	ACOT2	ACOT1	ACOT7	THEM4	THEM5			
Choline metabolism	PEMT	CHPT1	PLD1	PLD2	PLD3	SGMS1	SGMS2	PPAP2A	PPA2PB	PCYT1A
	PCYT1B	PHOSPHO1	PHOSPHO2	SMPD1	SMPD2	SMPD3	SMPD4	SPTLC1	SPTLC2	SPTLC3
	CERK	CERKL	UGCG	SGPP1	SGPP2	SPHK1	SPHK2	ASAH1	ASAH2	ACER1
ChoK enzymes	CHKA	CHKB								
Choline transporters	SLC22A1	SLC22A2	SLC22A3	SLC22A4	SLC44A1	SLC44A2	SLC44A3	SLC44A4	SLC44A5	CHAT
	SLC18A3									
Cholesterol metabolism	HMGCS2	HMGCR	PMVK	MVD	FDPS	SQLE				
LPA-DAG metabolism	CHDH	CEPT1	LPCAT1	LPCAT2	LPCAT3	LPCAT4	DGAT1	DGAT2	PEBP1	PTDSS1
	PLA1A	PLA2G10	PLA2G12A	PLA2G12B	PLA2G15	PLA2G16	PLA2G1B	PLA2G2A	PLA2G2C	PLA2G2D
	PLA2G2E	PLA2G2F	PLA2G3	PLA2G4A	PLA2G4B	PLA2G4C	PLA2G4D	PLA2G4E	PLA2G4F	PLA2G5
	PLA2G6	PLA2G7	PLCB1	PLCB2	PLCB3	PLCB4	PLCD1	PLCD3	PLCD4	PLCE1
	PLCG1	PLCG2	PLCH1	PLCH2	PLCL1	PLCL2	PLCXD1	PLCXD2	PLCXD3	PLCZ1
Fatty Acid	SLC27A1	SLC27A2	SLC27A3	SLC27A4	SLC27A5	SLC27A6	SLC27A7	SLC27A8	SLC27A9	SLC27A10
	SLC27A11	FFAR1	FFAR2	FFAR3	FFAR4	GPR84	FAPB1	FABP2	FABP3	FABP4
	FABP5	FABP6	FADS1	FADS2	FADS3	FADS6	FAT1	FAT2	FAT3	FAT4

Apolipoproteins	APOE	APOC1	APOL3	APOO	APOL4	APOBR	APOD	APOL2	APOM	APOL1

Amino acid metabolism

Amino acid	MAT2B	MPST	GNMT	BHMT2	DMGDH	CBS	GLDC	GOT1	SDS	AFMID
	TDO2	PHGDH	PAH	ALDH6A1	GPT	GPT2	CPS1	GLS2	ASS1	PRODH
	GLS	ASL	OAT	OTC	ARG2					
Polyamine	ODC1	ALDH18A1	PYCR1	PRODH	ODC1	PAOX	SMOX	APRT	SMS	SAT1
	AMD1	PRPSAP1	MAT2B							
AA transporters	SLC7A1	SLC7A2	SLC7A3	SLC7A4	SLC7A5	SLC7A6	SLC7A6OS	SLC7A7	SLC7A8	SLC7A9
	SLC7A10	SLC7A11	SLC7A13	SLC7A14	SLC7A15P					
Glutaminolysis	GLS	GDH	GLUD1	GOT1	GOT2	GPT	GPT2			
Glutamine transporters	SLC1A1	SLC1A2	SLC1A3	SLC1A4	SLC1A5	SLC1A6	SLC1A7	SLC7A8	SLC7A11	
Neutral AA	SLC38A1	SLC38A2	SLC38A3	SLC38A4	SLC38A5	SLC38A6	SLC38A7	SLC38A8	SLC38A9	SLC38A10
	SLC38A11	SLC3A1	SLC3A2							
AA and monoamines	SLC6A1	SLC6A2	SLC6A3	SLC6A4	SLC6A5	SLC6A6	SLC6A7	SLC6A8	SLC6A9	SLC6A10P
	SLC6A11	SLC6A12	SLC6A13	SLC6A14	SLC6A15	SLC6A16	SLC6A17	SLC6A18	SLC6A19	SLC6A20
	SLC6A21P									
Monoamines	SLC18A1	SLC18A2	SLC18A3	SLC18B1	SLC32A1	SLC33A1				
Large neutral AA	SLC43A1	SLC43A2	SLC43A3							
Peptides	SLC15A1	SLC15A2	SLC15A3	SLC15A4	SLC15A5	SLC36A1	SLC36A2	SLC36A3	SLC36A4	